Quantification of Location, Dispersion, Skewness and Kurtosis

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der KIT-Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

Andreas Eberl, M.Sc.
geboren in Bruchsal

Tag der mündlichen Prüfung:	13. Juli 2022
Referent:	PD Dr. Bernhard Klar
Korreferenten:	Prof. Dr. Alfred Müller
	Prof. Dr. Tilmann Gneiting

Oma und Opa gewidmet, für meine Wurzeln

Danksagung

Diese Arbeit ist während meiner Zeit als wissenschaftlicher Mitarbeiter am Institut für Stochastik am KIT entstanden. An dieser Stelle möchte ich mich bei den Personen bedanken, die mich auf dem Weg hierher unterstützt haben.

Der erste Dank geht an meinen Betreuer PD Dr. Bernhard Klar, der mich von meiner einführenden Statistik-Vorlesung an über beide Abschlussarbeiten bis zu diesem Punkt begleitet hat. Über meine gesamte Zeit am Institut bin ich bei ihm stets auf ein offenes Ohr gestoßen, egal ob zum Thema Lehre, Forschung oder darüber hinaus. Mir wurden große Freiheiten für eigene Ideen gelassen, aber ich wurde auch sanft gebremst, wenn mein Idealismus überhandgenommen hat. Nicht zuletzt hat er mich auch in meiner nicht immer einfachen Finanzierungssituation unterstützt. Ich erachte keinen dieser Punkte für selbstverständlich und schätze mich entsprechend glücklich, dass sie mir dennoch zuteil wurden.

Ich bedanke mich bei Prof. Dr. Alfred Müller für die Übernahme eines Korreferats. Seine umgängliche Art sowie seine Expertise auf dem Gebiet der stochastischen Ordnungen kamen mir an mehreren Stellen zugute.

Prof. Dr. Tilmann Gneiting danke ich ebenfalls für seine Bereitschaft, ein Korreferat zu übernehmen. Ihn durfte ich bereits in meiner Anfangszeit am Institut als überaus kompetenten, enthusiastischen und offenen Vertreter unseres Fachs kennenlernen.

In meiner gesamten Zeit am Institut für Stochastik hat sich mein Arbeitsplatz dort wie ein zusätzliches Zuhause angefühlt. Dafür danke ich allen aktuellen und ehemaligen Kollegen und Wegbegleitern, die dazu beigetragen haben. Die Freundschaften, die hier entsprungen sind, und die Momente, die hier mit mir geteilt wurden, sind von unschätzbarem Wert. Alle meine zukünftigen Arbeitsumfelder haben große Fußstapfen zu füllen.

Auch in meiner schulischen Ausbildung wurde meine Begeisterung für Mathematik gefördert. Besonderer Dank gilt hier Frau Seebach und Herrn Pfefferle, die diesen Enthusiasmus vorgelebt haben, auch über den alltäglichen Unterricht hinaus.

Ich kann die Dankbarkeit nicht in Worte fassen, die ich gegenüber meiner Familie empfinde, für ihre ständige Unterstützung und den Rückhalt, den sie mir bietet. Das bezieht sich in besonderem Maße auf meine Eltern sowie meine allerersten Mathelehrer, meine Geschwister.

Schließlich danke ich Julia dafür, dass sie nicht müde wird, mir die Bedeutung des Wortes „positiv“ zu zeigen, wie es die Mathematik niemals könnte.

Prior Publications

The contents of Section 3.2 are partly taken from the following prior publications:

Eberl and Klar (2022a),
Eberl and Klar (2022b).

Contents

Chapter 1

Introduction

1.1. Overview and Motivation

Probability distributions, along with directly related concepts like probability measures and random variables, are the most essential objects in the field of stochastics. Thus, being able to describe them as concisely as possible is of utmost importance for both probability theory and statistics. That description can either be in absolute terms or relative to other distributions. The most common tools to achieve this in a rigorous way are stochastic orders and measures, both of which focus on specific features or characteristics instead of the distribution as a whole.

The nature of these characteristics is dependent upon the set on which the probability distributions are defined. If we associate a distribution P with a random variable X, P describes how likely it is that X takes on certain values. The set of all possible values of X is the set on which the probability distribution P is defined. If X describes the outcome of a dice toss, that set could be written as $\{1, 2, 3, 4, 5, 6\}$; if X describes the filling level of a one meter high barrel, it could be written as $[0, 1]$. Note that it is not always possible to describe all potential outcomes of a random experiment with the natural or the real numbers. For example, if the location of a particle at a given time is described, the outcomes are multivariate. However, in this work, we restrict ourselves to probability distributions on the real line and subsets thereof.

The most basic characteristic of a distribution on the real line is its location. As the most popular measures of (central) location, the mean and the median were used as tools long before they were considered as concepts themselves, for example in ancient Greek mathematics (see Bakker and Gravemeijer, 2006). The usual stochastic order, which is both the most basic stochastic order and a location order, was introduced much later by Mann and Whitney (1947, p. 50). Subsequently, the usual stochastic order was used to define measures of location in general (see Bickel and Lehmann, 1975b, p. 474, or Doksum, 1975, p. 11). If X and Y are random variables, X precedes Y in the usual stochastic order (in short: $X \leq_{st} Y$), if $F(t) \geq G(t)$ for all $t \in \mathbb{R}$, where F and G denote the cumulative distribution functions (cdf's) of X and Y, respectively. Now the crucial property of a measure of central location ν is that $\nu(X) \leq \nu(Y)$ holds if $X \leq_{st} Y$. Verbally, ν is required to preserve any ordering between two distributions given by $\leq_{st}$. In this context, a random variable, its probability distribution, and the corresponding cdf are all used interchangeably. Note that it is possible for two random variables X and Y to satisfy neither $X \leq_{st} Y$ nor $Y \leq_{st} X$. The other property a measure of central location ν is required to fulfil, concerns the behaviour of ν under affine linear transformations of a random variable. Specifically, $\nu(a \cdot X + b) = a \cdot \nu(X) + b$ for all $a, b \in \mathbb{R}$.

Another quantity that is frequently associated with the central location of a distribution is its mode, i.e. the point that maximizes the likelihood over the given distribution, which is sometimes difficult to handle, e.g. for multimodal distributions. Furthermore, there are variations of mean and median such as trimmed means or generalized quantiles. Of the latter family, the most well known and useful representatives are expectiles, which were introduced by Newey and Powell (1987). Expectiles can be understood as quantiles of a transformed and thereby smoothed distribution (see Jones, 1994), so that they are always unique. It has been shown that the p-expectile for $p \in (0, 1)$ preserves the usual stochastic order (see Bellini, 2012) and, as a special case, the mean is included as the $\frac{1}{2}$-expectile. So, while this does not result in a new measure of central location, it is another useful way to characterize a distribution that conserves a number of desirable properties of quantiles while also offering some advantages. A helpful way to think of expectiles is that they are to quantiles what the mean is to the median.

The monographs of Shaked and Shanthikumar (2006), Müller and Stoyan (2002) and Belzunce et al. (2015) contain numerous properties of the usual stochastic order that are useful in different stochastic fields. In particular, the two latter references discuss in some detail what $F \leq_{st} G$ graphically means for two cdf's F and G. They state that $F \leq_{st} G$ is equivalent to the corresponding P-P-plot lying below the 45°-diagonal and, similarly, to the Q-Q-plot lying above the 45°-diagonal (see p. 4 in Müller and Stoyan, 2002 or pp. 35-36 in Belzunce et al., 2015). If F and G are continuous, the graph of the P-P-plot corresponds to that of the function $G \circ F^{-1}$; if F and G are strictly increasing, the graph of the Q-Q-plot corresponds to that of the function $G^{-1} \circ F$. The latter statement is contrary to Müller and Stoyan (2002, p. 4), who require F and G to be continuous. However, this error is rectified in Proposition 6.1 of this thesis. Here, F^{-1} and G^{-1} denote the corresponding quantile functions for which the usual definition

$$F^{-1}(p) = \inf\{t \in \mathbb{R} : F(t) \geq p\}, \quad p \in (0, 1)$$

is used. Further treatments of the usual stochastic order on discrete distributions are rare in the literature. Examples include Klar et al. (2010) and Klenke and Mattner (2010), which discuss the use of different measures of central location and equivalent characterizations of the stochastic order for a number of discrete distribution families.

Beside measures of central location, measures of dispersion are often among the first topics in an introductory statistics course. As the most popular such measure today, the standard deviation, along with the closely related variance, has been used at least since the beginning of the 19th century (see Kourkoulos and Tzanakis, 2010). Their first uses were mostly in the context of applications in the field of physics, using the method of least squares. However, a historically more relevant measure was the so-called 'probable error', which is closely related to the presently used interquartile range (see Hald, 1998, p. 360). Further popular dispersion measures include the mean absolute deviation and Gini's mean difference.

The first stochastic order with respect to dispersion was proposed by Birnbaum (1948) as an order of 'peakedness', which considers dispersion around a fixed point. Bickel and Lehmann (1976, 2012) differentiate between 'dispersion' in symmetric distributions and 'spread' in asymmetric distributions. In the latter work (first published in 1976), they defined what was later coined the dispersive order by Lewis and Thompson (1981). This stochastic order of dispersion, which is the most basic order of this kind, is reminiscent of an equivalent characterization of the usual stochastic order. Note that $F \leq_{st} G$ holds if and only if $F^{-1}(p) \leq G^{-1}(p)$ for all $p \in (0,1)$. In a similar way, F is said to precede G in the dispersive order (in short: $F \leq_{disp} G$), if

$$F^{-1}(p_1) - F^{-1}(p_0) \leq G^{-1}(p_1) - G^{-1}(p_0) \quad \forall\ 0 < p_0 \leq p_1 < 1.$$

Two previously published papers, which both utilize a preprint of the paper by Lewis and Thompson (1981), apply this order to specific distribution families (Saunders and Moran, 1978) and to point process theory (Saunders, 1978).

A number of properties of the dispersive order can be found in the monographs of Shaked and Shanthikumar (2006), Müller and Stoyan (2002) and Belzunce et al. (2015). Here, the fact that $\leq_{disp}$ is stronger than other orders of dispersion like the dilation order $\leq_{dil}$ is of particular interest (see Shaked and Shanthikumar, 2006, p. 154). X precedes Y in the dilation order if $\mathbb{E}[\varphi(X - \mathbb{E}[X])] \leq \mathbb{E}[\varphi(Y - \mathbb{E}[Y])]$ holds for all convex functions φ for which the expected values exist. Overall, under some regularity conditions, the dispersive order is the strongest commonly used order of dispersion.

Since measures of dispersion are one of the most widely used statistical concepts that is also frequently used in more applied scientific disciplines, their definition must not exhibit any lack in rigour. This definition is analogous to the definition of measures of central location with the crucial condition being that any dispersion measure τ preserves an order of dispersion. The canonical choice for the dispersion order to be used in that definition is the dispersive order $\leq_{disp}$ because, as the strongest sensible order, it imposes a minimal requirement on any measure to correctly identify differences in dispersion. Furthermore, the dilation order as the other major dispersion order, by definition, favours expectation- and moment-based measures of dispersion, whereas the dispersive order is solely based on a pointwise comparison between two cdf's. The dispersive order is often used as this kind of foundational order in the literature, e.g. by Bickel and Lehmann (1976) and Oja (1981). Oja also shows that a number of the previously mentioned popular dispersion measures preserve the dispersive order.

Most detailed treatments of the dispersive order assume throughout that all involved distributions are sufficiently regular, meaning that their supports are intervals and that the corresponding cdf's are (absolutely) continuous (see, e.g., Oja, 1981, Shaked, 1982 and Belzunce et al., 2015 in most results). However, Müller and Stoyan (2002, p. 41) prove that not only do a number of properties of $\leq_{disp}$ not hold for discrete distributions, but the order itself is virtually useless. Their seemingly minor result states that $\text{range}(G) \subseteq \text{range}(F)$ is a necessary condition for $F \leq_{disp} G$. Since the ranges of all discrete cdf's only consist of atoms, all discrete distributions that are not closely related to each other are not accounted for by the dispersive order. This implication even seems to surprise Müller and Stoyan themselves, who contradict their own result by falsely stating on p. 63 that the geometric distribution is generally compatible with the dispersive order. Thus, there exists no rigorous foundation for the use of traditional dispersion measures on discrete distributions as they are not known to preserve a fundamental order of dispersion. Although measures like the standard deviation of discrete distribution families are still standardly included in introductory textbooks and used for the evaluation and comparison of empirical distributions, virtually no literature addressing this problem can be found. An exception is posed by a paper by Oja (1983) treating the 'scatter' of multivariate empirical distribution.

A number of parallels can be found between the quantification methods of (central) location and dispersion, both in terms of stochastic orders and in terms of measures. First, the arguably most popular measures for both characteristics are constructed similarly. The mean is just the first moment while the standard deviation is the square root of the centralized second moment. Here, the centralization acts as a standardization with respect to the location, which is measured by the mean. A consequent continuation of this concept is to consider the third root of the third moment, which is standardized both with respect to location and dispersion. The resulting quantity

$$\sqrt[3]{\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]},$$

where μ and σ denote the mean and the standard deviation of X, was first considered by Pearson (1895) as a measure of skewness. As opposed to the central location ('Where does the centre of the distribution lie?') and dispersion ('How spread out is the distribution?'), the concept of the skewness of a distribution is slightly less easy to grasp. However, as Arnold and Groeneveld (1993) put it, 'skewness is asymmetry, plain and simple.' These concepts are

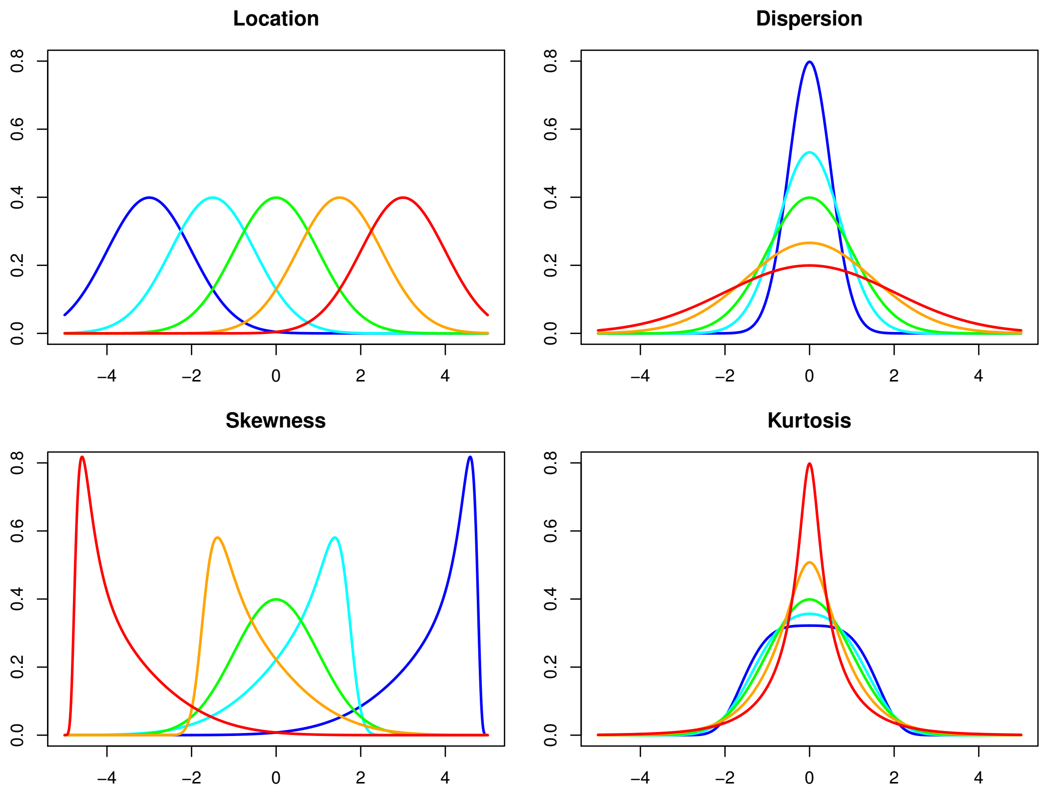

Figure 1.1.: Illustration of the zeroth to third convex characteristics through the densities of five distributions. In each panel, the corresponding characteristic is lowest in the blue distribution and highest in the red distribution.

illustrated in the upper left three panels of Figure 1.1.

Besides the moment-based measures, the popular quantile-based measures of central location, i.e. the median, and dispersion, i.e. the interquartile range, can also be continued in a similar fashion. Instead of the zeroth and the first order differences, the second order difference

$$\left(F^{-1}(\tfrac{3}{4}) - F^{-1}(\tfrac{1}{2})\right) - \left(F^{-1}(\tfrac{1}{2}) - F^{-1}(\tfrac{1}{4})\right) = F^{-1}(\tfrac{3}{4}) - 2F^{-1}(\tfrac{1}{2}) + F^{-1}(\tfrac{1}{4})$$

can be considered. This quantity, divided by the interquartile range, was proposed as a skewness measure by Bowley (1901) and later generalized by David and Johnson (1956), who substituted $\frac{1}{4}$ and $\frac{3}{4}$ for $\alpha \in (0, \frac{1}{2})$ and $1 - \alpha$. The division by the interquartile range or a generalization thereof is done to standardize with respect to dispersion. It is structurally obvious that this quantile-based skewness measure centres the distribution around the median and compares its right side up to a certain quantile to the corresponding left side in order to analyze for asymmetries.

The concept of skewness can also be obtained in the form of an ordering by continuing the ideas of orders of location and dispersion. There are two possible ways of doing this, based on the two major orders of dispersion $\leq_{disp}$ and $\leq_{dil}$, and resulting in two different orders of skewness. For this, we assume F and G to be two cdf's with interval support that are sufficiently often continuously differentiable. The usual stochastic order can now be equivalently characterized by $\Delta_{FG}(t) \geq 0$ for all $t \in \operatorname{supp}(F)$, where $\Delta_{FG}(t) = G^{-1}(F(t)) - t$ and $\operatorname{supp}(F)$ denotes the support of F (see Oja, 1981, p. 156). For other values of t, $F(t)$ takes values outside of the interval $(0, 1)$, on which the quantile function G^{-1} is not defined. The function Δ_{FG} was considered extensively by Oja (1981) and is closely related to the probability integral transform. The latter states that if $X \sim F$ and U is uniformly distributed on the unit interval (or $U \sim \mathcal{U}([0, 1])$ for short), then $F(X) \sim \mathcal{U}([0, 1])$ and $F^{-1}(U) \sim F$ hold under some regularity conditions (see Ferguson, 1967, p. 216 or Rüschendorf, 1981, p. 331). Thus, for $X \sim F$, $G^{-1}(F(X)) \sim G$ follows, so that $\Delta_{FG}(X)$ is the difference of an F-distributed and a G-distributed random variable, implying that the function Δ_{FG} is perfectly fit for the comparison of F and G. Functions of the form $G^{-1} \circ F$ were termed relative inverse distribution functions (RIDF's) by Müller and Stoyan (2002, p. 41), Δ_{FG} can be seen as a modified version of this. Alternatively to $G^{-1} \circ F$, which is associated with the Q-Q-plot, one can also compare two distributions using functions of the form $G \circ F^{-1}$, which are associated with the P-P-plot. The latter function is related to the so-called Lorenz curve, which is used to measure inequality (of wealth, income, etc.); a theoretical treatment with applications in social sciences can be found in Handcock and Morris (1998, 1999).

Similarly to the usual stochastic order, the dispersive order can also be rewritten in terms of Δ_{FG}. To be precise, $F \leq_{disp} G$ is equivalent to $\Delta_{FG}(s) - \Delta_{FG}(t) \geq 0$ for all $t, s \in \operatorname{supp}(F), t < s$, which is equivalent to Δ_{FG} being increasing on the support of F (see Müller and Stoyan, 2002, p. 41). Utilizing the ideas about general convexity put forth by Karlin and Novikoff (1963) and Karlin (1968) and applied to this topic by Oja (1981), non-negativity is equivalent to convexity of order zero and being increasing is equivalent to convexity of order one. Hence, F should precede G in a skewness order if Δ_{FG} is convex of order two. Since a function is convex of order $k \in \mathbb{N}_0$ if and only if its k-th derivative is non-negative, provided that derivative exists, convexity of order two is just usual convexity (see Karlin, 1968, p. 23). The idea to consider G at least as skewed as F, if Δ_{FG} is convex, has merit, as it was developed independently of location and dispersion orders by van Zwet (1964). This so-called convex transformation order, denoted by $\leq_c$ or $\leq_2$, has since been the most popular and fundamental order of skewness in the literature (see, e.g., Oja, 1981, MacGillivray, 1986 or Arnold and Groeneveld, 1993). The three cited papers also propose numerous of further skewness orders that are weaker than the convex transformation order. Most of them utilize concrete measures of location or dispersion in order to standardize the distributions in some way, which is undesirable for a fundamental order. One notable weakening is defined by the existence of an affine linear transformation $\tilde{G}$ of G such that $F - \tilde{G}$ changes sign exactly twice, from '+' to '−' to '+' (see Oja, 1981, p. 162). Based on an underlying result by Karlin (1968, p. 281-282), this kind of weakening also exists for other orders that are defined via convexity of order $k \in \mathbb{N}_0$ of Δ_{FG}.

An equivalent characterization of the stochastic order that is more reminiscent of the dilation order reads as follows: $X \leq_{st} Y$, if and only if $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(Y)]$ for all increasing functions φ for which the expectations exist. Recall that $X \leq_{dil} Y$ is defined by $\mathbb{E}[\varphi(X - \mathbb{E}[X])] \leq \mathbb{E}[\varphi(Y - \mathbb{E}[Y])]$ for all convex functions φ. This concept can also be continued one order of convexity higher by considering Y at least as skewed as X, if

$$\mathbb{E}\left[\varphi\left(\frac{X-\mu_X}{\sigma_X}\right)\right] \leq \mathbb{E}\left[\varphi\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right]$$

for all functions φ that are convex of order three. This latter order is seldom explicitly considered as a skewness order in the literature. However, a general result by Klar (2002, p. 13) implies that it is a weakening of the convex transformation order, which suggests that it may be used in that capacity. Furthermore, orders of this kind were considered by Fishburn (1976, 1980), Rolski (1976), and Rolski and Stoyan (1974) in the context of queuing theory; a brief summary can be found in Müller and Stoyan (2002, pp. 37–40).

Overall, a tight connection between notions of location, dispersion and skewness of a distribution can be found both in measure- and order-based quantification methods. Both considered types of orders connect these characteristics of a distribution to the concept of convexity of different orders. A connection from convex functions to the moment-based measures is made by van Zwet (1964) while a connection to the quantile-based measures can be founded on the interplay between higher-order convexity and the concept of divided differences, considered by Mühlbach (1973), among others. Since this thesis is mostly focused on orders like $\leq_{st}$, $\leq_{disp}$ and $\leq_c$, one might refer to the location of a distribution as its zeroth convex characteristic, to the dispersion as its first convex characteristic and to skewness as its second convex characteristic. So, is there also a third convex characteristics and are all aforementioned concepts generalizable one order higher?

This third characteristic is commonly referred to as kurtosis and it was first considered by Pearson (1895) in the form of the standardized fourth moment

$$\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right].$$

Compared to the lower-order characteristics, it is much more difficult to pinpoint what exactly is described by the notion of kurtosis. Consequently, this question has been controversially debated in the literature over the years. There exist a number of papers advocating for kurtosis to be understood as the peakedness (see, e.g., Crack, 2019) of the density of a unimodal distribution while others dispute this by stating that kurtosis has much more to do with tailweight (see, e.g., Westfall, 2014). A sensible compromise is found by Balanda and MacGillivray (1988, p. 116), who state that an increase in kurtosis corresponds to 'the location- and scale-free movement of probability mass from the shoulders of a distribution into its center and tails' (see bottom right panel of Figure 1.1).

Numerous works discussing the meaning of kurtosis understand that notion to be synonymous to the standardized fourth moment (see, e.g., Crack, 2019; Darlington, 1970; DeCarlo, 1997; Westfall, 2014). However, since this is analogous to identifying the notion of the central location of a distribution with its mean, kurtosis should be understood in a broader sense. In concurrence with this observation, other kurtosis measures as well as an underlying order-based approach have been proposed in the literature.

Most alternative measures of kurtosis are quantile-based and are constructed in a similar way as the quantile-based measure of skewness mentioned earlier. First, one might simply use the third-order difference of quantiles in the numerator and standardize it with some interquantile distance in the denominator, giving

$$\frac{F^{-1}(1-\alpha) - 3F^{-1}(1-\beta) + 3F^{-1}(\beta) - F^{-1}(\alpha)}{F^{-1}(1-\alpha) - F^{-1}(\alpha)}. \tag{1.1}$$

For specific values of $0 < \alpha < \beta < \frac{1}{2}$, this measure was proposed by Jones et al. (2011). Ruppert (1987) proposed the ratio of two different interquantile distances as a kurtosis measure, which is easily shown to be an equivalent measure to (1.1), see Section 4.3.2. This fits a more general approach by Bickel and Lehmann (1975a), who state that a kurtosis measure is a ratio of two suitable measures of dispersion without going into more detail. Another similarly constructed quantile-based measure was given by Moors (1988) with all of these measures being summarized in a general formula by Jones et al. (2011). Further proposals for kurtosis measures are rare in the literature, with a measure based on so-called L-moments by Hosking (1989, 1990) posing a notable exception.

Measures of dispersion usually contain some standardization with respect to location whereas measures of skewness are standardized with respect to location and dispersion. This is particularly apparent in the corresponding standardized moments, but can also be observed in the popular quantile-based measures. Overall, measures of the k-th convex characteristic are standardized with respect to all previous convex characteristics in order to not interfere with the intended measurement. However, all mentioned kurtosis measures are only standardized with respect to location and dispersion, but not with respect to skewness. While location can be shifted by addition of a constant to the random variable in question and dispersion can be rescaled by multiplication, there is no arithmetic operation to manipulate the skewness of a random variable. On one hand, this means that skewness, kurtosis and even higher convex characteristics are more intrinsic to the shape of a distribution. On the other hand, it means that, generally, kurtosis measurements remain entangled with skewness. This problem also becomes apparent in the well known inequality $\beta_2 \geq \beta_1^2 + 1$ connecting the third and fourth standardized moments β_1 and β_2 (see Pearson, 1916, p. 432). Attempts to rid kurtosis measures of skewness effects were made in the literature by Blest (2003) for moment-based measures and, in a more general way, by Jones et al. (2011) for quantile-based measures. Among other things, Jones et al. (2011) used the family of so-called sinh-arsinh distributions, defined by

Jones and Pewsey (2009), in which skewness and kurtosis can be varied independently from each other. Generally, however, the problem persists, also in the order-based foundation of kurtosis measures.

The first kurtosis order in the literature was proposed by van Zwet (1964) along with the convex transformation order and was only defined for symmetric distributions. That order, denoted by $\leq_s$, deems G more kurtotic than F if the RIDF $G^{-1} \circ F$ is concave for arguments smaller than the center of symmetry and convex for arguments larger than the center of symmetry. Van Zwet also proved that the standardized fourth moment preserves $\leq_s$. A number of other kurtosis orders were given in the literature since, see Arnold and Groeneveld (1993), Balanda and MacGillivray (1988, 1990), MacGillivray and Balanda (1988), and Oja (1981). Some of these works also restrict their attention to symmetric distributions while others do include asymmetric distributions. The mostly used approach to cope with asymmetries is to artificially centre the comparison of two distributions with respect to kurtosis around some value, usually the median. Balanda and MacGillivray (1990) do this by introducing the so-called spread function and using it as a substitute for the cdf. This spread function is obtained by folding the two parts left and right of the median together and then adding them up. In this way, a new, symmetric distribution is obtained to be analyzed with respect to kurtosis.

Since the most fundamental orders of the zeroth, first and second convex characteristics are all constructed in the same way, a canonical candidate for a kurtosis order is obtained by using the same principle. Then, G is deemed more kurtotic than F if $G^{-1} \circ F$ is convex of order three. Although this seems to be an obvious consideration, this order is barely considered in the literature. While Hosking (1989, p. 6) uses it to show the validity of his kurtosis measure based on L-moments, Oja (1981, p. 168) mentions it briefly, only to dismiss it because of its lack of transitivity. In spite of that observation, the apparent lack of further papers treating this order is surprising.

As mentioned before, the concepts of location and dispersion are much easier to grasp and also much more well known than the higher order characteristics skewness and kurtosis. While it is often very useful that probability distributions can easily be standardized with respect to location and dispersion, this also means that these two characteristics can be seen as not particularly intrinsic to the distribution. Contrarily, both the skewness and the kurtosis of a distribution are representative of it in a way that cannot be changed by a simple transformation. This is why skewness and kurtosis have often been used as the critical parameters for categorizing distributions with respect to their shape. The oldest and most well known of these attempts are the Pearson families of distributions proposed in Pearson (1895, 1901, 1916). These families cover the entirety of the so-called skewness-kurtosis-plane, which contains all possible combinations of values of the standardized third and fourth moments (see, e.g., Rhind, 1909, Johnson et al., 1994, pp. 15–25 and Lahcene, 2013). Unlike the skewness-kurtosis-plane in Figure 1.2, the x-axis usually describes the square

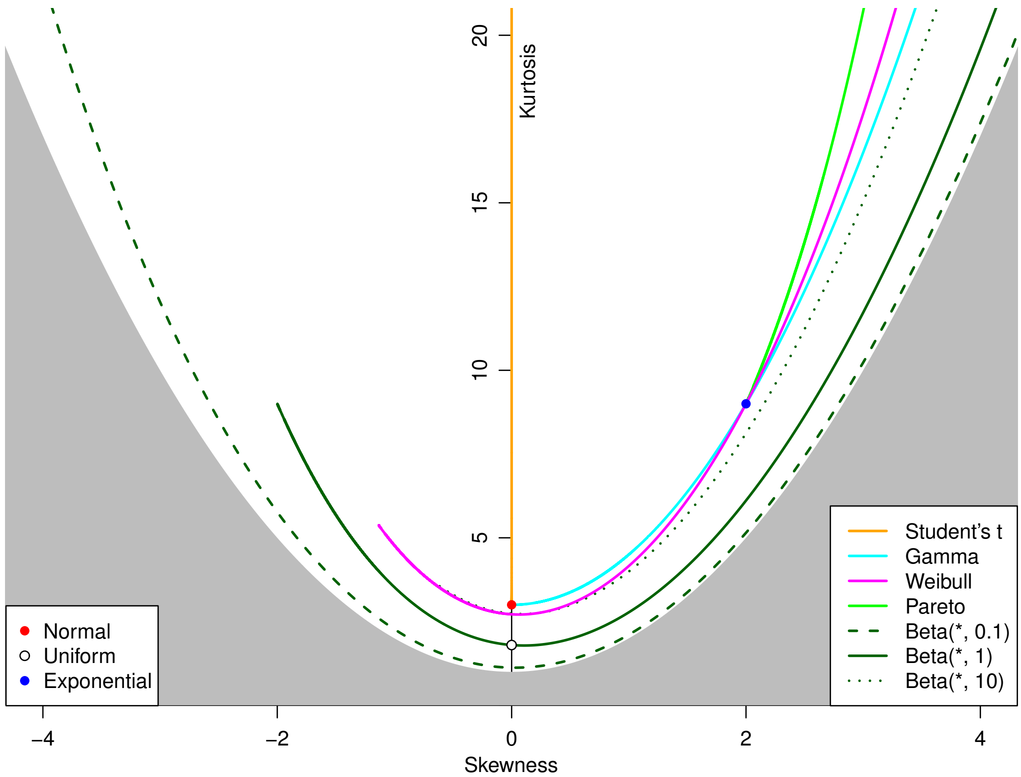

Figure 1.2.: Skewness-kurtosis-plane for a number of well-known distribution families. The skewness and kurtosis values represented by the grey area cannot be attained by any distribution.

of the standardized third moment in historical representations. Since the pioneering work by Pearson, a number of further distribution families have been proposed, which parametrize their shape via skewness and kurtosis. Examples include the skew-normal distributions by Azzalini and subsequent skew-t distributions (Azzalini, 1985; Azzalini and Capitanio, 2003), Tukey's g-and-h or g-and-k distributions (Hoaglin, 1985; MacGillivray and Cannon, 1997; Tukey, 1977) and the sinh-arsinh distributions (Jones and Pewsey, 2009).

The categorization of distributional shape via skewness and kurtosis is used in a variety of applications including finance, physics and other disciplines of science (see, e.g., Corrado and Su, 1996, Cristelli et al., 2012 or Martins, 1965). It can also be used to quantify how close an observed distribution is to known theoretical distributions. An example is given by a test of a given sample on normality, which is possible since the normal distributions only vary in location and dispersion (see, e.g., D'Agostino et al., 1990 or Hopkins and Weeks, 1990). In the vast majority of these applications, the moment-based measures of skewness and kurtosis are used.

1.2. Outline of the Thesis

This thesis is structured around two main parts, which cover two gaps in the literature identified in the overview given above. Their topics are the quantification of dispersion for discrete distributions and the quantification of kurtosis for asymmetric distributions. Before getting to these main parts, Chapter 2 introduces a number of general concepts that are of particular relevance for the subsequent considerations. These concepts range from basic knowledge and notation in stochastic order relations over the foundations of convex functions and convex characteristics to expectiles. A particular focus is laid on the order of the k-th convex characteristic and corresponding criteria. While Oja (1981) systematically analyzed these orders for $k \in \{0, 1, 2\}$, this methodology is generalized to $k \in \mathbb{N}$ in this thesis.

Chapters 3, 4 and 5 then comprise Part I in which we restrict ourselves to the consideration of sufficiently regular absolutely continuous distributions. This restriction yields a number of helpful simplifications in the description and handling of the involved orders and measures.

Since the convex characteristics hierarchically build on one another, the quantification of location, dispersion and skewness is discussed in Chapter 3 ahead of the first main part on kurtosis in Chapter 4. A number of approaches to measure these first three convex characteristics are presented. These include well established measures based on moments and quantiles, less prominent approaches like the so-called L-moments, and new classes of measures like the density-based measures in Section 3.1.3. The separate Section 3.2 is devoted to expectile-based ways of quantifying these characteristics, which includes both orders and measures. These are mostly defined analogously to quantile-based orders and measures. A number of results are proved, which clarify the relationship between these new ideas and traditional approaches.

The following Chapter 4 is the first of two main parts of the thesis. After starting with a brief summary of the discussion on the meaning of kurtosis in the literature, different kurtosis orders are considered. It is confirmed that $\leq_3$, the order of the third convex characteristic, is indeed generally not transitive, as noted by Oja (1981, p. 168). However, the notion of transitivity sets is then introduced, describing sets of distributions on which the order $\leq_3$ is transitive. All of the derived transitivity sets have in common that they contain distributions that have the same skewness in some sense. In particular, the set $\mathcal{S}$ of all symmetric distributions is also a transitivity set. Since the concept of kurtosis is usually only analyzed for symmetric distributions, this invalidates the dismissal of $\leq_3$ as a suitable kurtosis order. This observation is subsequently strengthened by the fact that concave-convex orders, which are the most popular kurtosis orders in the literature, have worse transitivity properties than $\leq_3$. Furthermore, the definition of the concave-convex orders $\leq_s$ and $\leq_a$ from the literature (see van Zwet, 1964 and MacGillivray and Balanda, 1988) are modified to be more adaptive to asymmetric distributions. We then apply both kinds of kurtosis orders to a number of well known distribution families, which vary in both skewness and kurtosis.

Particular emphasis is put on the sinh-arsinh distributions proposed by Jones and Pewsey (2009), which have four parameters for the first four convex characteristics. It is shown that the behaviour of both considered types of kurtosis orders is only dependent on the kurtosis parameters of the distributions and not on the skewness parameters.

The disentanglement of skewness and kurtosis also plays a central role in the analysis of several kurtosis measures in Section 4.3. The notion of kurtosis measures is used in a rather loose sense because the lack of transitivity of the underlying kurtosis orders can be used to show that no such measure exists in the traditional sense. Mappings that were identified as kurtosis measures in the literature or that arise from corresponding measures of location, dispersion and skewness in Section 3 are analyzed. Many of them preserve the order $\leq_3$, if they are restricted to some transitivity set like the symmetric distributions. Some families of measures, in particular those based on densities, also offer further insights on how large values and differences in skewness obscure the measurement of kurtosis.

Part I is concluded by Chapter 5, which summarizes the findings and points out arising questions that could possibly be analyzed in future research. Among other topics, this concerns the specific nature of convex characteristics of higher order than kurtosis and whether one of them is associated with bimodality.

Part II of the thesis, which consists of Chapters 6, 7 and 8 and some supplementary material in Appendix A, deals with the behaviour of orders of convex characteristics on discrete distributions as well as suitable replacements. The necessity to explore this topic in depth is established in Chapter 6. While the stochastic order in Section 6.1 only exhibits minor irregularities in RIDF-based characterizations for discrete distributions, the shortcomings of the dispersive order that are discussed in Section 6.2 are much more severe. Essentially, it is shown to not be a meaningful order of dispersion for discrete distributions, leaving the entire concept of dispersion without foundation.

Chapter 7 seeks to rectify this shortcoming through proposals of modifications of the dispersive order that are more suitable to discrete distributions. We carefully construct discrete dispersive orders from two major starting points. One is an equivalent characterization of the original dispersive order at the edge of its applicability for discrete distributions, and the other is a discrete analogue of how the dispersive order works on continuous distributions. The derivation of these discrete orders contains a number of examples, which are used to ensure that they are meaningful. Furthermore, a number of helpful relations and criteria are introduced in the process. Finally, we obtain two candidates for discrete dispersive orders and subsequently examine whether they fulfil the same properties as the original dispersive order $\leq_{disp}$. Two crucial properties are transitivity and whether the discrete version coincides with the original version on their joint area of applicability. Each candidate for the discrete dispersive order satisfies only one of these two properties. Numerous other properties of the original dispersive order can subsequently be replicated for both candidates.

In the final two sections of Chapter 7, the compatibility of the discrete orders with popular dispersion measures and with popular families of discrete distributions is analyzed. The first part is of particular importance, because it implies whether dispersion measures are suitable to be applied to discrete distributions. Here, we show that the interquantile range is not a meaningful dispersion measure for discrete distributions. The results of the application of the discrete dispersive orders to specific discrete distributions in Section 7.5 are satisfactory: the distributions are ordered in the direction that is heuristically plausible, but a certain difference between the parameter values of the considered distribution families is required.

The concluding Chapter 8 of Part II is subdivided into three sections. First, we discuss a number of alternative approaches to the development of a discrete dispersive order. Most of these approaches can be dismissed, while one of them seems to have enough merit to be pursued further. Second, a discrete skewness order is proposed using a similar approach as in Chapter 7. The fact that this is similarly necessary as for dispersion was demonstrated in Eberl and Klar (2019) and is also fairly obvious from the observations in Chapter 6. Finally, the general techniques that are employed to handle discrete distributions throughout Chapter 7 are discussed. Their merit is evaluated and open questions are posed.

The discussion of discrete distributions in Part II is complemented by Appendix A. Here, the concept of RIDF's, i.e. of functions of the form $G^{-1} \circ F$, where F and G are cdf's, is evaluated for discrete distributions. Similarly to the dispersive order, it is not meaningful for a large set of discrete distributions. We then examine the exact limits of the meaningful applicability of RIDF's and subsequently propose a possible generalization. While the obtained generalization is only partially successful in replicating the properties of RIDF's, interesting information about probability integral transforms is revealed in the process.

Chapter 2

General Concepts

2.1. Probability Distributions and Stochastic Orders

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be our underlying probability space, which is assumed to be rich enough to carry all random objects in this thesis. A random variable X is defined as a measurable mapping $X : \Omega \to \mathbb{R}$ as, for the purposes of this thesis, all random variables are real-valued. The probability distribution or probability measure of X is defined as $\mathbb{P}^X = \mathbb{P} \circ X^{-1}$, where X^{-1} denotes the inverse image function, which maps a (Borel-)measurable subset of $\mathbb{R}$ onto a set in $\mathcal{A}$. The set of all real-valued probability distributions is denoted by $\mathcal{P}$. The cumulative distribution function (cdf) of X is denoted by F and is defined by

$$F : \mathbb{R} \to [0, 1], \quad t \mapsto \mathbb{P}(X \leq t).$$

Analogously, the cdf of a random variable Y is denoted by G. For any random variable Z other than X and Y, the corresponding cdf is denoted by H_Z. Since there exists a one-to-one relationship between real-valued probability distributions and cdf's, we use these two notions interchangeably (see Kallenberg, 2021, pp. 42, 84). This means that a set $\mathcal{Q} \subseteq \mathcal{P}$ of probability distributions can also interpreted as the set of the corresponding cdf's and vice versa. The fact that X is distributed according to F or $\mathbb{P}^X$ is denoted by $X \sim F$ or $X \sim \mathbb{P}^X$.

The support of the probability distribution of X is defined by as the smallest closed subset $\operatorname{supp}(\mathbb{P}^X) \subseteq \mathbb{R}$ such that $\mathbb{P}(X \in \operatorname{supp}(\mathbb{P}^X)) = 1$ (see Bogachev, 2007, p. 77). Furthermore, define $\operatorname{supp}(F) = \operatorname{supp}(X) = \operatorname{supp}(\mathbb{P}^X)$. Heuristically, the support is the closure of the set of all points with positive probability mass on them. We also define the interval

$$D_F = \mathbb{R} \setminus F^{-1}(\{0, 1\}) = \{t \in \mathbb{R} : F(t) \in (0, 1)\},$$

as well as $D'_F = D_F \cup \{\sup(D_F)\}$. Sets of this kind were also considered by van Zwet (1964,

p. 6), who denoted it by I, and by Oja (1981, p. 155), who denoted it by S_F. The distinction between D_F and D'_F is useful for discrete F.

According to Lebesgue's decomposition theorem (see Hewitt and Stromberg, 1975, p. 337), each probability measure $\mathbb{P}^X$ can be uniquely decomposed as

$$\mathbb{P}^X = a_{ac} \cdot \mathbb{P}^X_{ac} + a_{sc} \cdot \mathbb{P}^X_{sc} + a_d \cdot \mathbb{P}^X_d, \tag{2.1}$$

where $\mathbb{P}^X_{ac}$ is an absolutely continuous probability measure (with respect to the Lebesgue measure), $\mathbb{P}^X_{sc}$ is a singular continuous probability measure, $\mathbb{P}^X_d$ is a discrete probability measure and the coefficients $a_{ac}, a_{sc}, a_d \in [0, 1]$ satisfy $a_{ac} + a_{sc} + a_d = 1$. Because of their limited relevance in applications, we disregard singular continuous measures for the purposes of this thesis. Absolutely continuous and discrete distributions are considered separately in the two Parts I and II. Any mixture of these two kinds of distributions can then be obtained as a linear combination using (2.1). X is an absolutely continuous random variable, if it has a Lebesgue density f such that

$$\mathbb{P}(X \in A) = \int_A f(t)\ \lambda^1(\mathrm{d}t) = \int_A f(t)\ \mathrm{d}t$$

holds for all measurable sets $A \subseteq \mathbb{R}$, where λ^1 denotes the one-dimensional Lebesgue measure. In that case, $F' = f$ holds almost everywhere, so the density function f is the derivative of the corresponding cdf F. If F is differentiable on D_F, f denotes that version of the density. X is a discrete random variable, if $\mathrm{supp}(\mathbb{P}^X)$ is at most countable. In that case, X has a probability mass function (pmf) f that satisfies $f(t) = 0$ for $t \notin \mathrm{supp}(\mathbb{P}^X)$ and

$$\mathbb{P}(X \in A) = \int_A f(s) \left(\sum_{t \in \mathrm{supp}(\mathbb{P}^X)} \delta_t \right) (\mathrm{d}s) = \sum_{t \in \mathrm{supp}(\mathbb{P}^X)} f(t) \cdot \delta_t(A),$$

where $\delta_t(A)$ evaluates the Dirac-measure and is equal to 1 if $t \in A$ and 0 otherwise. The pmf f is also characterized by $f : \mathbb{R} \to [0, \infty), t \mapsto \mathbb{P}(X = t)$. The density function (pmf) of an absolutely continuous (discrete) random variable Y is denoted by g. Note that $\mathbb{P}(X \in A)$ is given as an integral over f in both cases: with respect to the Lebesgue measure for absolutely continuous X, and with respect to a suitable counting measure for discrete X. In the following, we define a number of subsets of the set $\mathcal{P}$ of all real-valued probability distributions.

Definition 2.1. Let $F \in \mathcal{P}$ be a cdf and $k \in \mathbb{N}$. We define:

a) $F \in \mathcal{S}$, if F is symmetric, i.e., if there exists a $t_0 \in \mathbb{R}$ such that $F(t_0 - t) = 1 - F(t_0 + t)$ holds for all $t \in \mathbb{R}$.

b) $F \in \mathcal{D}$, if F is discrete, i.e., if $\mathrm{supp}(F)$ is at most countable.

c) $F \in \mathcal{C}$, if F is absolutely continuous with respect to the Lebesgue measure.

d) $F \in \mathcal{P}^k$, if $F \in \mathcal{C}$ and F is k times differentiable on D_F.

e) $F \in \mathcal{P}_I$, if $F \in \mathcal{C}$ and F has interval support, i.e., if $\operatorname{supp}(F) = \overline{D_F}$.

f) $F \in \mathcal{P}_I^k$, if $F \in \mathcal{C}$, F is k times differentiable on D_F and F has interval support.

g) $F \in \mathcal{L}^k$, if $\mathbb{E}[|X|^k] < \infty$ for $X \sim F$.

Note that, for $k \in \mathbb{N}$, the inclusions $\mathcal{P}_I^{k+1} \subseteq \mathcal{P}^{k+1} \subseteq \mathcal{P}^k \subseteq \mathcal{P}^1 \subseteq \mathcal{C}$ hold, the last one by definition. Furthermore, $\mathcal{P}_I^k = \mathcal{P}^k \cap \mathcal{P}_I$ and $\mathcal{D} \cap \mathcal{C} = \emptyset$. By definition of absolute continuity, all $F \in \mathcal{C}$ are continuous.

For $F \in \mathcal{D}$, $D_F = [\inf(\operatorname{supp}(F)), \sup(\operatorname{supp}(F)))$ and $D'_F = [\inf(\operatorname{supp}(F)), \sup(\operatorname{supp}(F))]$ follows. Thus, $\mathbb{P}(X \in D_F) < 1$ and $\operatorname{supp}(F) \subset D'_F$ hold in that case. For $F \in \mathcal{C}$, it follows that $D_F = (\inf(\operatorname{supp}(F)), \sup(\operatorname{supp}(F)))$. If the assumption is strengthened to $F \in \mathcal{P}_I$, $D_F = \operatorname{int}(\operatorname{supp}(F))$ holds.

The quantile function of a cdf $F \in \mathcal{P}$ is defined by

$$F^{-1} : (0,1) \to \mathbb{R}, \quad p \mapsto \inf\{t \in \mathbb{R} : F(t) \geq p\}. \tag{2.2}$$

If necessary, the quantile function can be extended to the domain of the closed interval $[0,1]$. In that case, the value $F^{-1}(1) = \sup(\operatorname{supp}(F)) = \sup(D_F)$ is obtained. For quantile functions evaluated at zero, we establish the convention $F^{-1}(0) = \inf(D_F)$. However, using the domain $(0,1)$ is usually more practical. Note that, if $F \in \mathcal{P}_I$, F is bijective on D_F, yielding that F^{-1} matches the inverse function of F restricted to the interval D_F.

A possible, more general characterization of the p-quantile Q_F^p, $p \in (0,1)$, is that at least $p \cdot 100\%$ of the probability mass lies on values no larger than Q_F^p and at least $(1-p) \cdot 100\%$ of the probability mass lies on values no smaller than Q_F^p. In short,

$$\mathbb{P}(X \leq Q_F^p) \geq p \quad \text{and} \quad \mathbb{P}(X \geq Q_F^p) \geq 1 - p \tag{2.3}$$

(see Georgii, 2013, p. 231). A necessary condition for either inequality being strict is $\mathbb{P}(X = Q_F^p) > 0$. If F is strictly increasing on D_F, all quantiles of F are unique. However, if F constantly assumes the value $p \in (0,1)$ on some interval, all definitions of the p-quantile within that interval

$$[\inf\{t \in \mathbb{R} : F(t) \geq p\}, \sup\{t \in \mathbb{R} : F(t) \leq p\}]$$

satisfy characterization (2.3) (see, e.g., Artzner et al., 1999, p. 216). Therefore, there is more than one possible definition of the quantile function for discrete distributions. We mostly use the convention (2.2), denoted by $F^{-1}(p)$; quantiles in the sense of (2.3) are denoted by Q_F^p and are discussed further in Section 2.3. Finally, note that the quantile function of F and the inverse image function of F are both denoted by F^{-1}. If the function argument is a number, the quantile function is being evaluated; if the function argument is a set, the inverse image function is being evaluated.

For $F, G \in \mathcal{P}$, the relative inverse distribution function (RIDF) from F to G is given by

$$R_{FG} : D_F \to \mathbb{R}, \quad t \mapsto (G^{-1} \circ F)(t) = G^{-1}(F(t))$$

(see Müller and Stoyan, 2002, p. 3, or Oja, 1981, p. 155). While this definition is sensible for absolutely continuous distributions, the usage of $\operatorname{supp}(F)$ or D'_F as alternative domains of R_{FG} is sometimes useful, e.g. for discrete distributions. The behaviour of RIDF's for discrete distributions is analyzed in detail in Chapter 6 and Appendix A. If necessary, the domain can also be extended to $\mathbb{R}$: for $t > \sup(D_F)$, we then have $R_{FG}(t) = G^{-1}(1) = \sup(D_G)$ and for $t < \inf(D_F)$, we have $R_{FG}(t) = G^{-1}(0) = \inf(D_G)$. The crucial property of a RIDF is that $R_{FG}(X) \sim G$, if $F, G \in \mathcal{C}$ (see van Zwet, 1964, p. 48, Rüschendorf, 1981, p. 331). Verbally, the RIDF from F to G transforms an F-distributed random variable into a G-distributed random variable. Note that the exact domain of R_{FG} is not relevant for this property as long as it almost surely contains all possible values of X. The assumption for this crucial property can be weakened according to Proposition A.3. The RIDF can be modified to

$$\Delta_{FG} : D_F \to \mathbb{R}, \quad t \mapsto R_{FG}(t) - t,$$

yielding that $\Delta_{FG}(X)$ is the difference of an F-distributed random variable and a G-distributed random variable. The ability of the function Δ_{FG} to compare two random variables is expanded upon in Section 2.2.

A stochastic order $\leq_o$ is defined as a binary relation on a suitable subset $\mathcal{Q} \subseteq \mathcal{P}$ of all probability distributions on the real line. We only refer to such a relation as a stochastic order if its purpose is to order distributions with respect to some characteristic, but this is no formal requirement. We sometimes apply stochastic orders to random variables instead of cdf's or distributions, so $X \leq_o Y$ for $X \sim F$ and $Y \sim G$ means $F \leq_o G$.

The main use of stochastic orders in the context of this thesis is to define measures of a certain characteristic of probability distributions. For example, the usual stochastic order $\leq_{st}$ is used the definition of location measures and the dispersive order $\leq_{disp}$ is used in the definition of dispersion measures. If $\leq_o$ denotes an order of a specific characteristic, the crucial property in the definition of a measure ν of this characteristic is given as follows: if $F \leq_o G$ holds for two cdf's F and G, then ν is required to satisfy $\nu(F) \leq \nu(G)$. In that case, we say that ν preserves the order $\leq_o$.

A stochastic order $\leq_o$ induces two further binary relations on the same underlying set $\mathcal{Q}$ of distributions. First, it induces the strict version $<_o$ of itself, which is defined by

$$F <_o G, \text{ if } F \leq_o G \text{ and } G \not\leq_o F \tag{2.4}$$

for $F, G \in \mathcal{Q}$, where $G \not\leq_o F$ simply denotes the negation of $G \leq_o F$. Second, it induces equivalence $=_o$ with respect to itself, which is defined by

$$F =_o G, \text{ if } F \leq_o G \text{ and } G \leq_o F \tag{2.5}$$

for $F, G \in \mathcal{Q}$.

There are two major desirable properties of stochastic orders. Since we generally formulate these orders in a non-strict way, they should be reflexive, i.e., $F \leq_o F$ should hold for all $F \in \mathcal{Q}$. The most crucial property of stochastic orders is transitivity, i.e., that $F \leq_o G$ and $G \leq_o H$ implies $F \leq_o H$ for all $F, G, H \in \mathcal{Q}$. Transitive orders generally behave more intuitively and are more easy to handle. Furthermore, lack of transitivity makes it virtually impossible that a measure ν that preserves the order in question exists. Note that values of ν for different distributions are compared via the transitive '$\leq$'. Therefore, $\leq_o$ not being transitive poses a problem for the existence of measures ν that preserve the order $\leq_o$. This issue is discussed in more detail for the specific order $\leq_3$ in Section 4.2.1.

On the other hand, there are also properties that cannot or should not be satisfied by most stochastic orders. The first property of this kind is totality, meaning that all pairs $(F, G) \in \mathcal{Q}^2$ satisfy $F \leq_o G$ or $G \leq_o F$. This property is usually undesirable because, for most underlying characteristics, there exist pairs of distributions that cannot be ordered unambiguously with respect to that characteristic. This is exemplified in the following for the characteristic of dispersion.

Example 2.2. Let $X \sim \mathcal{N}(0, 1)$ and $Y = 0.725\tilde{Y}$ with $\tilde{Y} \sim t_3$, where t_3 denotes Student's t-distribution with 3 degrees of freedom. The densities of both X and Y are depicted in Figure 2.1. From the plot, it is not obvious which distribution is more dispersed. While the density of Y has fatter tails and therefore more dispersion far away from the centre, the density of X is smaller around the centre, but larger for about $|t| \in (0.45, 2)$, so it seems more dispersed close to the centre.

As examples for popular dispersion measures, we consider the standard deviation (SD), the mean absolute deviation (MAD) and the interquartile distance (IQR). While the MAD's of X and Y are approximately equal (difference $< 0.1\%$), the SD of Y is about 25% larger than that of X, and the IQR of X is about 22% larger than that of Y. Overall, neither random variable is unambiguously more dispersed than the other.

The final property to be discussed here is antisymmetry, meaning that $F \leq_o G$ and $G \leq_o F$ implies $F = G$ for all $F, G \in \mathcal{Q}$. However, if a cdf F is merely shifted to obtain $\tilde{F} = F(\cdot - a)$ for some $a \in \mathbb{R}$, a dispersion order $\leq_D$ should still treat F and $\tilde{F}$ as the same. Therefore, the reflexivity dictates $F \leq_D \tilde{F}$ and $\tilde{F} \leq_D F$ although $F = \tilde{F}$ is obviously not true. A similar argument can be used for orders of other characteristics.

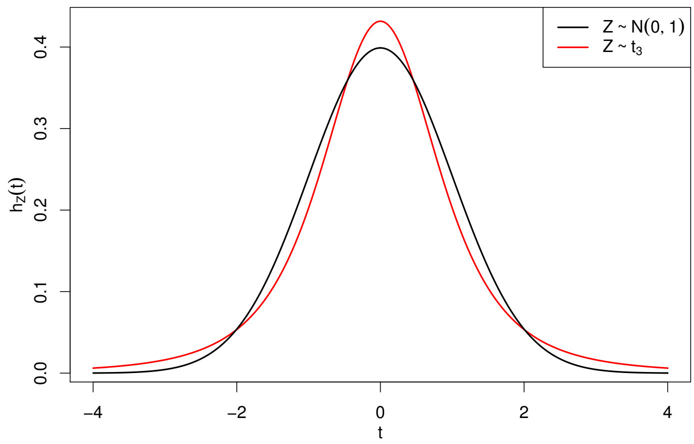

Figure 2.1.: Visualization of Example 2.2.

2.2. Generalized Convexity and Orders of Convex Characteristics

The characteristics of location, dispersion, skewness and kurtosis are related in multiple ways. The connection between basic orders of these characteristics can be established via generalized convex functions. Thus, the representatives of the arising family of orders are dubbed orders of convex characteristics.

The concept of generalized convexity is closely related to that of total positivity, both of which are discussed at length by Karlin (1968), among others. The special case of generalized convexity that is relevant for this thesis corresponds to the concept considered in Oja (1981). Accordingly, we begin with the following definition, derived from Karlin (1968, pp. 280–281) and Oja (1981, p. 155).

Definition 2.3. Let $k \in \mathbb{N}_0$ and let $D \subseteq \mathbb{R}$ have at least $k+1$ elements. Furthermore, let $\varphi : D \to \mathbb{R}$ be a function.

a) For $x_0, x_1, \ldots, x_k \in D$, define the $(k+1) \times (k+1)$-matrix $\Xi_\varphi^k(x_0, x_1, \ldots, x_k) = (\xi_{i,j})_{i,j=1,\ldots,k+1}$ by

$$\begin{aligned} \xi_{k+1,j} &= \varphi(x_{j-1}), \quad && j = 1, \ldots, k+1, \\ \xi_{i,j} &= x_{j-1}^{i-1}, && j = 1, \ldots, k+1,\ i = 1, \ldots, k. \end{aligned}$$

b) φ is said to be *convex of order k* or *k-convex* on D, if

$$\det\left(\Xi_\varphi^k(x_0, x_1, \ldots, x_k)\right) \geq 0 \tag{2.6}$$

holds for all $x_0, x_1, \ldots, x_k \in D$ with $x_0 < x_1 < \ldots < x_k$. Moreover, φ is said to be *strictly convex of order k* on D, if inequality (2.6) is strict.

It is shown in the following example that k-convexity corresponds to well-known concepts for small values of k.

Example 2.4. a) For $k = 0$, the matrix from Definition 2.3a) reduces to $\Xi^0_\varphi(x_0) = \varphi(x_0)$. It follows that the 0-convexity of $\varphi : D \to \mathbb{R}$ corresponds to $\varphi(x_0) \geq 0$ for all $x_0 \in D$. Hence, a function is 0-convex / strictly 0-convex, if and only if it is non-negative / positive.

b) For $k = 1$, we obtain

$$\Xi^1_\varphi(x_0, x_1) = \begin{pmatrix} 1 & 1 \\ \varphi(x_0) & \varphi(x_1) \end{pmatrix}.$$

Thus, the 1-convexity of φ is equivalent to $\varphi(x_1) - \varphi(x_0) \geq 0$ for all $x_0, x_1 \in D$ with $x_0 < x_1$, so it is equivalent to φ being increasing.

c) For $k = 2$, Definition 2.3a) yields

$$\Xi^2_\varphi(x_0, x_1, x_2) = \begin{pmatrix} 1 & 1 & 1 \\ x_0 & x_1 & x_2 \\ \varphi(x_0) & \varphi(x_1) & \varphi(x_2) \end{pmatrix}.$$

Hence, the 2-convexity of φ is equivalent to

$$\begin{aligned} \det(\Xi^2_\varphi(x_0, x_1, x_2)) &= x_1\varphi(x_2) + x_2\varphi(x_0) + x_0\varphi(x_1) - x_1\varphi(x_0) - x_2\varphi(x_1) - x_0\varphi(x_2) \\ &= \varphi(x_0)(x_2 - x_1) + \varphi(x_1)(x_0 - x_2) + \varphi(x_2)(x_1 - x_0) \\ &\geq 0 \end{aligned}$$

for all $x_0, x_1, x_2 \in D$ with $x_0 < x_1 < x_2$. That inequality can be transformed to

$$\varphi(x_1) \leq \frac{x_2 - x_1}{x_2 - x_0}\varphi(x_0) + \frac{x_1 - x_0}{x_2 - x_0}\varphi(x_2).$$

The substitution $\lambda = \frac{x_2 - x_1}{x_2 - x_0} \in (0, 1)$ yields

$$\varphi(\lambda x_0 + (1 - \lambda)x_2) \leq \lambda\varphi(x_0) + (1 - \lambda)\varphi(x_2).$$

It follows that 2-convexity is equivalent to convexity in the usual sense.

d) For $k \geq 3$, we have

$$\Xi_{\varphi}^{k}(x_0, x_1, \ldots, x_k) = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ x_0 & x_1 & x_2 & \cdots & x_{k-1} & x_k \\ x_0^2 & x_1^2 & x_2^2 & \cdots & x_{k-1}^2 & x_k^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_0^{k-1} & x_1^{k-1} & x_2^{k-1} & \cdots & x_{k-1}^{k-1} & x_k^{k-1} \\ \varphi(x_0) & \varphi(x_1) & \varphi(x_2) & \cdots & \varphi(x_{k-1}) & \varphi(x_k) \end{pmatrix}.$$

The definition of higher-order convexity is rather complex because of the corresponding determinant. Hence, it is usually preferable to utilize other characterizations or implications of k-convexity. The concept of 3-convexity is explored further in Chapter 4.

If the function φ is smooth enough, its k-convexity is equivalently characterized via its k-th derivative (see Karlin, 1968, p. 281 or Popoviciu, 1933, p. 41). We denote the first, second and third derivative of a univariate function φ by φ', φ'' and φ'''; the k-th derivative is denoted by $\varphi^{(k)}$, $k \in \mathbb{N}_0$.

Proposition 2.5. *Let $D \subseteq \mathbb{R}$ be an open interval and $\varphi : D \to \mathbb{R}$ be k times differentiable. Then, the k-convexity of φ is equivalent to $\varphi^{(k)}(x) \geq 0$ for all $x \in D$.*

Especially for larger values of k, Proposition 2.5 is not only helpful for working with k-convex functions, but also for visualizing what k-convexity means. The same is true for the intersection characterization, which is noted in the following. We start with a slightly modified definition by Karlin (1968, p. 20). Preliminarily, we define the sign function as

$$\operatorname{sgn} : \mathbb{R} \to \{-1, 0, 1\}, \quad x \mapsto \mathbb{1}_{(0,\infty)}(x) - \mathbb{1}_{(-\infty,0)}(x) = \begin{cases} -1 & \text{, if } x < 0, \\ 0 & \text{, if } x = 0, \\ 1 & \text{, if } x > 0. \end{cases}$$

Definition 2.6. a) For $n \in \mathbb{N}$, the function

$$S_n^- : \mathbb{R}^n \to \{0, \ldots, n-1\}, \quad (x_1, \ldots, x_n)^\top \mapsto |\{i \in \{1, \ldots, n-1\} : \operatorname{sgn}(x_i \cdot x_{i+1}) = -1\}|$$

counts the number of direct sign changes in a given n-dimensional vector.

b) Let $D \subseteq \mathbb{R}$ be an interval and let $\mathrm{M}(D, \mathbb{R})$ denote the set of all mappings with domain D and codomain $\mathbb{R}$. Then, the function

$$S^- : \mathrm{M}(D, \mathbb{R}) \to \mathbb{N}_0, \quad \varphi \mapsto \sup\{S_n^-(\varphi(x_1), \ldots, \varphi(x_n)) : n \in \mathbb{N}, (x_1, \ldots, x_n)^\top \in D^n\}$$

counts the number of sign changes of a function.

For the following result, we use the convention that the polynomial that is constantly equal to zero has degree $-\infty$.

Theorem 2.7. *Let $k \in \mathbb{N}_0$, let $D \subseteq \mathbb{R}$ be an open interval and $\varphi : D \to \mathbb{R}$ be a continuous function. Then, φ is k-convex, if and only if φ is a polynomial of degree $\leq k-1$ or both of the following conditions are satisfied:*

(i) $S^-(\varphi - \pi) \leq k$ for all polynomials π of degree $\leq k-1$.

(ii) For all polynomials π of degree $\leq k-1$ with $S^-(\varphi - \pi) = k$, the first non-zero value of $\varphi - \pi$ has sign $(-1)^k$.

Proof. First, note that the set of the zeroth to the $(k-1)$-th monomial, $\{m_0, \ldots, m_{k-1}\}$, constitutes a so-called Tchebycheff system or T-system (see Karlin, 1968, pp. 24–26). With that in mind, Theorem 2.1 on pages 281 and 282 in the same reference gives the result. ▨

The concept of k-convex functions is now applied to modified RIDF's. Therefore, let $F, G \in \mathcal{P}$ and let Δ_{FG} be defined as in Section 2.1. Using the k-convexity of that function, the notion of the k-th convex characteristic of a distributions shall be introduced as rigorously as possible. This is done based on a fundamental ordering with respect to that characteristic. As mentioned in Section 2.1, $\Delta_{FG}(X)$ is the difference of an F-distributed random variable and a G-distributed random variable, and is therefore well suited to compare these two distributions. This is the basis of the approach by Oja (1981), who uses the k-convexity of Δ_{FG} as a foundation for comparing distributions with respect to the k-th convex characteristic.

However, the distributional result concerning $\Delta_{FG}(X)$ is only true under the assumption $G(D_G) \subseteq \overline{F(D_F)}$, as proved in Proposition A.3. Thus, a sufficient condition is given by $F, G \in \mathcal{C}$. In order to establish the convex characteristics as general as possible, we do not require Δ_{FG} to be k-convex for comparison with respect to the k-th convex characteristic, but instead give the following definition.

Definition 2.8. Let $k \in \mathbb{N}_0$ and $F, G \in \mathcal{P}$.

a) For $p_0, p_1, \ldots, p_k \in (0,1)$, define the $(k+1) \times (k+1)$-matrix $\tilde{\Xi}^k_{F^{-1},G^{-1}}(p_0, p_1, \ldots, p_k) = \left(\tilde{\xi}_{i,j}\right)_{i,j=1,\ldots,k}$ by

$$\begin{aligned} \tilde{\xi}_{k+1,j} &= G^{-1}(p_{j-1}) - F^{-1}(p_{j-1}), \quad j = 1, \ldots, k+1, \\ \tilde{\xi}_{i,j} &= \left(F^{-1}(p_{j-1})\right)^{i-1}, \quad j = 1, \ldots, k+1,\ i = 1, \ldots, k. \end{aligned}$$

b) We say that F *precedes G in the order of the k-th convex characteristic*, denoted by $F \leq_k G$, if

$$\det\left(\tilde{\Xi}^k_{F^{-1},G^{-1}}(p_0, p_1, \ldots, p_k)\right) \geq 0 \tag{2.7}$$

holds for all $0 < p_0 < p_1 < \ldots < p_k < 1$.

This definition does not use the function Δ_{FG} and, therefore, is not dependent upon the necessary and sufficient condition $G(D_G) \subseteq \overline{F(D_F)}$ for its crucial property. Instead, the two quantile functions of F and G are compared. This is generally more complex to deal with because the interplay of two functions needs to be considered instead of just one function. However, since $\leq_k$ just compares two quantile functions in a pointwise way, no assumptions are necessary. The following result connects the two approaches.

Proposition 2.9. *Let $k \in \mathbb{N}_0$, $F \in \mathcal{P}_I$ and $G \in \mathcal{P}$. Then, $F \leq_k G$ holds, if and only if Δ_{FG} is k-convex on D_F.*

Proof. Note that D_F is an open interval because of $F \in \mathcal{P}_I$. We proceed to compare the two $(k \times k)$-matrices $\Xi^k_{\Delta_{FG}}(x_0, x_1, \ldots, x_k) = (\xi_{i,j})_{i,j=1,\ldots,k+1}$ and $\tilde{\Xi}^k_{F^{-1},G^{-1}}(p_0, p_1, \ldots, p_k) = \left(\tilde{\xi}_{i,j}\right)_{i,j=1,\ldots,k}$.

For each $x_\ell \in D_F, \ell = 0, \ldots, k$, there exists a unique $p_\ell \in (0,1)$ such that $x_\ell = F^{-1}(p_\ell)$. Hence, for $j = 1, \ldots, k+1,\ i = 1, \ldots, k$, we have

$$\begin{aligned}
\xi_{k+1,j} &= \Delta_{FG}(x_{j-1}) = \Delta_{FG}(F^{-1}(p_{j-1})) = G^{-1}(F(F^{-1}(p_{j-1}))) - F^{-1}(p_{j-1}) \\
&= G^{-1}(p_{j-1}) - F^{-1}(p_{j-1}) = \tilde{\xi}_{k+1,j}, \\
\xi_{i,j} &= x_{j-1}^{i-1} = \left(F^{-1}(p_{j-1})\right)^{i-1} = \tilde{\xi}_{i,j}.
\end{aligned}$$

Note that $F \circ F^{-1} = \mathrm{id} = F^{-1} \circ F$, where id denotes the identity function, follows from $F \in \mathcal{P}_I$, since both F and F^{-1} are strictly increasing. It follows that, for each choice $x_0, \ldots, x_k \in D_F$ with $x_0 < \ldots < x_k$, there exist $0 < p_0 < \ldots < p_k < 1$ such that $\Xi^k_{\Delta_{FG}}(x_0, x_1, \ldots, x_k) = \tilde{\Xi}^k_{F^{-1},G^{-1}}(p_0, p_1, \ldots, p_k)$. This proves that $F \leq_k G$ implies the k-convexity of Δ_{FG} on D_F.

Conversely, for each $p_\ell \in (0,1), \ell = 0, \ldots, k,\ F \in \mathcal{P}_I$ yields the existence of a unique $x_\ell \in D_F$ such that $p_\ell = F(x_\ell)$. Hence, for $j = 1, \ldots, k+1,\ i = 1, \ldots, k$, we have

$$\begin{aligned}
\tilde{\xi}_{k+1,j} &= G^{-1}(p_{j-1}) - F^{-1}(p_{j-1}) = G^{-1}(F(x_{j-1})) - F^{-1}(F(x_{j-1})) \\
&= G^{-1}(F(x_{j-1})) - x_{j-1} = \Delta_{FG}(x_{j-1}) = \xi_{k+1,j}, \\
\tilde{\xi}_{i,j} &= \left(F^{-1}(p_{j-1})\right)^{i-1} = \left(F^{-1}(F(x_{j-1}))\right)^{i-1} = x_{j-1}^{i-1} = \xi_{i,j}.
\end{aligned}$$

It follows that, for each choice $0 < p_0 < \ldots < p_k < 1$, there exist $x_0, \ldots, x_k \in D_F$ with $x_0 < \ldots < x_k$ such that $\Xi^k_{\Delta_{FG}}(x_0, x_1, \ldots, x_k) = \tilde{\Xi}^k_{F^{-1},G^{-1}}(p_0, p_1, \ldots, p_k)$. This proves that the k-convexity of Δ_{FG} on D_F implies $F \leq_k G$. ▨

Hürlimann (2002, p. 10) also considered the family of orders $\leq_k, k \in \mathbb{N}_0$, calling the k-th representative the 'degree k relative inverse convex order'. Like Oja (1981), he restricted attention to $\mathcal{P}_I$ and defined the orders using the characterization in Proposition 2.9. As opposed to that characterization, Definition 2.8b) matches the usual definitions in the literature for special cases exactly. This is shown in the following example.

Example 2.10. a) For $k = 0$, the matrix from Definition 2.8a) reduces to $\tilde{\Xi}^0_{F^{-1},G^{-1}}(p_0) = G^{-1}(p_0) - F^{-1}(p_0)$. Hence, $F \leq_0 G$ holds, if and only if

$$F^{-1}(p_0) \leq G^{-1}(p_0) \quad \forall p_0 \in (0,1). \tag{2.8}$$

Since this is an equivalent characterization of the usual stochastic order (see Belzunce et al., 2015, p. 30), it follows that $F \leq_0 G$ is equivalent to $F \leq_{st} G$. Müller and Stoyan (2002, p. 4) note that $F \leq_{st} G$ is generally not equivalent to the 0-convexity of Δ_{FG}, which is the same as the non-negativity of Δ_{FG} (see Example 2.4a)). This issue is explored further in Section 6.1 and specifically in Proposition 6.1.

b) For $k = 1$, we have

$$\tilde{\Xi}^1_{F^{-1},G^{-1}}(p_0, p_1) = \begin{pmatrix} 1 & 1 \\ G^{-1}(p_0) - F^{-1}(p_0) & G^{-1}(p_1) - F^{-1}(p_1) \end{pmatrix},$$

yielding that $F \leq_1 G$ is equivalent to

$$F^{-1}(p_1) - F^{-1}(p_0) \leq G^{-1}(p_1) - G^{-1}(p_0) \quad \forall\, 0 < p_0 < p_1 < 1. \tag{2.9}$$

This matches the definition of the dispersive order (see Müller and Stoyan, 2002, p. 40), meaning that $F \leq_1 G$ is equivalent to $F \leq_{disp} G$. The fact that this is not equivalent to Δ_{FG} being increasing is proved with the use of counterexamples in Example 6.5.

c) The matrix $\tilde{\Xi}^2_{F^{-1},G^{-1}}(p_0, p_1, p_2)$ is given by

$$\begin{pmatrix} 1 & 1 & 1 \\ F^{-1}(p_0) & F^{-1}(p_1) & F^{-1}(p_2) \\ G^{-1}(p_0) - F^{-1}(p_0) & G^{-1}(p_1) - F^{-1}(p_1) & G^{-1}(p_2) - F^{-1}(p_2) \end{pmatrix}.$$

It follows that $F \leq_2 G$ is equivalent to

$$\begin{aligned}
&\det\left(\tilde{\Xi}^2_{F^{-1},G^{-1}}(p_0, p_1, p_2)\right) \\
&= F^{-1}(p_0)(G^{-1}(p_1) - F^{-1}(p_1)) - F^{-1}(p_0)(G^{-1}(p_2) - F^{-1}(p_2)) \\
&\quad + F^{-1}(p_1)(G^{-1}(p_2) - F^{-1}(p_2)) - F^{-1}(p_1)(G^{-1}(p_0) - F^{-1}(p_0)) \\
&\quad + F^{-1}(p_2)(G^{-1}(p_0) - F^{-1}(p_0)) - F^{-1}(p_2)(G^{-1}(p_1) - F^{-1}(p_1)) \\
&= G^{-1}(p_0)(F^{-1}(p_2) - F^{-1}(p_1)) + G^{-1}(p_1)(F^{-1}(p_0) - F^{-1}(p_2)) \\
&\quad + G^{-1}(p_2)(F^{-1}(p_1) - F^{-1}(p_0)) \\
&= (G^{-1}(p_2) - G^{-1}(p_1))(F^{-1}(p_1) - F^{-1}(p_0)) \\
&\quad - (G^{-1}(p_1) - G^{-1}(p_0))(F^{-1}(p_2) - F^{-1}(p_1)) \geq 0
\end{aligned} \tag{2.10}$$

for all $0 < p_0 < p_1 < p_2 < 1$. This, in turn, is equivalent to

$$\frac{F^{-1}(p_2) - F^{-1}(p_1)}{F^{-1}(p_1) - F^{-1}(p_0)} \leq \frac{G^{-1}(p_2) - G^{-1}(p_1)}{G^{-1}(p_1) - G^{-1}(p_0)} \tag{2.11}$$

for all $0 < p_0 < p_1 < p_2 < 1$ with $F^{-1}(p_0) < F^{-1}(p_2)$ and $G^{-1}(p_0) < G^{-1}(p_2)$. Here, we allow division by zero and assign the value ∞ in that case.

Note that if $G^{-1}(p_0) = G^{-1}(p_1)$ holds, both (2.10) and (2.11) are trivially satisfied; if $F^{-1}(p_0) = F^{-1}(p_1)$ holds, each inequality is satisfied if and only if $G^{-1}(p_0) = G^{-1}(p_1)$. If $F^{-1}(p_0) = F^{-1}(p_2)$ or $G^{-1}(p_0) = G^{-1}(p_2)$ holds, (2.10) holds as both sides of the inequality are zero.

The commonly used fundamental skewness order in the literature is the convex transformation order by van Zwet (1964, p. 48), denoted by $\leq_c$. It is usually only considered for $F, G \in \mathcal{P}_I$ or under even stronger assumptions and $F \leq_c G$ is defined by the convexity of Δ_{FG} on D_F. Under this assumption, the definition by van Zwet is equivalent to $F \leq_2 G$, according to Proposition 2.9.

It is shown in Lemma 3.9 that $F \leq_2 G$ can also be characterized via a uniform ordering of second differences of F^{-1} and G^{-1}.

Due to the observations in Example 2.10, we formally define the three well-known orders in the following way.

Definition 2.11. Let $F, G \in \mathcal{P}$. The *usual stochastic order* $\leq_{st}$, *the dispersive order* $\leq_{disp}$ and *the convex transformation order* $\leq_c$ are defined as the order of the zeroth convex characteristic $\leq_0$, first convex characteristic $\leq_1$ and second convex characteristic $\leq_2$, respectively.

Note that the usual stochastic order is usually defined differently: $F \leq_{st} G$, if $G(t) \leq F(t)$ for all $t \in \mathbb{R}$. However, this is easily seen to be equivalent to (2.8) (see, e.g., Belzunce et al., 2015, p. 30).

For the order of the k-th convex characteristic, $k \in \mathbb{N}_0$, Propositions 2.5 and 2.9 yield the following corollary.

Corollary 2.12. *Let $k \in \mathbb{N}_0$ and $F, G \in \mathcal{P}_I^k$. Then, $F \leq_k G$ holds, if and only if $\Delta_{FG}^{(k)}(x) \geq 0$ for all $x \in D_F$. If $k \geq 2$, $F \leq_k G$ is also equivalent to $R_{FG}^{(k)}(x) \geq 0$ for all $x \in D_F$.*

Proof. Since $G \in \mathcal{P}_I^k$ implies that G is continuous and strictly increasing on D_G, its quantile function G^{-1} coincides with its inverse function. Furthermore, G^{-1} inherits from G that it is k times differentiable on $(0, 1)$. As a composition of k times differentiable functions, Δ_{FG} is also k times differentiable. The first statement then follows from Propositions 2.5 and 2.9.

For $k \geq 2$, we have

$$\Delta_{FG}^{(k)}(x) = R_{FG}^{(k)}(x) - \mathrm{id}^{(k)}(x) = R_{FG}^{(k)}(x)$$

for all $x \in D_F$. ∎

Note that the strict version of the order $\leq_k$ is not defined via the strict version of inequality (2.7). Instead, it is defined via the general construction principle given in (2.4). These two definitions, in general, do not yield the same strict order. This is analyzed in more detail for the order of the third convex characteristic at the end of Section 4.2.1.

Another characterization of the order of the k-th convex characteristic is obtained by employing divided differences, a concept that is related to generalized convexity (see, e.g., Mühlbach, 1973). The recursive definition of divided differences is given in the following (cp. Nørlund, 1926, p. 1).

Definition 2.13. Let $\varphi : D \to \mathbb{R}$ be a function with $D \subseteq \mathbb{R}, |D| \geq k+1$, and let $x_0, x_1, \ldots, x_k \in D$ with $x_0 < x_1 < \ldots < x_k$ for some $k \in \mathbb{N}_0$. Then, the *zeroth and k-th divided difference, respectively, of φ at $x_0, \ldots, x_k$* is defined by

$$\begin{aligned}[x_0|\varphi] &= \varphi(x_0),\\ [x_0, \ldots, x_k|\varphi] &= \frac{[x_1, \ldots, x_k|\varphi] - [x_0, \ldots, x_{k-1}|\varphi]}{x_k - x_0}.\end{aligned}$$

The following result connects divided differences to the concept of k-convexity.

Proposition 2.14. *Let $k \in \mathbb{N}_0$, let $\varphi : D \to \mathbb{R}$ be a function with $D \subseteq \mathbb{R}$ and let $F \in \mathcal{P}_I$, $G \in \mathcal{P}$.*

a) For any choice of $x_0, \ldots, x_k \in D$ with $x_0 < \ldots < x_k$, the identity

$$[x_0, \ldots, x_k|\varphi] = \frac{\det\left(\Xi^k_{\varphi}(x_0, \ldots, x_k)\right)}{\det\left(\Xi^k_{m_k}(x_0, \ldots, x_k)\right)},$$

holds, where $m_k : D \to \mathbb{R}, x \mapsto x^k$ denotes the k-th monomial.

b) φ is k-convex, if and only if $[x_0, \ldots, x_k|\varphi] \geq 0$ for all $x_0, \ldots, x_k \in D$ with $x_0 < \ldots < x_k$.

c) The following statements are all equivalent:

(i) $F \leq_k G$,

(ii) $[x_0, \ldots, x_k|\Delta_{FG}] \geq 0$ for all $x_0, \ldots, x_k \in D_F$ with $x_0 < \ldots < x_k$,

(iii) $[F^{-1}(p_0), \ldots, F^{-1}(p_k)|\Delta_{FG}] \geq 0$ for all $0 < p_0 < \ldots < p_k < 1$.

Proof. a) See Mühlbach (1973, p. 165) or Nørlund (1926, p. 2). Note that the ratio of determinants is well defined because the denominator, which is known as the Vandermonde determinant, is given by

$$\det\left(\Xi^k_{m_k}(x_0, \ldots, x_k)\right) = \prod_{0 \leq i < j \leq k} (x_j - x_i) > 0$$

for all $x_0, \ldots, x_k \in D$ with $x_0 < \ldots < x_k$ (see, e.g., Horn and Johnson, 2013, p. 37).

b) Follows directly from part a) and Proposition 2.9.

c) The equivalence of (i) and (ii) follows from Proposition 2.9 and part b). The equivalence of (ii) and (iii) follows from the fact that the function F^{-1} is a bijective and strictly increasing mapping from $(0,1)$ to D_F.

$\blacksquare$

In the following, the characterization from Proposition 2.14c) is explored for the order of the second convex characteristics. Here, the divided quantile differences are calculated via the recursion formula. For the order of the third convex characteristic, it is explored in Section 4.3.2.

Example 2.15. Let $k = 2$, $F \in \mathcal{P}_I$, $G \in \mathcal{P}$ and $0 < p_0 < p_1 < p_2 < 1$. By definition, the corresponding divided difference of Δ_{FG} at $F^{-1}(p_0), F^{-1}(p_1), F^{-1}(p_2)$ is given by

$$\begin{aligned}
[F^{-1}(p_0), F^{-1}(p_1), F^{-1}(p_2)|\Delta_{FG}] &= \frac{[F^{-1}(p_1), F^{-1}(p_2)|\Delta_{FG}] - [F^{-1}(p_0), F^{-1}(p_1)|\Delta_{FG}]}{F^{-1}(p_2) - F^{-1}(p_0)} \\
&= \frac{\frac{G^{-1}(p_2)-G^{-1}(p_1)}{F^{-1}(p_2)-F^{-1}(p_1)} - \frac{G^{-1}(p_1)-G^{-1}(p_0)}{F^{-1}(p_1)-F^{-1}(p_0)}}{F^{-1}(p_2) - F^{-1}(p_0)}
\end{aligned}$$

Overall, Corollary 2.14c)(iii) yields the same characterization of $F \leq_2 G$ as before, namely (2.11).

The intersection characterization for k-convex functions, given in Theorem 2.7, also yields a necessary condition for the order of the k-th convex characteristic.

Corollary 2.16. *Let $k \in \mathbb{N}_0$ and $F, G \in \mathcal{P}_I$. $F \leq_k G$ holds, if and only if there exists a polynomial π_0 of degree $\leq k-1$ such that $F = G \circ (\mathrm{id} + \pi_0)$, or both of the following conditions are satisfied:*

(i) $S^-(F - (G \circ (\mathrm{id} + \pi))) \leq k$ for all polynomials π of degree $\leq k-1$.

(ii) For all polynomials π of degree $\leq k-1$ with $S^-(F - (G \circ (\mathrm{id} + \pi))) = k$, the first non-zero value of $F - (G \circ (\mathrm{id} + \pi))$ has sign $(-1)^k$.

Proof. First, note that $G \circ G^{-1} = \mathrm{id}$ holds because of $G \in \mathcal{P}_I$. The statement of the corollary follows directly from Theorem 2.7 and Proposition 2.9, if $F - (G \circ (\mathrm{id} + \pi))$ is substituted by $\Delta_{FG} - \pi$. Since only the signs of the function $\Delta_{FG} - \pi$ are relevant for that statement, the equivalences

$$\begin{aligned}
\Delta_{FG}(x) - \pi(x) \gtreqless 0 &\Leftrightarrow G^{-1}(F(x)) - x - \pi(x) \gtreqless 0 \\
&\Leftrightarrow G^{-1}(F(x)) \gtreqless x + \pi(x) \\
&\Leftrightarrow F(x) = G(G^{-1}(F(x))) \gtreqless G(x + \pi(x)) = (G \circ (\mathrm{id} + \pi))(x)
\end{aligned}$$

for all $x \in D_F$ justify that substitution. ▨

Another result concerning the number of sign changes of the difference of two cdf's was proved by MacGillivray (1985) and is given in the first part of the following proposition. It connects the number of sign changes to the moments of the distributions. The second part of the proposition is a direct implication.

Proposition 2.17. *Let $k \in \mathbb{N}_0$ and let the first k moments of $F, G \in \mathcal{P}_I, F \neq G$, be finite.*

a) If $\mathbb{E}[X^j] = \mathbb{E}[Y^j]$ holds for all $j \in \{1, \ldots, k\}$, then the following two statements are true:

(i) $S^-(F - G) \geq k$.

(ii) If $S^-(F-G) = k$ and the last non-zero value of $F-G$ is positive, then $\mathbb{E}[X^{k+2\ell-1}] < \mathbb{E}[Y^{k+2\ell-1}]$ holds for all $\ell \in \mathbb{N}$ such that the moments exist.

b) If $F \leq_k G$ and $\mathbb{E}[X^j] = \mathbb{E}[Y^j]$ holds for all $j \in \{1, \ldots, k\}$, then $S^-(F - G) = k$ follows, the last non-zero value of $F - G$ is positive and $\mathbb{E}[X^{k+2\ell-1}] < \mathbb{E}[Y^{k+2\ell-1}]$ holds for all $\ell \in \mathbb{N}$ such that the moments exist.

Proof. a) See MacGillivray (1985, p. 413).

b) The statement $S^-(F - G) = k$ follows directly from part a)(i) and Corollary 2.16(i) for $\pi \equiv 0$. Furthermore, it follows from Corollary 2.16(ii) that the first non-zero value of $F - G$ has sign $(-1)^k$. Since $F - G$ has exactly k sign changes, the last non-zero value of $F - G$ has sign $(-1)^{2k} = +1$. Part a)(ii) then yields the asserted inequality of the higher moments.

▨

Another family of stochastic orders that is based upon the concept of k-convex functions are the k-convex orders. The following definition and equivalent characterization can be found in Müller and Stoyan (2002, p. 39). For $x \in \mathbb{R}$, let $x_+ = \max\{0, x\}$ and $x_- = \max\{0, -x\}$ denote its positive part and negative part, respectively.

Definition 2.18. Let $k \in \mathbb{N}$ and $F, G \in \mathcal{P}$. Then, F is said to precede G in the *k-convex order*, denoted by $F \leq_{k-cx} G$, if $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(Y)]$ holds for all k-convex functions φ for which the expectations exist.

Proposition 2.19. *Let $k \in \mathbb{N}_{\geq 2}$ and $F, G \in \mathcal{L}^{k-1}$. Then, $F \leq_{k-cx} G$ holds, if and only if the following two conditions are satisfied:*

(i) $\mathbb{E}[X^j] = \mathbb{E}[Y^j]$ for all $j \in \{1, \ldots, k - 1\}$,

(ii) $\mathbb{E}[(X - t)_+^{k-1}] \leq \mathbb{E}[(Y - t)_+^{k-1}]$ for all $t \in \mathbb{R}$.

Through the assumption of equal moments, the k-convex order is connected to the order of the k-th convex characteristic by the following result. The first part can be found in Klar (2002, p. 13) or Rolski (1976, p. 16), the second part follows directly with Proposition 2.17b).

Proposition 2.20. *Let $k \in \mathbb{N}_0$ and let the first k moments of $F, G \in \mathcal{P}_I, F \neq G$, be finite and satisfy $\mathbb{E}[X^j] = \mathbb{E}[Y^j]$ for all $j \in \{1, \ldots, k\}$.*

a) If $S^-(F-G) = k$ holds and the last non-zero value of $F-G$ is positive, then $F \leq_{(k+1)-cx} G$ follows.

b) If $F \leq_k G$, then $F \leq_{(k+1)-cx} G$ follows.

For $k \in \mathbb{N}_0$, if the first k moments of two distributions are equal, the $(k+1)$-convex order is a weakening of the order of the k-th convex characteristic. Hence, for $k \in \mathbb{N}$, the k-convex order compares two distributions with respect to $(k-1)$-th convex characteristic, if the moment conditions are fulfilled.

Example 2.21. a) Let $F, G \in \mathcal{P}$. $F \leq_{1-cx} G$ holds, if and only if $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(Y)]$ is satisfied for all increasing functions φ for which the expectations exist. It can be shown that this is equivalent to $F \leq_{st} G$ (see, e.g., Belzunce et al., 2015, pp. 31–32). Hence, the statement of Proposition 2.20b) can be improved in the case $k = 0$, as $F \leq_0 G$ and $F \leq_{1-cx} G$ are equivalent. Overall, this confirms that the 1-convex order indeed compares with respect to the zeroth convex characteristic, which means that it is a order of location. The characterization from Proposition 2.19 is not applicable for the 1-convex order.

b) Since 2-convexity coincides with the usual notion of convexity, the 2-convex order $\leq_{2-cx}$ is also simply called the convex order and denoted by $\leq_{cx}$. We proceed to consider the characterization of $F \leq_{cx} G$ given by Proposition 2.19. In order to circumvent dealing with the requirement of equal expectation, it often makes sense to instead consider the random variables that are standardized with respect to their expectations. The order arising from this is called the *dilation order* and is denoted by $\leq_{dil}$. For $F, G \in \mathcal{L}^1$, $F \leq_{dil} G$ is said to hold, if

$$\pi_X(t + \mathbb{E}[X]) = \mathbb{E}[(X - \mathbb{E}[X] - t)_+] \leq \mathbb{E}[(Y - \mathbb{E}[Y] - t)_+] = \pi_Y(t + \mathbb{E}[Y]) \tag{2.12}$$

is satisfied for all $t \in \mathbb{R}$, where $\pi_Z : \mathbb{R} \to [0, \infty), t \mapsto \mathbb{E}[(Z - t)_+]$ denotes the so-called *stop-loss transform* of a random variable Z. Note that $\leq_{dil}$ and $\leq_{cx}$ are not the same order since $\mathbb{E}[X] = \mathbb{E}[Y]$ is a necessary condition for $F \leq_{cx} G$ but not for $F \leq_{dil} G$. Both the convex order and the dilation order are widely used in financial and actuarial mathematics as orders of dispersion.

c) The situation for the 3-convex order is similar to that for the 2-convex order, although the 3-convex order is not as widely used. One can define an equivalent to the dilation order, which we call the 3-dilation order, by standardizing the random variables not only with respect to the expectation but also with respect to the standard deviation, which corresponds to the second moment. Hence, for $F, G \in \mathcal{L}^2$, $F \leq_{3-dil} G$ is said to hold, if

$$\mathbb{E}\left[\left(\frac{X-\mathbb{E}[X]}{\sqrt{\mathbb{V}(X)}}-t\right)_+^2\right] \leq \mathbb{E}\left[\left(\frac{Y-\mathbb{E}[Y]}{\sqrt{\mathbb{V}(Y)}}-t\right)_+^2\right]$$

is satisfied for all $t \in \mathbb{R}$. The usage of this order in queuing theory is summarized by Müller and Stoyan (2002, pp. 37–40), see also Rolski (1976). The 3-dilation order is seldom explicitly considered as a order of skewness although it can be regarded as such.

d) For k-convex orders with $k \geq 4$, the corresponding k-dilation order cannot be defined in the same way as for $k = 2$ and $k = 3$. This is due to the fact that a random variable cannot be standardized with respect to its third moment, as opposed to the first two. This observation makes the characterization in Proposition 2.19 as well as the corresponding k-convex order much less useful. Therefore, these orders are not explored any further here.

2.3. Generalized Quantiles and Expectiles

The most widely used definition of quantiles, which we mostly use in this thesis, is given in (2.2). Furthermore, (2.3) gives a more general characterization of quantiles that is not unique for all $p \in (0, 1)$, if the corresponding cdf F is not strictly increasing on D_F. The latter characterization also arises, if quantiles are not defined as in (2.2), but as the solution of a minimization problem. For $p \in (0, 1)$ and any integrable random variable $X \sim F \in \mathcal{P}$,

$$Q_p^F = \operatorname*{argmin}_{t \in \mathbb{R}} \{p\mathbb{E}[(X-t)_+] + (1-p)\mathbb{E}[(X-t)_-]\}$$

coincides with the characterization in (2.3) (see Ferguson, 1967, p. 51 with solution, or Koenker, 2005, pp. 5–6). This concept can be generalized by applying arbitrary convex loss functions to the positive and the negative part of $(X - t)$. If $\phi_1, \phi_2 : [0, \infty) \to [0, \infty)$ are convex and strictly increasing functions with $\phi_1(0) = \phi_2(0) = 0$ and $\phi_1(1) = \phi_2(1) = 1$, any thereby obtained minimizer

$$\operatorname*{argmin}_{t \in \mathbb{R}} \{p\mathbb{E}[\phi_1((X-t)_+)] + (1-p)\mathbb{E}[\phi_2((X-t)_-)]\}$$

is called a *generalized p-quantile*, $p \in (0, 1)$, following Bellini et al. (2014, p. 42). They also proved that the generalized p-quantile is unique for all choices of $p \in (0, 1)$, if ϕ_1 and ϕ_2 are strictly convex.

If $\phi_1(x) = \phi_2(x) = x^2, x \in [0, \infty)$, the generalized p-quantile is called the *p-expectile*, which was first discussed by Newey and Powell (1987). This specific kind of generalized quantile has a number of desirable properties, e.g., it arises from the only choice of ϕ_1 and ϕ_2 for which a coherent risk measure is obtained (see Bellini et al., 2014, p. 44). Furthermore, the strict convexity of ϕ_1 and ϕ_2 directly implies that all expectiles are unique, which is particularly noteworthy for discrete distributions. For $X \in L^2$ and $p \in (0,1)$, the corresponding p-expectile is defined as

$$e_X(p) = \operatorname*{argmin}_{t\in\mathbb{R}} \left\{\mathbb{E}[\ell_p(X-t)]\right\},$$

where

$$\ell_p(x) = \begin{cases} px^2, & \text{if } x \geq 0, \\ (1-p)x^2, & \text{if } x < 0. \end{cases}$$

As before, we use cdf's and correspondingly distributed random variables interchangeably, i.e. $e_F = e_X$. The assumption can be weakened to $X \in L^1$ by instead defining

$$e_X(p) = \operatorname*{argmin}_{t\in\mathbb{R}} \left\{\mathbb{E}[\ell_p(X-t) - \ell_p(X)]\right\}$$

(see Newey and Powell, 1987, p. 823). The p-expectile is also uniquely characterized by the first order condition

$$p\mathbb{E}[(X - e_X(p))_+] = (1-p)\mathbb{E}[(X - e_X(p))_-]. \tag{2.13}$$

This directly implies the useful property that the $\frac{1}{2}$-expectile is equal to the expected value. So, heuristically, the relationship between a p-expectile and the p-quantile is similar to the much more tangible relationship between the mean and the median. There are a number of further characterizations of expectiles, most of which can be found in Bellini et al. (2014) and Bellini et al. (2018a). Here, we limit ourselves to mentioning that the p-expectile $e_X(p)$, $p \in (0,1)$, is the p-quantile of the suitably transformed cdf

$$\breve{F}(t) = \frac{\mathbb{E}[(X-t)_-]}{\mathbb{E}[|X-t|]}, \quad t \in \mathbb{R} \tag{2.14}$$

(see Jones, 1994, pp. 149–150). Finally, the following proposition states a number of properties of expectiles, collected from Newey and Powell (1987), Bellini et al. (2014) and Bellini et al. (2018a).

Proposition 2.22. *Let $F \in \mathcal{L}^1$ and $p \in (0,1)$. Then:*

a) $e_{X+a}(p) = e_X(p) + a$ *for all* $a \in \mathbb{R}$*,*

b) $e_{\lambda X}(p) = \lambda e_X(p)$ *for all* $\lambda > 0$*,*

c) $e_X(p)$ *is strictly increasing with respect to* p*,* *d)* $e_X(p)$ *is continuous with respect to* p*,*

e) $e_{-X}(p) = -e_X(1-p)$*,*

f) if F *is continuous,* e_X *has the derivative*

$$e'_X(p) = \frac{\mathbb{E}\left[|X - e_X(p)|\right]}{(1-p)F(e_X(p)) + p\left(1 - F(e_X(p))\right)}.$$

PART I

Continuous Setting

Chapter 3

Quantification of Location, Dispersion and Skewness

In the entirety of Part I, we only consider distributions within the set $\mathcal{P}_I$. Generally, let $F, G \in \mathcal{P}_I$. The purpose of this chapter to utilize the fundamental orders of location, dispersion and skewness established in Chapter 2.2 to define measures of these characteristics. This is done using the general approach by Oja (1981). Subsequently, multiple different families of measures are considered. Moment-based and quantile-based measures, the topics of Sections 3.1.1 and 3.1.2, are widely used in the literature as well as in applications. The density-based measures introduced in Section 3.1.3 do not lend themselves particularly well to applications, but they are closely connected to the fundamental orders. Thus, they can be used as a tool to better understand the orders themselves. Measures based on the mode are then mentioned rather briefly in Section 3.1.4, mainly to be referenced in Chapter 4.

Finally, expectile-based orders and measures are considered in Section 3.2. The measures can be treated in a similar way as quantile-based measures, but bring some additional advantages to the table. However, while quantile-based measures are closely related to the underlying orders of the convex characteristics, the foundation for expectile-based measures is more difficult to establish. This is why we also consider a number of expectile-based orders and their relationship with the traditional orders.

We only discuss theoretical properties of the considered measures. For results concerning the asymptotic and empirical properties of most of the skewness measures, we refer to Eberl and Klar (2020, 2022a, pp. 384–393).

3.1. Measuring Location, Dispersion and Skewness

We start out by formally defining measures of location, dispersion and skewness, using the approach by Oja (1981, pp. 157, 159, 163). Recall the convention $X \sim F$, $Y \sim G$.

Definition 3.1. Let $\mathcal{Q} \subseteq \mathcal{P}_I$.

a) A mapping $\nu : \mathcal{Q} \to \mathbb{R}$ is said to be a *measure of central location*, if the two following properties are satisfied:

   (L1) $\nu(aX + b) = a \cdot \nu(X) + b$ for all $a, b \in \mathbb{R}$ and $F \in \mathcal{Q}$.

   (L2) $\nu(F) \leq \nu(G)$ for all $F, G \in \mathcal{Q}$ such that $F \leq_{st} G$.

b) A mapping $\tau : \mathcal{Q} \to \mathbb{R}$ is said to be a *measure of dispersion*, if the two following properties are satisfied:

   (D1) $\tau(aX + b) = |a| \cdot \tau(X)$ for all $a, b \in \mathbb{R}$ and $F \in \mathcal{Q}$.

   (D2) $\tau(F) \leq \tau(G)$ for all $F, G \in \mathcal{Q}$ such that $F \leq_{disp} G$.

c) A mapping $\gamma : \mathcal{Q} \to \mathbb{R}$ is said to be a *measure of skewness*, if the two following properties are satisfied:

   (S1) $\gamma(aX + b) = \mathrm{sgn}(a) \cdot \tau(X)$ for all $a, b \in \mathbb{R}$ and $F \in \mathcal{Q}$.

   (S2) $\gamma(F) \leq \gamma(G)$ for all $F, G \in \mathcal{Q}$ such that $F \leq_c G$.

Note that part a) only defines measures of central location and not measures of location generally. A *measure of non-central location* is defined as in Definition 3.1a), but (L1) is only required to hold for all $a > 0$ instead of for all $a \in \mathbb{R}$. Additionally, one might require the existence of an $F \in \mathcal{Q}$ such that $\nu(-X) \neq -\nu(X)$. This ensures that any measure of central location is not also a measure of non-central location. The most well-known example of a measure of non-central location is the p-quantile for $p \in (0,1) \setminus \{\frac{1}{2}\}$. This can easily be verified using Lemma 3.5d). It can be shown in a similar way that the p-expectile is also a measure of non-central location for $p \in (0,1) \setminus \{\frac{1}{2}\}$. The idea that other characteristics may also be measured in a non-central way is not formally defined, but exemplified in a note following Corollary 3.13.

That the first condition in all three parts of Definition 3.1 dictates the behaviour of a corresponding measure under affine transformations, thereby only covering closely related distributions. Hence, in all three definitions, the crucial property is the second one. Since these properties are based on the three orders $\leq_{st}$, $\leq_{disp}$ and $\leq_c$, we discuss the meaning and some properties of these orders before coming back to the measures. According to Corollary 2.12, the following equivalences hold:

$$
\begin{aligned}
F \leq_{st} G \quad &\Leftrightarrow \quad \Delta_{FG}(t) \geq 0 \;\; \forall t \in D_F \quad \Leftrightarrow \quad R_{FG}(t) \geq t \;\; \forall t \in D_F, \\
F \leq_{disp} G \quad &\Leftrightarrow \quad \Delta'_{FG}(t) \geq 0 \;\; \forall t \in D_F \quad \Leftrightarrow \quad R'_{FG}(t) \geq 1 \;\; \forall t \in D_F, \qquad (3.1) \\
F \leq_c G \quad &\Leftrightarrow \quad \Delta''_{FG}(t) \geq 0 \;\; \forall t \in D_F \quad \Leftrightarrow \quad R''_{FG}(t) \geq 0 \;\; \forall t \in D_F. \qquad (3.2)
\end{aligned}
$$

For the validity of (3.1) and (3.2), the additional assumption $F, G \in \mathcal{P}_I^1$ and $F, G \in \mathcal{P}_I^2$ is needed, respectively.

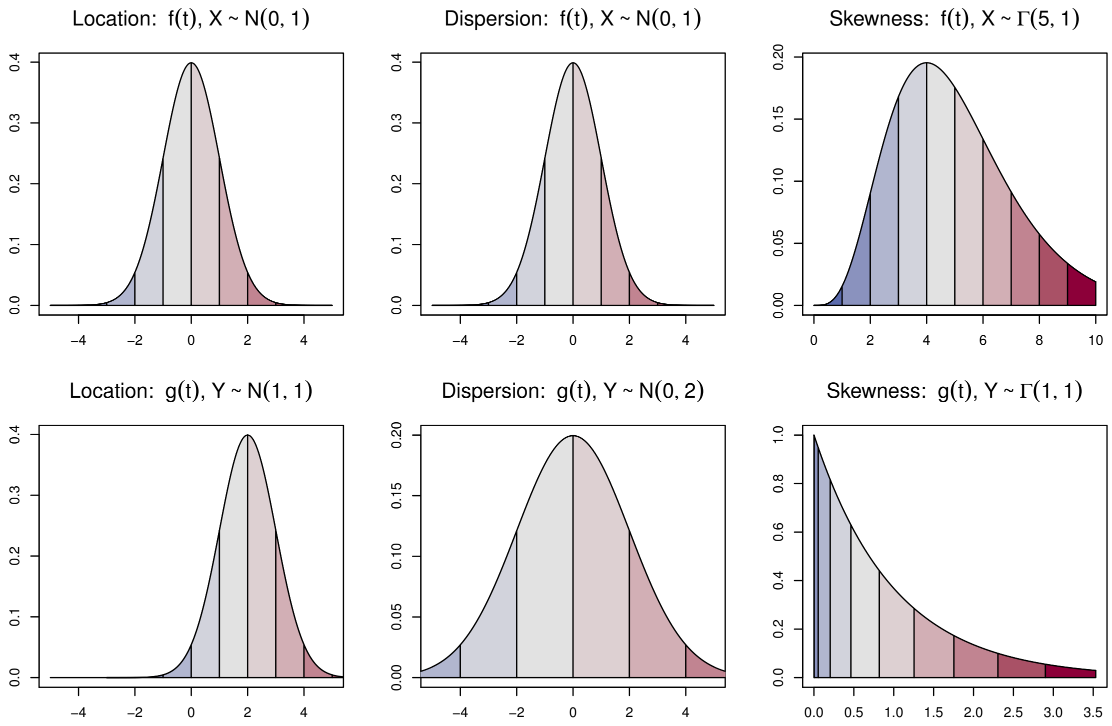

Figure 3.1.: Illustration on how pieces of probability mass are redistributed by R_{FG}, if $F \leq_{st} G$ (order of location), $F \leq_{disp} G$ (order of dispersion) or $F \leq_{c} G$ (order of skewness) holds. Each piece of probability mass is uniquely identified by the colour of the corresponding area.

In order to see that the behaviour of these orders coincides with heuristic ideas of location, dispersion and skewness, we consider how a piece of probability mass of F is transformed by R_{FG} and what requirements the orders $\leq_{st}$, $\leq_{disp}$ and $\leq_{c}$ impose on that transformation. Let the unit interval be partitioned by $0 = p_0 < p_1 < \ldots < p_{n-1} < p_n = 1$ for some sufficiently large $n \in \mathbb{N}$. We consider the intervals $[F^{-1}(p_{i-1}), F^{-1}(p_i)]$, $i = 1, \ldots, n$. On the i-th interval, $100 \cdot (p_i - p_{i-1})\%$ of the probability mass of F is located. If $F \leq_{st} G$ holds, it follows that $G^{-1}(p_i) \geq F^{-1}(p_i)$ for all $i \in \{1, \ldots, n-1\}$. Thus, the transformation R_{FG} shifts all intervals and therefore also all considered pieces of probability mass to the right. This corresponds to the fact that F precedes G with respect to location. The two panels on the left side of Figure 3.1 illustrate this for specific distributions by dividing the probability mass of F into areas of equal width and showing the location and shape of the same pieces for G. In order to visualize probability mass as areas under the curve, the densities of F and G are plotted.

If $F \leq_{disp} G$ holds, $G^{-1}(p_{i-1}) - G^{-1}(p_i) \geq F^{-1}(p_{i-1}) - F^{-1}(p_i)$ follows for all $i \in \{2, \ldots, n-1\}$. Thus, the transformation R_{FG} increases the width of the considered pieces of probability mass. This means that the cdf G is more stretched out, i.e., more dispersed than F. An illustration using the densities of specific distributions is given in the two panels in the middle of Figure 3.1.

Finally, $F \leq_c G$ implies

$$\frac{G^{-1}(p_{i+1}) - G^{-1}(p_i)}{G^{-1}(p_i) - G^{-1}(p_{i-1})} \geq \frac{F^{-1}(p_{i+1}) - F^{-1}(p_i)}{F^{-1}(p_i) - F^{-1}(p_{i-1})}$$

for all $i \in \{2, \ldots, n-2\}$. As for dispersion, R_{FG} increases the length of all considered intervals. This time, however, it does so relative to the length of the interval before instead of extending it in an absolute sense. The larger the values on which the probability mass lies, the more stretched out this probability mass becomes when it is transformed from F to G. This is illustrated in the two panels on the right side of Figure 3.1. Thus, $F \leq_c G$ states that G is more skewed to the right than F, meaning that the right tail of G is longer than the left tail, relative to the corresponding tail ratio in F.

A number of basic properties of the three orders are given in the following results.

Proposition 3.2. *The orders $\leq_{st}$, $\leq_{disp}$ and $\leq_c$ are all reflexive and transitive.*

Proof. The transitivity of is shown separately for each order by Oja (1981, pp. 156, 157, 161). The reflexivity follows from the fact that the function $\Delta_{FF} = (F^{-1} \circ F) - \mathrm{id} \equiv 0$ is non-negative, increasing and convex. ▨

Proposition 3.3. *a) $F =_{st} G$ is equivalent to $F = G$.*

b) $F =_{disp} G$ is equivalent to the existence of an $a \in \mathbb{R}$ such that $F(t) = G(t + a)$ for all $t \in \mathbb{R}$.

c) $F =_c G$ is equivalent to the existence of $a > 0, b \in \mathbb{R}$ such that $F(t) = G(a \cdot t + b)$ for all $t \in \mathbb{R}$.

Proof. Part a) follows directly from the definition of $\leq_{st}$. For parts b) and c), see Oja (1981, pp. 157, 161). ▨

Proposition 3.4. *a) If F is symmetric around $t_0 \in \mathbb{R}$, then $\nu(F) = t_0$ follows for all measures ν of central location.*

b) $\tau(F) \geq 0$ holds for all dispersion measures τ.

c) If F is symmetric, then $\gamma(F) = 0$ follows for all skewness measures γ.

Proof. a) Consider the cdf $\tilde{F}$ of $\tilde{X} = X - t_0$, which is symmetric around zero. Because of $\tilde{F}(t) = 1 - \tilde{F}(-t) = \mathbb{P}(\tilde{X} \geq -t) = \mathbb{P}(-\tilde{X} \leq t) = H_{-\tilde{X}}(t)$ for all $t \in \mathbb{R}$, $\tilde{X}$ and $-\tilde{X}$ have the same distribution. Hence, $\nu(\tilde{X}) = \nu(-\tilde{X}) = -\nu(\tilde{X})$ follows from (L1). This implies $0 = \nu(\tilde{X}) = \nu(X - t_0)$, yielding $\nu(X) = t_0$.

b) Let H_Z be the cdf of Z with $\mathbb{P}(Z = 0) = 1$ and let $F \in \mathcal{P}$. It follows that the quantile function H_Z^{-1} is constant and its first-order difference is constantly zero. Hence, $H_Z \leq_{disp} F$ holds. By (S1), this implies $0 = \tau(0 \cdot X) = \tau(Z) \leq \tau(X)$.

c) Consider $\tilde{X}$ as defined in the proof of a) and note that $\gamma(X) = \gamma(\tilde{X})$ follows from (S1). Furthermore, $\gamma(\tilde{X}) = \gamma(-\tilde{X}) = -\gamma(\tilde{X})$ yields $\gamma(\tilde{X}) = 0$.

∎

The statements of Proposition 3.4 were noted by Oja (1981, pp. 157, 159, 163), but not explicitly proved. Part c) is sometimes considered as an additional requirement in the definition of a skewness measure (see, e.g., Groeneveld and Meeden, 1984, p. 393), which is shown to be redundant by the above result.

Proposition 3.4 states that the value of central location and skewness measures is fixed for symmetric distributions, but, as opposed to dispersion measures, they can take negative values. Specifically, since $F \leq_c G$ means that G is more skewed to the right than F, any skewness measure γ is positive for right-skewed distributions and negative for left-skewed distributions. One might say that location and skewness are asymmetric characteristics while dispersion is a symmetric characteristic. In fact, early works on dispersion by Bickel and Lehmann (1976, 2012) treat dispersion of symmetric distributions as a different concept from dispersion of asymmetric distributions, which Bickel and Lehmann termed 'spread'.

If any dispersion measure τ is also required to preserve the strict version $<_{disp}$ of the dispersive order, so that $F <_{disp} G$ implies $\tau(F) < \tau(G)$, the statement of Proposition 3.4b) can be improved. Then, $\tau(F) = 0$ holds, if and only if all probability mass of F is concentrated on one point, and $\tau(F) > 0$ holds for all other cdf's F.

For the specific measures considered in the following subsections, this basic result is helpful for proving the properties (L1), (D1) and (S1).

Lemma 3.5. *Let $F \in \mathcal{P}_I$ and $a, b \in \mathbb{R}$.*

a) If $a \neq 0$, then $H_{a\cdot X+b}(t) = \left\{\begin{array}{ll} F\left(\dfrac{t-b}{a}\right), & \text{if } a > 0, \\ 1 - F\left(\dfrac{t-b}{a}\right), & \text{if } a < 0 \end{array}\right\}$ *holds for all $t \in \mathbb{R}$.*

b) If $a \neq 0$, then $h_{a\cdot X+b}(t) = \frac{1}{|a|} f\left(\frac{t-b}{a}\right)$ holds for all $t \in \mathbb{R}$.

c) If $a \neq 0$ and $F \in \mathcal{P}_I^{k+1}$, then $h^{(k)}_{a\cdot X+b}(t) = \frac{1}{|a|\cdot a^k} f^{(k)}\left(\frac{t-b}{a}\right)$ holds for all $t \in \mathbb{R}$ and all $k \in \mathbb{N}_0$.

d) $H^{-1}_{a\cdot X+b}(p) = \left\{\begin{array}{ll} a \cdot F^{-1}(p) + b, & \text{if } a \geq 0, \\ a \cdot F^{-1}(1-p) + b. & \text{if } a \leq 0 \end{array}\right\}$ *for all $p \in (0, 1)$.*

Proof. a) For all $t \in \mathbb{R}$, it holds that

$$H_{a\cdot X+b}(t) = \mathbb{P}(a \cdot X + b \leq t) = \begin{cases} \mathbb{P}\left(X \leq \frac{t-b}{a}\right) = F\left(\frac{t-b}{a}\right), & \text{if } a > 0, \\ \mathbb{P}\left(X \geq \frac{t-b}{a}\right) = 1 - F\left(\frac{t-b}{a}\right), & \text{if } a < 0. \end{cases}$$

b) Almost everywhere, F is differentiable and its derivative coincides with its density (see Kallenberg, 2021, pp. 42–43); the same is true for $H_{a\cdot X+b}$ by part a). In the following, we only consider the $t \in \mathbb{R}$ in which F is differentiable; for all other values of t, the two densities can be chosen to be equal. In the case $F \in \mathcal{P}_I^1$, they are unique. We have

$$\begin{aligned} h_{a\cdot X+b}(t) &= H'_{a\cdot X+b}(t) = \frac{\mathrm{d}}{\mathrm{d}t} F\left(\frac{t-b}{a}\right) \\ &= \left.\begin{cases} \frac{1}{a} f\left(\frac{t-b}{a}\right), & \text{if } a > 0, \\ -\frac{1}{a} f\left(\frac{t-b}{a}\right), & \text{if } a < 0 \end{cases}\right\} = \frac{1}{|a|} f\left(\frac{t-b}{a}\right). \end{aligned}$$

c) Since the densities are unique and sufficiently often differentiable, the assertion follows directly from b).

d) Because of $F \in \mathcal{P}_I$ and part a), the quantile functions coincide with the corresponding inverse functions. In the case $a = 0$, $a \cdot X + b = b$ holds almost surely, meaning that $H^{-1}_{a\cdot X+b} \equiv 0$. Now, let $a \neq 0$ and $p \in (0,1)$. Then, there exists a unique $t \in \mathbb{R}$ such that $p = H_{a\cdot X+b}(t)$. With part a), it follows that

$$\begin{aligned} H^{-1}_{a\cdot X+b}(p) &= H^{-1}_{a\cdot X+b}(H_{a\cdot X+b}(t)) = t = a \cdot \frac{t-b}{a} + b = a \cdot F^{-1}\left(F\left(\frac{t-b}{a}\right)\right) + b \\ &= \begin{cases} a \cdot F^{-1}\left(H_{a\cdot X+b}(t)\right) + b = a \cdot F^{-1}(p) + b, & \text{if } a > 0, \\ a \cdot F^{-1}\left(1 - H_{a\cdot X+b}(t)\right) + b = a \cdot F^{-1}(1-p) + b, & \text{if } a < 0. \end{cases} \end{aligned}$$

3.1.1. Moment-Based Measures

The most popular and well-known measures of central location, dispersion and skewness are based on moments. More specifically, they are given by the first, second and third standardized moment. All of them preserve the order of the corresponding convex characteristic.

Theorem 3.6. *a) The mapping*

$$\nu_M : \mathcal{L}^1 \to \mathbb{R}, \quad F \mapsto \mathbb{E}[X] = \int_{\mathbb{R}} t \,\mathrm{d}F(t)$$

is a measure of central location. It is also denoted by $\mu_X = \nu_M(X)$.

b) The mapping

$$\tau_M : \mathcal{L}^2 \to [0,\infty), \quad F \mapsto \sqrt{\mathbb{V}(X)} = \sqrt{\mathbb{E}[(X-\mu_X)^2]}$$

is a measure of dispersion. It is also denoted by $\sigma_X = \tau_M(X)$.

c) The mapping

$$\gamma_M : \mathcal{L}^3 \to \mathbb{R}, \quad F \mapsto \mathbb{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^3\right]$$

is a measure of skewness.

Proof. a) This follows directly from the representation $\mathbb{E}[X] = \int_0^1 F^{-1}(p)\,\mathrm{d}p$ and the linearity of the expectation, see Oja (1981, p. 157).

b) Oja (1981, p. 159) proved the chain of implications $F \leq_{disp} G \Rightarrow F \leq_{dil} G \Rightarrow \tau_M(F) \leq \tau_M(G)$, this proves (D2). (D1) again follows from the linearity of the expectation and the fact that $\sqrt{x^2} = |x|$ for any $x \in \mathbb{R}$.

c) (S2) was proved directly by van Zwet (1964, pp. 10–15). Oja (1981, pp. 163–164) offered a proof for an intersection-based skewness order that was shown to be weaker than $\leq_c$ on page 162. (S1) follows similarly to before by using part b).

Properties (L2), (D2) and (S2) for parts a), b) and c), respectively, can also be proved directly using Proposition 2.17b).

Of course, μ_X is called the expected value of X and σ_X is called the standard deviation of X. A historically popular notation for the moment skewness γ_M is β_1 (see, e.g., Pearson, 1895). Note that without the square root in the definition of τ_M, one would not obtain a measure of dispersion. While the crucial property (D2) would still be satisfied because the root functions are monotone on $[0, \infty)$, property (D1) would then by violated. In contrast, $\sqrt[3]{\gamma_M}$ is also a measure of skewness.

Although these moment-based measures are the most widely used measures for their respective characteristic, they have a number of disadvantages. First, they are very sensitive to the tails of distributions. This implies that the measures do not exist for sufficiently heavy-tailed distributions like the Cauchy distribution. The severity of this problem increases for higher moments, so γ_M is affected the most out of the three measures considered here. To this end, Eberl and Klar (2020, pp. 9–11) noted that, compared to other skewness measures, γ_M takes significantly larger values on markedly skewed distributions with bigger tails. Hosking (1990, p. 110) observed that γ_M 'is so sensitive to [...] extreme tails [...] that it is difficult to estimate in practice when the distribution is markedly skew'. This is confirmed by simulation results in Eberl and Klar (2020, pp. 11–14, 2022, pp. 390–393), where the empirical version of γ_M performs particularly bad for heavy-tailed distributions. While empirical and asymptotic properties are not explicitly considered in this thesis, these are still arguments against the use of γ_M in applications, at least for some distributions. To this end, ν_M, τ_M and γ_M are also not robust as empirical measures, meaning that they are sensitive to outliers in the data.

Hosking (1989, 1990) proposed measures of central location, dispersion and skewness that are based on so-called L-moments instead of traditional moments. A corresponding measure of kurtosis was also proposed, and is discussed in Chapter 4.3.1 of this thesis. L-moments are specific linear combinations of the order statistic of an iid sample. Many of the disadvantages of traditional moments are not exhibited by the corresponding L-moments: all of them exist, if the expected value exists, and they are more robust and less sensitive to heavy tails and outliers (see Hosking, 1990, pp. 105–107). Furthermore, any distribution with finite mean is uniquely characterized by its L-moments (see Hosking, 1990, pp. 107–108). The same statement is not true for traditional moments.

Let $k \in \mathbb{N}$, $F \in \mathcal{L}^1$ and let $X_{1:k} \leq \ldots \leq X_{k:k}$ be the order statistic of $X_1, \ldots, X_k \overset{\text{iid}}{\sim} F$. Then, the k-th L-moment is defined as

$$\lambda_k^F = \frac{1}{k} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} \mathbb{E}[X_{k-j:k}].$$

Since the expectation of the j-th component of an order statistic of sample size k is given by

$$\mathbb{E}[X_{j:k}] = \frac{k!}{(j-1)!(k-j)!} \int_{-\infty}^{\infty} t \cdot (F(t))^{j-1} \cdot (1 - F(t))^{k-j} \, \mathrm{d}F(t),$$

the k-th L-moment can also be written as

$$\lambda_k = \int_0^1 F^{-1}(p) \cdot P_{k-1}^*(p) \, \mathrm{d}p,$$

where $P_k^*(p) = \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} \binom{k+j}{j} p^j$, $p \in (0,1)$, denotes the k-th shifted Legendre polynomial.

The first two L-moments correspond to well-known measures of central location and dispersion. It follows directly from the definition that $\lambda_1^F = \mathbb{E}[X]$ holds, i.e., the first L-moment is equal to the first traditional moment, the mean. For the second L-moment, it holds that

$$\begin{aligned}
\lambda_2^F &= \tfrac{1}{2} \mathbb{E}[X_{2:2} - X_{1:2}] \\
&= \tfrac{1}{2} \left(\mathbb{P}(X_1 > X_2) \cdot \mathbb{E}[X_1 - X_2 | X_1 > X_2] + \mathbb{P}(X_2 > X_1) \cdot \mathbb{E}[X_2 - X_1 | X_2 > X_1] \right) \\
&= \tfrac{1}{2} \mathbb{E}[|X_1 - X_2|],
\end{aligned}$$

where $X_1, X_2 \overset{\text{iid}}{\sim} F$. There exist a few different names for the quantity $\mathbb{E}[|X_1 - X_2|]$; we call it Gini's mean difference. Note that, because of the symmetry between X_1 and X_2, $\lambda_2^F = \frac{1}{2} \mathbb{E}[X_1 - X_2 | X_1 > X_2]$ also holds.

As stated by the following result, the first three L-moments can be used to measure location, dispersion and skewness.

Theorem 3.7. *a) The mapping* $\nu_{LM} : \mathcal{L}^1 \to \mathbb{R}, F \mapsto \lambda_1^F$ *is a measure of central location.*

b) The mapping $\tau_{LM} : \mathcal{L}^1 \to [0, \infty), F \mapsto \lambda_2^F$ *is a measure of dispersion.*

c) The mapping $\gamma_{LM} : \mathcal{L}^1 \to \mathbb{R}, F \mapsto \frac{\lambda_3^F}{\lambda_2^F}$ is a measure of skewness.

Proof. a) Note $\nu_{LM} = \nu_M$ and see Theorem 3.6a).

b) For (D2), see Oja (1981, p. 160). For (D1), let $a, b \in \mathbb{R}$. It follows that

$$\tau_{LM}(a \cdot X + b) = \tfrac{1}{2}\mathbb{E}[|(a \cdot X_1 + b) - (a \cdot X_2 + b)|] = |a| \cdot \tau_{LM}(X).$$

c) For (S2), see Hosking (1989, pp. 6–7). For (S1), let $a, b \in \mathbb{R}$. It follows that

$$\begin{aligned}\lambda_3^{a \cdot X + b} &= \tfrac{1}{3}\mathbb{E}[(a \cdot X_{3:3} + b) - 2(a \cdot X_{2:3} + b) + (a \cdot X_{1:3} + b)] \\ &= a \cdot \lambda_3^X.\end{aligned}$$

Note that, if the affine transformation is applied to the order statistic in the case $a < 0$, $X_{1:3}$ and $X_{3:3}$ swap roles. The above equation holds in that case because the third L-moment is symmetric in these two random variables. The assertion now follows with part b).

▨

While the central location measure and the dispersion measure based on L-moments are equal to the L-moments themselves, the skewness measure γ_{LM} is only obtained after division by λ_2^F. Otherwise, it would not be a skewness measure. Compared to γ_M, γ_{LM} has the advantage that it is normalized. Specifically, $-1 < \gamma_{LM}(F) < 1$ holds for all $F \in \mathcal{L}^1$ (see Hosking, 1990, p. 108).

By definition, γ_{LM} uses two different order statistics: one of sample size three and one of sample size two. This implies that γ_{LM} exhibits a lack of intuitiveness and interpretability. However, Hosking (1990, p. 110) rectified this problem by using an identity by Sillitto (1951, p. 378) to show

$$\gamma_{LM}(F) = \frac{\mathbb{E}[X_{3:3}] - 2\mathbb{E}[X_{2:3}] + \mathbb{E}[X_{1:3}]}{\mathbb{E}[X_{3:3}] - \mathbb{E}[X_{1:3}]}$$

for any cdf $F \in \mathcal{L}^1$, where $X_{1:3} \leq X_{2:3} \leq X_{3:3}$ is the order statistic of $X_1, X_2, X_3 \overset{\text{iid}}{\sim} F$.

3.1.2. Quantile-Based Measures

The most well-known examples of quantile-based measures of convex characteristics are the median as a measure of central location and the interquartile distance as a measure of dispersion. A similarly constructed quantile-based measure of skewness dates back to Bowley (1901). The dispersion measure and the skewness measure can both be generalized by replacing the lower and upper quartiles with the α- and the $(1 - \alpha)$-quantile for an $\alpha \in (0, \frac{1}{2})$. All of these measures satisfy the corresponding properties given in Definition 3.1.

Theorem 3.8. *Let $\alpha \in (0, \frac{1}{2})$.*

a) The mapping $\nu_Q : \mathcal{P}_I \to \mathbb{R}, F \mapsto F^{-1}(\frac{1}{2})$ *is a measure of central location.*

b) The mapping $\tau_Q^\alpha : \mathcal{P}_I \to [0,\infty), F \mapsto F^{-1}(1-\alpha) - F^{-1}(\alpha)$ *is a measure of dispersion.*

c) The mapping

$$\gamma_Q^\alpha : \mathcal{P}_I \to \mathbb{R}, F \mapsto \frac{F^{-1}(1-\alpha) - 2F^{-1}(\frac{1}{2}) + F^{-1}(\alpha)}{F^{-1}(1-\alpha) - F^{-1}(\alpha)}$$

is a measure of skewness.

Proof. Parts a) and b) follows directly from the definitions of $\leq_{st}$ and $\leq_{disp}$. For part c), Groeneveld and Meeden (1984, pp. 393–394) proved (S1) generally and (S2) under some differentiability assumptions. However, the fact that (S2) is also satisfied without these assumptions follows from (2.11) in Example 2.10c) and the following Lemma 3.9. ▨

The following lemma gives an equivalent characterization of the convex transformation order based on quantities similar to γ_Q^α. For reference in the discrete Part II of this thesis, it is proved for general cdf's in $\mathcal{P}$ instead of only $\mathcal{P}_I$.

Lemma 3.9. *Let* $F, G \in \mathcal{P}$. *Then,* $F \leq_c G$ *is equivalent to*

$$\frac{F^{-1}(p_2) - 2F^{-1}(p_1) + F^{-1}(p_0)}{F^{-1}(p_2) - F^{-1}(p_0)} \leq \frac{G^{-1}(p_2) - 2G^{-1}(p_1) + G^{-1}(p_0)}{G^{-1}(p_2) - G^{-1}(p_0)} \tag{3.3}$$

for all $0 < p_0 < p_1 < p_2 < 1$ *with* $F^{-1}(p_0) < F^{-1}(p_2)$ *and* $G^{-1}(p_0) < G^{-1}(p_2)$.

Proof. If $F^{-1}(p_0) < F^{-1}(p_1) < F^{-1}(p_2)$ holds as well as $G^{-1}(p_0) < G^{-1}(p_1) < G^{-1}(p_2)$, the identity

$$\begin{aligned}
&\frac{F^{-1}(p_2) - 2F^{-1}(p_1) + F^{-1}(p_0)}{F^{-1}(p_2) - F^{-1}(p_0)} \\
&= \frac{(F^{-1}(p_2) - F^{-1}(p_1)) - (F^{-1}(p_1) - F^{-1}(p_0))}{F^{-1}(p_2) - F^{-1}(p_0)} \\
&= \left(\frac{F^{-1}(p_2) - F^{-1}(p_0)}{F^{-1}(p_2) - F^{-1}(p_1)}\right)^{-1} - \left(\frac{F^{-1}(p_2) - F^{-1}(p_0)}{F^{-1}(p_1) - F^{-1}(p_0)}\right)^{-1} \\
&= \left(1 + \frac{F^{-1}(p_1) - F^{-1}(p_0)}{F^{-1}(p_2) - F^{-1}(p_1)}\right)^{-1} - \left(1 + \frac{F^{-1}(p_2) - F^{-1}(p_1)}{F^{-1}(p_1) - F^{-1}(p_0)}\right)^{-1}
\end{aligned}$$

proves that (2.11) and (3.3) are equivalent via the increasing transformation $(0,\infty) \to (-1,1)$, $t \mapsto (1+t^{-1})^{-1} - (1+t)^{-1}$. In the cases $F^{-1}(p_1) = F^{-1}(p_2)$ and $G^{-1}(p_0) = G^{-1}(p_1)$, both (2.11) and (3.3) are always true (and therefore equivalent) as the left hand sides take the minimal possible values and the right hand sides take the maximal possible values, respectively.

Two cases remain to be considered: $F^{-1}(p_0) = F^{-1}(p_1)$ and $G^{-1}(p_1) = G^{-1}(p_2)$. If $F^{-1}(p_0) = F^{-1}(p_1)$ holds, then the left hand side of (2.11) is infinite and the inequality is satisfied if and only if $G^{-1}(p_0) = G^{-1}(p_1)$. In the same case, the left hand side of (3.3) is equal to 1, so the inequality is satisfied if and only if the right side is also equal to 1, which is equivalent to $G^{-1}(p_0) = G^{-1}(p_1)$. The proof of the assertion in the remaining case of $G^{-1}(p_1) = G^{-1}(p_2)$ is analogous as both right hand sides are minimal. Consequently, both (2.11) and (3.3) are equivalent to the left sides being minimal, which is equivalent to $F^{-1}(p_1) = F^{-1}(p_2)$ for both inequalities. ▨

In addition to its contribution to the proof of Theorem 3.8c), Lemma 3.9 states that the mapping

$$\gamma_{QA}^{\alpha} : \mathcal{P}_I \to [0, \infty], F \mapsto \frac{F^{-1}(1-\alpha) - F^{-1}(\frac{1}{2})}{F^{-1}(\frac{1}{2}) - F^{-1}(\alpha)}$$

also preserves property (S2) for skewness measures. It does, however, not satisfy (S1) since $\gamma_{QA}^{\alpha}(F) = 1$ holds for $F \in \mathcal{S}$, which contradicts the necessary condition from Proposition 3.4c).

The measures ν_Q, τ_Q^{α} and γ_Q^{α} are directly connected to the definitions of $\leq_{st}$, $\leq_{disp}$ and $\leq_c$ from Example 2.10 and Definition 2.11. In the case of γ_Q^{α} and $\leq_c$, this connection is made by Lemma 3.9. However, the three measures can also be understood as differences of zeroth, first and second order of the quantile function. The evaluation points of these differences are always chosen symmetrically around $\frac{1}{2}$. Since ν_Q is the difference of zeroth order, it arises from just one evaluation of F^{-1} at $\frac{1}{2}$, so the evaluation point is fixed and there is no parameter α to be varied. The dispersion measure τ_Q^{α} is the first order difference, meaning that the two evaluation points have equal distance to $\frac{1}{2}$ and can be varied in $\alpha \in (0, \frac{1}{2})$. The second order difference used for measuring skewness constitutes the numerator of the corresponding measure γ_Q^{α} and is given by

$$(F^{-1}(1-\alpha) - F^{-1}(\tfrac{1}{2})) - (F^{-1}(\tfrac{1}{2}) - F^{-1}(\alpha)) = F^{-1}(1-\alpha) - 2F^{-1}(\tfrac{1}{2}) + F^{-1}(\alpha).$$

This difference is then divided by the corresponding dispersion measure τ_Q^{α} in order to obtain a measure that is invariant to dispersion. The fact that the limit of the suitably normalized k-th order difference of a function ($k \in \mathbb{N}_0$) is the k-th derivative of the same function, which is directly connected to the k-th convex characteristic, is discussed in Section 3.1.3.

The three quantile-based measures from Theorem 3.8 are structurally very similar to the L-moment-based measures from Theorem 3.7, which can be written as

$$\begin{aligned}
\nu_{LM}(F) &= \mathbb{E}[X_{1:1}], \\
\tau_{LM}(F) &= \mathbb{E}[X_{2:2}] - \mathbb{E}[X_{1:2}], \\
\gamma_{LM}(F) &= \frac{\mathbb{E}[X_{3:3}] - 2\mathbb{E}[X_{2:3}] + \mathbb{E}[X_{1:3}]}{\mathbb{E}[X_{3:3}] - \mathbb{E}[X_{1:3}]}.
\end{aligned}$$

The basic shape of the L-moment-based measures is the same and can also be characterized via k-th order differences, $k = 0, 1, 2$. The only difference is that the lower, middle and upper quantiles are replaced by the expected values of corresponding lower, middle and upper components of an order statistic. This structural similarity of the skewness measures also explains that, just like γ_{LM}, γ_Q^α is normalized, since $-1 \leq \gamma_Q^\alpha \leq 1$ was shown to hold for all $\alpha \in (0, \frac{1}{2})$ by Groeneveld and Meeden (1984, p. 394).

The dependence of the measures τ_Q^α and γ_Q^α on the parameter $\alpha \in (0, \frac{1}{2})$ has both negative and positive consequences. On the negative side, one is required to choose a specific value for α in order to utilize the measures, specifically in applications. This can be circumvented by integrating with respect to α over all possible choices. For the dispersion measure, we obtain

$$\begin{aligned}\tau_{IQ}(F) &= \int_0^{\frac{1}{2}} \tau_Q^\alpha(F)\,\mathrm{d}\alpha = \int_0^{\frac{1}{2}} (F^{-1}(1-\alpha) - F^{-1}(\alpha))\,\mathrm{d}\alpha \\ &= \int_{\frac{1}{2}}^1 (F^{-1}(\alpha) - F^{-1}(\tfrac{1}{2}))\,\mathrm{d}\alpha + \int_0^{\frac{1}{2}} (F^{-1}(\tfrac{1}{2}) - F^{-1}(\alpha))\,\mathrm{d}\alpha \\ &= \int_0^1 |F^{-1}(\alpha) - F^{-1}(\tfrac{1}{2})|\,\mathrm{d}\alpha \\ &= \mathbb{E}[|X - F^{-1}(\tfrac{1}{2})|],\end{aligned} \tag{3.4}$$

i.e., the mean absolute deviation from the median, a well known dispersion measure in the literature. The fact that τ_{IQ} is a measure of dispersion as specified in Definition 3.1b) follows directly from Theorem 3.8 and the linearity of the integral. This result can also be found in the literature (see, e.g., Hürlimann, 2002, p. 15). Integrating the skewness measure γ_Q^α in the same way obviously also yields another skewness measure. However, the corresponding quantity is considered in a preprint by Arachchige and Prendergast (2019) and does not have a closed form, which it is not very appealing. Instead, Groeneveld and Meeden (1984, p. 392) have proposed integrating the numerator and the denominator separately, which leads to the much more accessible quantity

$$\begin{aligned}\gamma_{IQ}(F) &= \frac{\int_0^{\frac{1}{2}} (F^{-1}(1-\alpha) - 2F^{-1}(\frac{1}{2}) + F^{-1}(\alpha))\,\mathrm{d}\alpha}{\int_0^{\frac{1}{2}} (F^{-1}(1-\alpha) - F^{-1}(\alpha))\,\mathrm{d}\alpha} \\ &= \frac{\int_0^1 F^{-1}(\alpha)\,\mathrm{d}\alpha - F^{-1}(\frac{1}{2})}{\tau_{IQ}(F)} \\ &= \frac{\mathbb{E}[X] - F^{-1}(\frac{1}{2})}{\mathbb{E}[|X - F^{-1}(\frac{1}{2})|]}.\end{aligned} \tag{3.5}$$

It was originally proven to be a skewness measure by Groeneveld and Meeden (1984, p. 394) and again by Eberl and Klar (2019, p. 271) under weakened assumptions. Note that the similar quantity

$$\frac{\mathbb{E}[X] - F^{-1}(\frac{1}{2})}{\sigma_X}, \tag{3.6}$$

proposed by Pearson (1895, p. 370) and Yule (1911, p. 150), does not satisfy (S2) and is therefore not a measure of skewness. A discrete counterexample was given by van Zwet (1964, pp. 16–17), for continuous examples see Groeneveld and Meeden (1984, p. 396, using the gamma distribution) or Eberl and Klar (2019, p. 275, using the Weibull distribution). The results concerning the integrated quantile measures are summarized in the following theorem.

Theorem 3.10. *a) The mapping* $\tau_{IQ} : \mathcal{L}^1 \to [0, \infty), F \mapsto \mathbb{E}[|X - F^{-1}(\frac{1}{2})|]$ *is a measure of dispersion.*

b) The mapping $\gamma_{IQ} : \mathcal{L}^1 \to \mathbb{R}, F \mapsto \frac{\mathbb{E}[X]-F^{-1}(\frac{1}{2})}{\mathbb{E}[|X-F^{-1}(\frac{1}{2})|]}$ *is a measure of skewness.*

The fact that, by varying α, γ_Q^α identifies a whole family of quantile-based skewness measures also has a positive effect. By considering the entire family $\{\gamma_Q^\alpha : \alpha \in (0, \frac{1}{2})\}$, an equivalent characterization for the symmetry of a distribution is obtained.

Proposition 3.11. $\gamma_Q^\alpha(F) = 0$ *for all* $\alpha \in (0, \frac{1}{2})$ *is equivalent to* $F \in \mathcal{S}$.

Proof. Since $\gamma_Q^\alpha(X)$ is invariant under shifts of X, we assume without restriction that $F^{-1}(\frac{1}{2}) = 0$. Thus, $F \in \mathcal{S}$ is equivalent to $F(t) = 1 - F(-t)$ for all $t \in \mathbb{R}$.

For each $p \in (0, 1)$, there exists a $t \in \mathbb{R}$ such that $p = F(t)$, yielding

$$F^{-1}(1 - p) = F^{-1}(1 - F(t)) = F^{-1}(F(-t)) = -t = -F^{-1}(p),$$

if F is symmetric. Conversely, $F^{-1}(1-p) = -F^{-1}(p)$ for all $p \in (0, 1)$ implies $F(t) = 1-F(-t)$ for all $t \in D_F$ in a similar way, as well as $F(t) = 0$ for $t < \inf(D_F)$ and $F(t) = 1$ for $t > \sup(D_F)$. Overall, the equivalence

$$F \in \mathcal{S} \Leftrightarrow F^{-1}(1 - p) = -F^{-1}(p) \ \forall p \in (0, 1) \tag{3.7}$$

holds.

Note that, for $\alpha \in (0, \frac{1}{2})$, $\gamma_Q^\alpha(F) = 0$ is equivalent to $F^{-1}(1 - \alpha) = -F^{-1}(\alpha)$. The assertion now follows from equivalence (3.7) by choosing $p = \alpha$ for $p \in (0, \frac{1}{2})$ and $p = 1 - \alpha$ for $p \in (\frac{1}{2}, 1)$. ∎

This kind of result does not and cannot hold for γ_M, because it requires a family of skewness measures instead of just a single measure. While one implication is true for all skewness measures due to Proposition 3.4c), counterexamples for the other implication are given by Johnson et al. (1980) and Ramberg et al. (1979) using the generalized lambda distribution and by Klar (2015) using the gamma difference distribution.

3.1.3. Density-Based Measures

As stated in Corollary 2.12, the order of the k-th convex characteristic can be equivalently described using the k-th derivative of the modified RIDF Δ_{FG}, if F and G are sufficiently smooth. In that case, the corresponding derivative of Δ_{FG} can be written using the derivatives of F and G. Since their first derivatives are their densities, this representation usually involves f and g and derivatives thereof. If the F-terms and the G-terms in the k-th derivative of Δ_{FG} can be disentangled in a symmetric way, a measure of the k-th convex characteristic based on densities is obtained.

Throughout Section 3.1.3 and whenever density-based measures are considered within Part I of this thesis, we assume the densities of all considered distributions to be strictly positive on the interior of their supports. This additional assumption is required because evaluations of the densities appear in denominators of density-based measures.

Theorem 3.12. *a) The mapping $\nu_D : \mathcal{P}_I \to \mathbb{R}, F \mapsto F^{-1}(\frac{1}{2})$ is a measure of central location.*

b) The mapping

$$\tau_D : \mathcal{P}_I^1 \to [0, \infty), \quad F \mapsto \frac{1}{f(F^{-1}(\frac{1}{2}))}$$

is a measure of dispersion.

c) The mapping

$$\gamma_D : \mathcal{P}_I^2 \to \mathbb{R}, \quad F \mapsto -\frac{f'(F^{-1}(\frac{1}{2}))}{(f(F^{-1}(\frac{1}{2})))^2}$$

is a measure of skewness.

Proof. a) Note that $\nu_D = \nu_Q$ and see Theorem 3.8a).

b) Using

$$R'_{FG}(t) = (G^{-1})'(F(t)) \cdot f(t) = \frac{f(t)}{g(G^{-1}(F(t)))} = \frac{f(t)}{g(R_{FG}(t))}$$

for all $t \in \mathrm{supp}(F)$, we obtain that the following equivalences hold for $F, G \in \mathcal{P}_I^1$:

$$\begin{aligned} F \leq_1 G &\Leftrightarrow \Delta'_{FG}(t) \geq 0 \;\forall t \in \mathrm{supp}(F) \Leftrightarrow R'_{FG}(t) \geq 1 \;\forall t \in \mathrm{supp}(F) \\ &\Leftrightarrow \frac{f(t)}{g(R_{FG}(t))} \geq 1 \;\forall t \in \mathrm{supp}(F) \Leftrightarrow \frac{1}{g(G^{-1}(p))} \geq \frac{1}{f(F^{-1}(p))} \;\forall p \in (0,1). \end{aligned}$$

Thus, τ_D satisfies (D2). For (D1), let $a, b \in \mathbb{R}$ and $p \in (0,1)$. Then, Lemma 3.5 yields for all $F \in \mathcal{P}_I^1$ that

$$\frac{1}{h_{a\cdot X+b}(H^{-1}_{a\cdot X+b}(p))} = \begin{cases} \frac{1}{\frac{1}{|a|} f\left(a \cdot \frac{F^{-1}(p)-b}{a} + b\right)} = |a| \cdot \frac{1}{f(F^{-1}(p))}, & \text{if } a \geq 0, \\ \frac{1}{\frac{1}{|a|} f\left(a \cdot \frac{F^{-1}(1-p)-b}{a} + b\right)} = |a| \cdot \frac{1}{f(F^{-1}(1-p))}, & \text{if } a \leq 0 \end{cases}$$

holds. Note that, for $p \in (0,1)$, $p = 1 - p$ is equivalent to $p = \frac{1}{2}$.

c) Let $F, G \in \mathcal{P}_I^2$. Because of

$$R''_{FG}(t) = \frac{f'(t) \cdot (g(R_{FG}(t)))^2 - f^2(t) \cdot g'(R_{FG}(t))}{(g(R_{FG}(t)))^3}$$

for all $t \in \operatorname{supp}(F)$, the equivalences

$$\begin{aligned}
F \leq_2 G &\Leftrightarrow R''_{FG}(t) \geq 0 \ \forall t \in \operatorname{supp}(F) \\
&\Leftrightarrow f'(t) \cdot (g(R_{FG}(t)))^2 \geq f^2(t) \cdot g'(R_{FG}(t)) \ \forall t \in \operatorname{supp}(F) \\
&\Leftrightarrow \frac{f'(t)}{f^2(t)} \geq \frac{g'(R_{FG}(t))}{(g(R_{FG}(t)))^2} \ \forall t \in \operatorname{supp}(F) \\
&\Leftrightarrow -\frac{g'(G^{-1}(p))}{(g(G^{-1}(p)))^2} \geq -\frac{f'(F^{-1}(p))}{(f(F^{-1}(p)))^2} \ \forall p \in (0,1)
\end{aligned}$$

hold, proving that (S2) is satisfied. (S1) can be shown similarly to (D1) in b), using the identity

$$\begin{aligned}
&-\frac{h'_{a\cdot X+b}(H^{-1}_{a\cdot X+b}(p))}{(h_{a\cdot X+b}(H^{-1}_{a\cdot X+b}(p)))^2} \\
&= \begin{cases} -\frac{|a|^2}{|a|\cdot a} \cdot \frac{f'(F^{-1}(p))}{(f(F^{-1}(p)))^2} = \operatorname{sgn}(a) \cdot \left(-\frac{f'(F^{-1}(p))}{(f(F^{-1}(p)))^2}\right), & \text{if } a \geq 0, \\ -\frac{|a|^2}{|a|\cdot a} \cdot \frac{f'(F^{-1}(1-p))}{(f(F^{-1}(1-p)))^2} = \operatorname{sgn}(a) \cdot \left(-\frac{f'(F^{-1}(1-p))}{(f(F^{-1}(1-p)))^2}\right), & \text{if } a \leq 0, \end{cases}
\end{aligned}$$

which is true for all $a, b \in \mathbb{R}$ and all $p \in (0,1)$.

▨

The proof of Theorem 3.12 also implies two further results.

Corollary 3.13. *Let $p \in (0,1)$.*

a) The mapping $\nu_D^p : \mathcal{P}_I \to \mathbb{R}, F \mapsto F^{-1}(p)$ satisfies (L2). It satisfies (L1), if and only if $p = \frac{1}{2}$.

b) The mapping

$$\tau_D^p : \mathcal{P}_I^1 \to [0,\infty), \quad F \mapsto \frac{1}{f(F^{-1}(p))}$$

satisfies (D2). It satisfies (D1), if and only if $p = \frac{1}{2}$.

c) The mapping

$$\gamma_D^p : \mathcal{P}_I^2 \to \mathbb{R}, \quad F \mapsto -\frac{f'(F^{-1}(p))}{(f(F^{-1}(p)))^2}$$

satisfies (S2). It satisfies (S1), if and only if $p = \frac{1}{2}$.

Obviously, we have $\nu_D^{1/2} = \nu_D$, $\tau_D^{1/2} = \tau_D$ and $\gamma_D^{1/2} = \gamma_D$. Note that the statements of Corollary 3.13 are stronger than those of Theorem 3.12. As noted after Definition 3.1, ν_D^p is a measure of non-central location for $p \neq \frac{1}{2}$. Similarly, one might say that τ_D^p and γ_D^p measure dispersion and skewness in a non-central way for $p \neq \frac{1}{2}$.

Corollary 3.14. *a) Let $F, G \in \mathcal{P}_I$. $F \leq_{st} G$ holds, if and only if $\nu_D^p(F) \leq \nu_D^p(G)$ for all $p \in (0,1)$.*

b) Let $F, G \in \mathcal{P}_I^1$. $F \leq_{disp} G$ holds, if and only if $\tau_D^p(F) \leq \tau_D^p(G)$ for all $p \in (0,1)$.

a) Let $F, G \in \mathcal{P}_I^2$. $F \leq_c G$ holds, if and only if $\gamma_D^p(F) \leq \gamma_D^p(G)$ for all $p \in (0,1)$.

The statements of Corollary 3.14 are also stronger than those of Theorem 3.12 concerning the properties (L2), (D2) and (S2). Specifically, equivalent conditions for the fundamental orders of location, dispersion and skewness are given, using measures of these characteristics as well as the closely related quantities τ_D^p and γ_D^p.

A result similar to Corollary 3.14 is also true for the quantile-based measures given in Theorem 3.8. Suppose that the evaluation points there are not chosen symmetrically around $\frac{1}{2}$ and are instead arbitrary but still ordered. Then, one just obtains the definitions of the three orders $\leq_{st}$, $\leq_{disp}$ and $\leq_c$ given in Example 2.10 and Definition 2.11. However, while these definitions require up to three evaluation points of the quantile function for each choice of parameters, all equivalent characterizations in Corollary 3.14 only consider one evaluation of the corresponding measure. This reduction of evaluation points is reflected in the equivalent characterization of the order of the k-th convex characteristic via the k-th derivative of the (modified) RIDF, given in Corollary 2.12. To this end, it is shown in the following that the density-based measures can also be obtained as limiting values of the quantile-based measures in the form of derivatives of the quantile function. This is only done for the measures of dispersion and skewness since $\nu_Q = \nu_D$ holds.

Remark 3.15. a) Let $F \in \mathcal{P}_I^1$. Obviously, $\lim_{\alpha \nearrow \frac{1}{2}} \tau_Q^\alpha(F) = 0$ holds. In order to obtain a derivative, we multiply with a factor that only depends on α and not on F itself, which vanished for $\alpha \nearrow \frac{1}{2}$. This yields

$$\begin{aligned}\lim_{\alpha \nearrow \frac{1}{2}} \frac{1}{1-2\alpha} \cdot \tau_Q^\alpha(F) &= \lim_{\alpha \nearrow \frac{1}{2}} \frac{F^{-1}(1-\alpha) - F^{-1}(\alpha)}{(1-\alpha)-\alpha} \\ &= \left(F^{-1}\right)'\left(\frac{1}{2}\right) = \frac{1}{f(F^{-1}(\frac{1}{2}))} \\ &= \tau_D(F).\end{aligned}$$

b) Let $F \in \mathcal{P}_I^2$. After multiplying with a similar α-dependent factor as before, we now obtain the limiting value for $\alpha \nearrow \frac{1}{2}$ by rewriting the resulting term as the ratio of a second-order and a first-order difference quotient

$$\lim_{\alpha \nearrow \frac{1}{2}} \frac{4}{1-2\alpha} \cdot \gamma_Q^\alpha(F) = 4 \cdot \lim_{\alpha \nearrow \frac{1}{2}} \frac{\frac{F^{-1}(1-\alpha)-2F^{-1}(\frac{1}{2})+F^{-1}(\alpha)}{(1-2\alpha)^2}}{\frac{F^{-1}(1-\alpha)-F^{-1}(\alpha)}{1-2\alpha}} = \frac{\left(F^{-1}\right)''\left(\frac{1}{2}\right)}{\left(F^{-1}\right)'\left(\frac{1}{2}\right)}$$

$$= -\frac{f'(F^{-1}(\frac{1}{2}))}{(f(F^{-1}(\frac{1}{2})))^2} = \gamma_D(F),$$

where we used that

$$\left(F^{-1}\right)''(p) = \left(\frac{1}{f(F^{-1}(p))}\right)' = -\frac{f'(F^{-1}(p))}{(f(F^{-1}(p)))^3}$$

for $p \in (0,1)$. Hence, the density-based skewness measure is a limiting value of a (rescaled) quantile-based skewness measure, analogous to the dispersion measures. The representation of $\gamma_D(F)$ using the derivatives of F^{-1} is of a familiar structure. All skewness measures considered this far consist of a quantity measuring skewness in the numerator and a dispersion measure in the denominator for standardization. Since $(F^{-1})'(\frac{1}{2})$ is a dispersion measure (see part a)), this structure can also be found in γ_D with $(F^{-1})''(\frac{1}{2})$ being the part that actually measures skewness.

While the use of a single evaluation point is a positive in Corollary 3.14, it is more of a negative, when the density-based measures ν_D, τ_D and γ_D are utilized. While they are very robust, they only consider the distribution at one point, which is often not representative enough. One approach to rectify this is by utilizing the generalized quantities from Corollary 3.13. Since they also satisfy (L2) / (D2) / (S2), aggregating them in a way that is symmetric around $\frac{1}{2}$ should generally yield another measure of the corresponding characteristic. This is done in the following remark by using integration.

Remark 3.16. Let μ be a symmetric, finite measure on the set $(0,1)$.

a) Let $F \in \mathcal{P}_I$ and define $\nu_{ID}^{\mu}(F) = \int_0^1 \nu_D^p(F)\,\mu(\mathrm{d}p)$. Because of the monotonicity of the integral, ν_{ID}^{μ} inherits property (L2) from ν_D^p. Concerning (L1), note that, for all $a < 0$, $b \in \mathbb{R}$,

$$\begin{aligned}\nu_{ID}^{\mu}(aX+b) &= \int_0^1 \nu_D^p(aX+b)\,\mu(\mathrm{d}p) = a \cdot \int_0^1 F_X^{-1}(1-p)\,\mu(\mathrm{d}p) + b \\ &= a \cdot \int_0^1 F_X^{-1}(p)\,\mu(\mathrm{d}p) + b = a \cdot \int_0^1 \nu_D^p(X)\,\mu(\mathrm{d}p) + b = a \cdot \nu_{ID}^{\mu}(X) + b\end{aligned}$$

holds because of the symmetry of μ. Since the same identity is obviously true in the case $a \geq 0$, ν_{ID}^{μ} satisfies (L1) and, therefore, is a measure of central location. Specific versions of $\nu_{ID}^{\mu}(F)$ include the mean $\mathbb{E}[X]$ for $\mu = \mathcal{U}(0,1)$ and the α-truncated mean $\mathbb{E}[X|_{[F^{-1}(\alpha),F^{-1}(1-\alpha)]}]$ for $\mathcal{U}(\alpha, 1-\alpha), \alpha \in (0, \frac{1}{2})$. Note that there are also discrete versions of ν_{ID}^{μ}, e.g. the median $F^{-1}(\frac{1}{2}) = \nu_D^{1/2}$ for $\mu = \delta_{1/2}$ or the arithmetic mean over any number of symmetrically chosen quantiles.

b) Let $F \in \mathcal{P}_I^1$ and define $\tau_{ID}^{\mu}(F) = \int_0^1 \tau_D^p(F)\,\mu(\mathrm{d}p)$. With analogous reasoning to a), τ_{ID}^{μ} satisfies both (D1) and (D2), implying that it is a measure of dispersion. The most notable family of examples is obtained by choosing $\mu = \mathcal{U}(\alpha, 1-\alpha)$ for $\alpha \in [0, \frac{1}{2})$, then

$$\tau_{ID}^{\mu}(F) = \int_{\alpha}^{1-\alpha} \frac{1}{f(F^{-1}(p))}\,\mathrm{d}p = \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \frac{1}{f(t)} \cdot f(t)\,\mathrm{d}t = F^{-1}(1-\alpha) - F^{-1}(\alpha)$$

follows. In the particular cases $\alpha = 0$ and $\alpha = \frac{1}{4}$, τ_{ID}^{μ} is the absolute range of the distribution and its interquartile range, respectively.

c) Let $F \in \mathcal{P}_I^2$. Similarly to parts a) and b), one can proof that the mapping $\gamma_{ID}^{\mu}(F) = \int_0^1 \gamma_D^p(F)\,\mu(\mathrm{d}p)$ is a skewness measure. To the knowledge of the author, no skewness measure of this kind has been mentioned in the literature so far. While the method of obtaining the measures in parts a) and b) also seems to be new, the thereby obtained measures are most certainly not. Since the special case $\mu = \mathcal{U}(\alpha, 1-\alpha)$ yields well-known measures for central location and dispersion in the previous parts, we also consider it here and obtain

$$\begin{aligned}\gamma_{ID}^{\mu}(F) &= -\int_{\alpha}^{1-\alpha} \frac{f'(F^{-1}(p))}{(f(F^{-1}(p)))^2}\,\mathrm{d}p = -\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \frac{f'(t)}{(f(t))^2} \cdot f(t)\,\mathrm{d}t \\ &= \log(f(F^{-1}(\alpha))) - \log(f(F^{-1}(1-\alpha))) \\ &= \log\left(\frac{f(F^{-1}(\alpha))}{f(F^{-1}(1-\alpha))}\right).\end{aligned}$$

In the examples from Remark 3.16 that integrate with respect to the measure $\mu = \mathcal{U}(\alpha, 1-\alpha)$, α acts as a trade-off parameter. For small α, the resulting measure is less robust but takes the tails of the distribution more into account; for large α, the opposite is the case. This is exemplified by ν_{ID}^{μ}, which is the mean for $\alpha \searrow 0$, the median for $\alpha \nearrow \frac{1}{2}$, and a trimmed mean in between.

3.1.4. Measures Based on the Mode

The mode of a distribution is the value that maximizes its likelihood. It is often only defined for unimodal distributions, meaning that the density or pmf of the distribution has a unique local maximum. In order to avoid unclear definitions, we assume $F \in \mathcal{P}_{I,um}^2$ throughout Section 3.1.4, where $\mathcal{P}_{I,um}^2$ denotes the set of all unimodal cdf's in $\mathcal{P}_I^2$. Now, we define the mode of an $F \in \mathcal{P}_{I,um}^2$, denoted by M_F, as the unique local maximum of F.

The mode is often associated with the mean and the median and it is usually assumed to also be a measure of central location or central tendency. It does, however, not satisfy condition (L2) and is therefore neither a measure of central location nor of non-central location in the sense of Definition 3.1a) and the subsequent remarks.

Theorem 3.17. *The mapping $M_\cdot : \mathcal{P}_{I,um}^2 \to \mathbb{R}, F \mapsto M_F$ does not satisfy (L2) and is not a measure of central location.*

Proof. We construct a counterexample from a beta distribution and a log-normal distribution. Let X be symmetrically beta distributed with shape parameter 2, so it has the density $f(t) = \frac{1}{6}t^2(1-t)^2 \cdot \mathbb{1}_{[0,1]}(t), t \in \mathbb{R}$. Furthermore, let Y be standard log-normally distributed, so $Y = \exp(Z)$ with $Z \sim \mathcal{N}(0,1)$. Then, $X \leq_{st} Y$ can be confirmed numerically as $F(t) \geq G(t)$ holds for all $t \in \mathbb{R}$ (see lower panel of Figure 3.2)). The mode of X is given by $M_F = \frac{1}{2}$ because of its symmetry while the mode of Y can be calculated by differentiating the log-normal density and is given by $M_G = \exp(-1) \approx 0.368$. Thus, $M_G < M_F$ holds (see upper panel of Figure 3.2), which contradicts (L2). ▨

Note that a counterexample could also be constructed using two cdf's with the same support. For that, one could simply remove a bit of probability mass from the peak of f and remodel it in the form of a long, thin tail on the interval $(1, \infty)$.

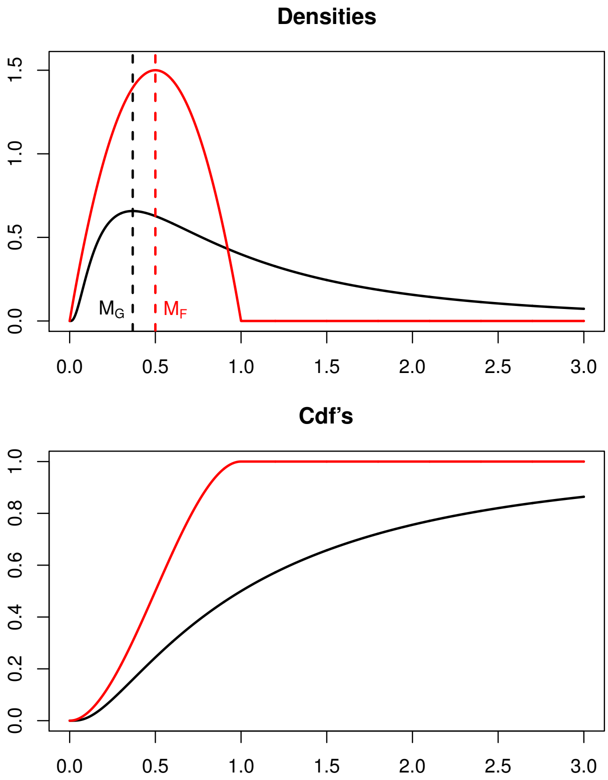

Figure 3.2.: Illustration of the counterexample from the proof of Theorem 3.17.

Although the mode does not qualify as a measure of location, it also finds use in measures for other characteristics, specifically for skewness. A number of results have been published on how skewness is represented in the relationship between mean, median and mode. The classical statement on this topic is that, for a right-skewed cdf F, the inequalities $M_F \leq F^{-1}(\frac{1}{2}) \leq \mathbb{E}[X]$ hold (see Runnenburg (1978) for a historical overview). The reverse chain of inequalities $\mathbb{E}[X] \leq F^{-1}(\frac{1}{2}) \leq M_F$ was proven to hold by van Zwet (1979), if the condition

$$F(F^{-1}(\tfrac{1}{2}) - t) + F(F^{-1}(\tfrac{1}{2}) + t) \geq 1 \quad \forall t \in \mathbb{R}$$

is fulfilled. Van Zwet also gives a stronger sufficient condition based on the convex transformation order, namely that $F \leq_c \tilde{F}$, where $\tilde{F}$ is the mirrored cdf given by $\tilde{F}(t) = 1 - F(-t), t \in \mathbb{R}$. Both conditions can be interpreted as F being skewed to the left in some sense with the first one being notably weaker. There are also publications opposing this classical rule by citing counterexamples (see Abadir, 2005 and von Hippel, 2005). However, the former paper uses a narrow definition of skewness by identifying it with the measure γ_M, and the latter paper mostly relies on discrete examples. The fact that the convex transformation order $\leq_c$ is virtually useless for discrete distributions is demonstrated in Eberl and Klar (2019) and is briefly discussed in Section 8.2 of this thesis, leaving the concept of skewness without a foundation. The problems that arise when quantile-based measures are applied to discrete distributions are discussed for dispersion in Sections 7.4 and 7.5.

A skewness measure based on the mode was proposed by Pearson (1895, p. 370) and Yule (1911, p. 150), who considered

$$\frac{\mathbb{E}[X] - M_F}{\sigma_X} \tag{3.8}$$

in the same context as the quantity in (3.6). However, Arnold and Groeneveld (1995, p. 34) proved that (3.8) does not satisfy property (S2) and, therefore, is not a skewness measure. They instead proposed another quantity based on the mode, which indeed qualifies as a measure of skewness according to Definition 3.1.

Theorem 3.18. *The mapping $\gamma_{Mode} : \mathcal{P}^2_{I,um} \to \mathbb{R}, F \mapsto 1 - 2F(M_F)$ is a measure of skewness.*

Proof. See Arnold and Groeneveld (1995, p. 35). ▨

Like a number of other skewness measures, γ_{Mode} is normalized in the sense that it only takes values in the interval $[-1, 1]$. Note that $\gamma_{Mode}(F)$ increases as the difference $F^{-1}(\frac{1}{2}) - M_F$ between the median and the mode increases; the same is true for $\gamma_{IQ}(F)$ and the difference $\mathbb{E}[X] - F^{-1}(\frac{1}{2})$ between the mean and the median. The fact that both γ_{Mode} and γ_{IQ} satisfy (S1) and (S2) and, therefore, measure skewness in a well-founded way further supports the classical connection between skewness on one side and mean, median and mode on the other.

3.2. Orders and Measures Based on Expectiles

Expectiles are a specific family of generalized quantiles that possess a number of desirable properties. They are formally introduced and subsequently discussed in Section 2.3. Since the expectiles of a cdf F are only defined in the case $F \in \mathcal{L}^1$, this is assumed to be true throughout Section 3.2. Expectiles can occupy a similar role in measuring convex characteristics as L-moments: they offer a sensible compromise between the robustness of quantile-based measures and the more holistic perspective of moment-based measures.

Obvious choices for expectile-based measures are obtained by replacing each quantile in the measures ν_Q, τ_Q^α and γ_Q^α by the corresponding expectile. However, while the quantile-based measures preserve the order of the corresponding convex characteristic by definition, this is not as simple for expectile-based measures of the same structure. In order to bridge this gap, it is convenient to define expectile-based orders of location, dispersion and skewness. The relationships between these orders and more traditional orders also shed some light on how the quantification of these characteristics via expectiles differs from the approaches discussed in Section 3.1.

3.2.1. Location

The only location order considered in this thesis so far is the usual stochastic order $\leq_{st}$, where $F \leq_{st} G$ is defined by $F^{-1}(p) \leq G^{-1}(p)$ for all $p \in (0,1)$. A straightforward expectile version of this order has been introduced by Bellini et al. (2018a, p. 859) and is defined in the following.

Definition 3.19. For $F, G \in \mathcal{L}^1$, F is said to precede G in the *expectile order*, denoted by $F \leq_e G$, if $e_F(p) \leq e_G(p)$ holds for all $p \in (0,1)$.

Bellini (2012, p. 2020) proved that the expectile order is weaker than the usual stochastic order by using an order-theoretic comparative static approach. In the following, a rather elementary proof for the same result is given.

Theorem 3.20. *Let $F, G \in \mathcal{L}^1$. Then $F \leq_{st} G$ implies $F \leq_e G$.*

Proof. As introduced in Example 2.21b), let $\pi_X(t) = \mathbb{E}[(X-t)_+]$ and $\pi_Y(t) = E](Y-t)_+]$ denote the stop-loss transforms of X and Y. Then,

$$\lim_{t\to\infty} \pi_X(t) = 0 \quad \text{and} \quad \lim_{t\to-\infty} (\pi_X(t)+t) = \mathbb{E}[X],$$

and analogously for Y (see, e.g., Müller and Stoyan, 2002, p. 20, Thm. 1.5.10). Hence,

$$\lim_{t\to-\infty} (\pi_Y(t) - \pi_X(t)) = \mathbb{E}[Y] - \mathbb{E}[X]. \tag{3.9}$$

Now, assume $X \leq_{st} Y$. This holds, if and only if the function $t \mapsto \pi_Y(t) - \pi_X(t)$ is decreasing (see Müller and Stoyan, 2002, p. 22, Thm. 1.5.13). In particular,

$$\pi_X(t) \leq \pi_Y(t) \quad \forall t \in \mathbb{R}. \tag{3.10}$$

Bellini et al. (2018a, p. 860) proved that

$$\pi_X(t)\,(t - \mathbb{E}[Y]) \leq \pi_Y(t)\,(t - \mathbb{E}[X]) \quad \forall t \in \mathbb{R} \tag{3.11}$$

is a sufficient condition for $X \leq_e Y$. Since $X \leq_{st} Y$, we have $\mathbb{E}[X] \leq \mathbb{E}[Y]$, and (3.11) is obviously satisfied for $\mathbb{E}[X] \leq t \leq \mathbb{E}[Y]$. For $t > \mathbb{E}[Y]$, $t - \mathbb{E}[X] \geq t - \mathbb{E}[Y] > 0$, and (3.11) holds by (3.10). Next, consider the case $t < \mathbb{E}[X]$. Then, (3.11) is equivalent to

$$\pi_X(t)\,(\mathbb{E}[Y] - t) \geq \pi_Y(t)\,(\mathbb{E}[X] - t) \quad \forall t \in \mathbb{R}$$

or

$$\pi_X(t)\,(\mathbb{E}[Y] - \mathbb{E}[X]) \geq (\pi_Y(t) - \pi_X(t))\,(\mathbb{E}[X] - t) \quad \forall t \in \mathbb{R}. \tag{3.12}$$

Since $t \to \pi_Y(t) - \pi_X(t)$ is decreasing, $\pi_Y(t) - \pi_X(t) \leq \mathbb{E}[Y] - \mathbb{E}[X]$ holds by (3.9). On the other hand, Jensen's inequality implies

$$(\mathbb{E}[X] - t) \leq (\mathbb{E}[X] - t)_+ \leq \mathbb{E}[(X - t)_+] = \pi_X(t)$$

and (3.12) follows. Hence, (3.11) is valid for all $t \in \mathbb{R}$, which completes the proof. ▨

Since the stochastic order is the fundamental location order, this result also puts the expectile order in the category of location orders. Furthermore, it implies that the p-expectile satisfies condition (L2) for all $p \in (0,1)$. However, it follows from Proposition 2.22 that (L1) is satisfied, if and only if $p = \frac{1}{2}$. Therefore, the p-expectile is a measure of non-central location for all $p \in (0,1) \setminus \{\frac{1}{2}\}$, just like the corresponding p-quantile. The $\frac{1}{2}$-expectile coincides with the mean and is therefore already known to be a measure of central location. For the sake of consistency, we define $\nu_E(F) = e_F(\frac{1}{2}) = \mu_F, F \in \mathcal{L}^1$.

Further results concerning the relationship between $\leq_{st}$ and $\leq_e$ were proved by Bellini et al. (2018a): a counterexample for the reverse implication of Theorem 3.20 using Lomax distributions is given on page 869, while the two orders are shown to be equivalent for normal distributions on page 861.

The expectile order and the result from Theorem 3.20 can be rewritten in terms of the transformed expectile cdf given in (2.14), which is the inverse function of the expectile function. Namely, $F \leq_e G$ is equivalent to $e_G(\breve{F}(t)) \geq t$ for all $t \in \mathbb{R}$. The function $\breve{R}_{FG} = R_{\breve{F}\breve{G}} = e_G \circ \breve{F}$ may be called the expectile RIDF from F to G. Theorem 3.20 connects the modified expectile RIDF $\breve{\Delta}_{FG} = \breve{R}_{FG} - \mathrm{id}$ to the usual modified RIDF Δ_{FG} by stating that the non-negativity of Δ_{FG} implies the non-negativity of $\breve{\Delta}_{FG}$.

3.2.2. Dispersion

We start this subsection by defining a straightforward definition of an expectile dispersive order. However, since obtaining results on that order is difficult, we also define a suitable weakening along with the same weakening of the traditional dispersive order.

Definition 3.21. a) F is said to precede G in the *weak dispersive order*, denoted by $F \leq_{w-disp} G$, if

$$F^{-1}(p_1) - F^{-1}(p_0) \leq G^{-1}(p_1) - G^{-1}(p_0) \quad \forall\, 0 < p_0 < \tfrac{1}{2} < p_1 < 1.$$

b) For $F, G \in \mathcal{L}^1$, F is said to precede G in the *expectile dispersive order*, denoted by $F \leq_{e-disp} G$, if

$$e_F(p_1) - e_F(p_0) \leq e_G(p_1) - e_G(p_0) \quad \forall\, 0 < p_0 < p_1 < 1.$$

c) For $F, G \in \mathcal{L}^1$, F is said to precede G in the *weak expectile dispersive order*, denoted by $F \leq_{we-disp} G$, if

$$e_F(p_1) - e_F(p_0) \leq e_G(p_1) - e_G(p_0) \quad \forall\, 0 < p_0 < \tfrac{1}{2} < p_1 < 1.$$

The weak dispersive order is known in the literature: $F \leq_1^m G$ was defined in MacGillivray (1986, p. 1002, Def. 2.6) by

$$F^{-1}(p) - F^{-1}(\tfrac{1}{2}) \underset{\leq}{\geq} G^{-1}(p) - G^{-1}(\tfrac{1}{2}) \quad \forall p \underset{\geq}{\leq} \tfrac{1}{2} \tag{3.13}$$

is easily seen to be equivalent to $F \leq_{w-disp} G$. Before that, Bickel and Lehmann (1976, p. 499) defined the same order only for symmetric distributions by $|X - F^{-1}(\frac{1}{2})| \leq_{st} |Y - G^{-1}(\frac{1}{2})|$. If all quantiles are replaced by expectiles and $\leq_{st}$ is replaced by $\leq_e$, both alternative characterizations are also applicable to the weak expectile dispersive order $\leq_{we-disp}$. Note that the notion 'weak dispersive order' is also used for another order in Part II of this thesis (defined on p. 182 without specific notation).

It is obvious that the two implications $F \leq_{disp} G \Rightarrow F \leq_{w-disp} G$ and $F \leq_{e-disp} G \Rightarrow F \leq_{we-disp} G$ hold. Furthermore, a dispersive analogue of Theorem 3.20 would be that the dispersive order implies the expectile dispersive order. While this could not be proved so far, the following example shows that the reverse implication is not true. Contrary to the general setting in Part I of this thesis, the example is constructed using discrete distributions. However, the statement of this example remains valid, if the distributions of both X and Y are sufficiently closely approximated by continuous distributions (e.g. by linear interpolation).

Example 3.22. Let $p_X, p_Y \in (0,1)$, $p_X \neq p_Y$ and $0 < a_X < a_Y$. Furthermore, let $\tilde{X} \sim \mathrm{Bin}(1, p_X), \tilde{Y} \sim \mathrm{Bin}(1, p_Y)$ and $X = a_X \cdot \tilde{X}$ and $Y = a_Y \cdot \tilde{Y}$. It follows directly that F and G, the cdf's of X and Y, are not comparable with respect to $\leq_{disp}$ since $\mathrm{range}(F) \subseteq \mathrm{range}(G)$ is a necessary condition for $F \leq_{disp} G$ (Müller and Stoyan, 2002, p. 41, Thm. 1.7.3). Specifically,

$$F^{-1}(1 - p_X + \varepsilon) - F^{-1}(1 - p_X - \varepsilon) = a_X > 0 = G^{-1}(1 - p_X + \varepsilon) - G^{-1}(1 - p_X - \varepsilon)$$

holds for $\varepsilon > 0$ sufficiently small. Further, a simple calculation yields $\pi_X(t) = p_X(a_X - t)$ for $t \in [0, a_X]$. It follows that

$$\breve{F}(t) = \frac{t(1 - p_X)}{p_X a_X + t(1 - 2p_X)} \quad \text{and} \quad e_F(\alpha) = \frac{\alpha p_X a_X}{(1 - \alpha) + p_X(2\alpha - 1)}$$

for $t \in [0, a_X]$ and $\alpha \in (0,1)$ with analogous results for Y. Overall,

$$\breve{R}_{FG}(t) = (e_G(\breve{F}(t)) = \frac{p_Y(1 - p_X)a_Y t}{p_X(1 - p_Y)a_X + t(p_Y - p_X)}$$

for $t \in [0, a_X]$. Note that $F \leq_{e-disp} G$ is equivalent to $\breve{F} \leq_{disp} \breve{G}$ and that $\breve{F}, \breve{G} \in \mathcal{P}_I^\infty$ holds.

According to Corollary 2.12, $F \leq_{e-disp} G$ is also equivalent to $\breve{\Delta}'_{FG} \geq 0$ or $\breve{R}'_{FG} \geq 1$. Because of $\lim_{p_X \to p_Y} \breve{R}_{FG}(t) = \frac{a_Y}{a_X} t$ for all $t \in [0, a_X]$, $X \leq_{e-disp} Y$ holds if the difference between p_X and p_Y is sufficiently small.

In order to obtain a positive result that connects traditional dispersive orders and their expectile-based counterparts, we move to the weak expectile dispersive order. The proof of the following result uses the idea from the proof of Theorem 14 by Arab et al. (2022, p. 6).

Theorem 3.23. *For $F, G \in \mathcal{L}^1$, $F \leq_{we-disp} G$ is equivalent to $F \leq_{dil} G$.*

Proof. Define $\tilde{X} = X - \mathbb{E}[X], \tilde{Y} = Y - \mathbb{E}[Y]$. Then, $X \leq_{dil} Y$ is equivalent to $\tilde{X} \leq_{cx} \tilde{Y}$, and therefore to $\pi_{\tilde{X}}(t) \leq \pi_{\tilde{Y}}(t)$ for all $t \in \mathbb{R}$. This is equivalent to $\frac{\pi_{\tilde{X}}(t)}{|t|} \leq \frac{\pi_{\tilde{Y}}(t)}{|t|}$ for all $t \neq 0$, and hence to

$$\frac{\pi_{\tilde{X}}(t)}{t} \geq \frac{\pi_{\tilde{Y}}(t)}{t} \; \forall t < 0 \quad \text{and} \quad \frac{\pi_{\tilde{X}}(t)}{t} \leq \frac{\pi_{\tilde{Y}}(t)}{t} \; \forall t > 0. \tag{3.14}$$

Note that, using the properties of the stop-loss transform (Müller and Stoyan, 2002, p. 20, Thm. 1.5.10), $\frac{\pi_{\tilde{X}}(t)}{t} \leq -1$ for $t < 0$, and $\frac{\pi_{\tilde{X}}(t)}{t} \geq 0$ for $t > 0$. Now, applying the transformation $h(x) = (x+1)/(2x+1), x \neq -1/2$, which is decreasing for $x < -1/2$ as well as for $x > -1/2$, to both sides of the inequalities shows that (3.14) is equivalent to

$$\breve{H}_{\tilde{X}}(t) \leq \breve{H}_{\tilde{Y}}(t) \; \forall t < 0 \quad \text{and} \quad \breve{H}_{\tilde{X}}(t) \geq \breve{H}_{\tilde{Y}}(t) \; \forall t > 0. \tag{3.15}$$

In turn, (3.15) is equivalent to

$$e_{\tilde{X}}(p) \geq e_{\tilde{Y}}(p) \; \forall p < \tfrac{1}{2} \quad \text{and} \quad e_{\tilde{X}}(p) \leq e_{\tilde{Y}}(p) \; \forall p > \tfrac{1}{2}.$$

This means that

$$e_X(p) - \mathbb{E}[X] \geq e_Y(p) - \mathbb{E}[Y] \; \forall p < \tfrac{1}{2} \quad \text{and} \quad e_X(p) - \mathbb{E}[X] \leq e_Y(p) - \mathbb{E}[Y] \; \forall p > \tfrac{1}{2},$$

which is equivalent to $X \leq_{we-disp} Y$ via the representation of the weak dispersive order in (3.13). ∎

One of the implications of Theorem 3.23 can also easily be derived from a result by Bellini (2012, p. 2020). His Theorem 3(b) proves that $X \leq_{cx} Y$ implies $e_X(p) \geq e_Y(p)$ for all $p \leq \frac{1}{2}$ and $e_X(p) \leq e_Y(p)$ for all $p \geq \frac{1}{2}$. These statements combined then imply $X \leq_{we-disp} Y$. However, the proof given by Bellini is not as elementary as that of Theorem 3.23.

The equivalence of $\leq_{dil}$ and $\leq_{we-disp}$ is heuristically plausible as both are dispersion orders that are centred around the expectation, which coincided with the $\frac{1}{2}$-expectile. Still, because the two orders are weakenings of the dispersive order and the expectile dispersive order, respectively, Theorem 3.23 bridges the gap between traditional and expectile-based orders of dispersion. Because $\leq_{dil}$ is a weakening of $\leq_{disp}$ (see Proposition 2.20b) and Example 2.21b)), the result also strengthens $\leq_{disp}$ in its role as foundational order of dispersion. Although it is defined using quantiles, it is not fixed on such and also implies expectile-based orders.

Corollary 3.24. *a) For $F, G \in \mathcal{L}^1$, $F \leq_{disp} G$ implies $F \leq_{we-disp} G$.*

b) Let $\alpha \in (0, \frac{1}{2})$. The mapping $\tau_E^\alpha : \mathcal{L}^1 \to [0, \infty), F \mapsto e_F(1-\alpha) - e_F(\alpha)$ is a measure of dispersion.

Proof. a) This follows from Theorem 3.23 and Proposition 2.20b).

b) (D2) follows directly from a) and (D1) follows from Proposition 2.22a), b), e).

∎

Corollary 3.24b) states that the expectile equivalent of the interquantile range is a measure of dispersion. We call τ_E the *interexectile range*. The evaluation points of the expectile function are restricted to symmetric choices for (D1) to be satisfied.

Analogously to how the limiting value of the interquantile range is considered in Remark 3.16a), we can consider the same limit for the interexpectile range. In doing so, we invoke Proposition 2.22f) to obtain

$$\begin{aligned}\lim_{\alpha \nearrow \frac{1}{2}} \frac{1}{1-2\alpha} \tau_E^\alpha(F) &= \lim_{\alpha \nearrow \frac{1}{2}} \frac{e_F(1-\alpha) - e_F(\alpha)}{(1-\alpha) - \alpha} \\ &= (e_F)' \left(\frac{1}{2}\right) = \frac{\mathbb{E}[|X - e_F(\frac{1}{2})|]}{(1-\frac{1}{2})F(e_F(\frac{1}{2})) + \frac{1}{2}(1 - F(e_F(\frac{1}{2})))} \\ &= 2\mathbb{E}[|X - \mu_X|], \end{aligned} \tag{3.16}$$

i.e., double the mean absolute deviation from the mean (MAD). In the following, we disregard the factor two. It inherits properties (D1) and (D2) directly from τ_E^α, which yields the following corollary. Note that $\tilde{\tau}_E^\alpha = \frac{\tau_E^\alpha}{1-2\alpha}$ is also a measure of dispersion for all $\alpha \in (0, \frac{1}{2})$.

Corollary 3.25. *The mapping $\tau_{EL} : \mathcal{L}^1 \to [0, \infty), F \mapsto \mathbb{E}[|X - \mu_X|]$ is a measure of dispersion.*

This result can also be found in the literature, see, e.g., Hürlimann (2002, p. 15). The notation is chosen this way for the sake of consistency throughout Part I, with the index meaning that the measure is obtained as the limiting value of the expectile-based measure. A downside of expectile-based measures is that they are less intuitive and more difficult to interpret that other, more traditional measures like τ_{EL}. For values of α that are sufficiently close to $\frac{1}{2}$, τ_E^α is known to be close to the more accessible measure τ_{EL}. This observation is generalized by the following result, which rescales τ_{EL} to obtain lower and upper boundaries for τ_E^α in dependence of α.

Proposition 3.26. *For all $\alpha \in (0, \frac{1}{2})$,*

$$\frac{1-2\alpha}{1-\alpha}\tau_{EL}(X) < \tau_E^\alpha(X) < \frac{1-2\alpha}{\alpha}\tau_{EL}(X)$$

holds.

Proof. For any $p \in (0,1) \setminus \{\frac{1}{2}\}$, the first order condition (2.13) for the p-expectile can be rewritten as

$$\mathbb{E}[(X - e_X(p))_+] = \frac{1-p}{1-2p}(\mu_X - e_X(p)), \tag{3.17}$$
$$\mathbb{E}[(X - e_X(p))_-] = \frac{p}{1-2p}(\mu_X - e_X(p)). \tag{3.18}$$

The representation

$$\mathbb{E}[(X-t)_+] = \int_t^\infty (1 - F(z))\,\mathrm{d}z$$

shows that $t \mapsto \mathbb{E}[(X-t)_+]$ is strictly decreasing on $\{t \in \mathbb{R} : F(t) < 1\}$. Similarly, $t \mapsto \mathbb{E}[(X-t)_-]$ is strictly increasing on $\{t \in \mathbb{R} : F(t) > 0\}$. Since, by Proposition 2.22c), $e_X(\alpha) < \mu_X < e_X(1-\alpha)$, we obtain

$$\mathbb{E}[(X - e_X(1-\alpha))_+] < \mathbb{E}[(X-\mu_X)_+] < \mathbb{E}[(X - e_X(\alpha))_+], \tag{3.19}$$
$$\mathbb{E}[(X - e_X(\alpha))_-] < \mathbb{E}[(X-\mu_X)_-] < \mathbb{E}[(X - e_X(1-\alpha))_-]. \tag{3.20}$$

Adding the terms in (3.19) and (3.20), and using equations (3.17) and (3.18) then yields

$$\frac{\alpha}{1-2\alpha}\left(e_X(1-\alpha) - e_X(\alpha)\right) < E[|X-\mu_X|] < \frac{1-\alpha}{1-2\alpha}\left(e_X(1-\alpha) - e_X(\alpha)\right),$$

or

$$\frac{1-2\alpha}{1-\alpha}E[|X-\mu_X|] < e_X(1-\alpha) - e_X(\alpha) < \frac{1-2\alpha}{\alpha}E[|X-\mu_X|].$$

∎

Note that the difference between the bounds given in Proposition 3.26 is decreasing in α: for $\alpha = \frac{1}{4}$, we have $\frac{2}{3}\tau_{EL}(X) < \tau_E^\alpha(X) < 2\tau_{EL}(X)$; and for $\alpha \searrow 0$, the lower bound converges to $\tau_{EL}(X)$ and the upper bound diverges. For $\alpha \nearrow \frac{1}{2}$, all three quantities converge to zero. However, if the result is reformulated in terms of $\tilde{\tau}_E^\alpha = \frac{\tau_E^\alpha}{1-2\alpha}$, we obtain

$$\frac{1}{1-\alpha}\tau_{EL}(X) < \tilde{\tau}_E^\alpha(X) < \frac{1}{\alpha}\tau_{EL}(X),$$

and these bounds both converge to $2\tau_{EL}(X)$ for $\alpha \nearrow \frac{1}{2}$.

The interquantile range τ_Q^α is another dispersion measure that is closely related to τ_E^α. A general statement that connects the two measures in a similar way as in Proposition 3.26 could not be obtained here. However, if attention is restricted to symmetric distributions with log-concave densities, Corollary 7 and the preceding statements by Arab et al. (2022, p. 4) directly yield the following result.

Proposition 3.27. *Let $F \in \mathcal{S}$ with a log-concave density, i.e., it satisfies $f(\lambda s + (1-\lambda)t) \geq f(s)^{\lambda} f(t)^{1-\lambda}$ for all $s, t \in \operatorname{supp}(F)$ and all $\lambda \in (0,1)$. Then, $\tau_E^{\alpha}(F) < \tau_Q^{\alpha}(F)$ holds for all $\alpha \in (0, \frac{1}{2})$.*

A well-known example of a symmetric distribution with a log-concave density is any normal distribution. Hence, any interexpectile range of a normal distribution is smaller than the corresponding interquantile range.

3.2.3. Skewness

As for location and dispersion, the most obvious candidate for an expectile-based skewness measure is obtained by replacing all quantiles in the quantile-based measure γ_Q^{α} by expectiles. Thus, for $F \in \mathcal{L}^1$ and $\alpha \in (0, \frac{1}{2})$, we have

$$\tilde{\gamma}_E^{\alpha}(F) = \frac{e_F(1-\alpha) - 2e_F(\frac{1}{2}) + e_F(\alpha)}{e_F(1-\alpha) - e_F(\alpha)} = \frac{e_F(1-\alpha) - 2\mu_F + e_F(\alpha)}{e_F(1-\alpha) - e_F(\alpha)}.$$

The discussion of the properties (S1) and (S2) is postponed to later in this section since it is aided by the definition of expectile-based skewness orders.

Most of the skewness measures considered so far are normalized to the interval $[-1, 1]$. This is also true for $\tilde{\gamma}_E^{\alpha}$ because its numerator is the difference of the two positive quantities $e_F(1-\alpha) - e_F(\frac{1}{2})$ and $e_F(\frac{1}{2}) - e_F(\alpha)$ and its denominator is the sum of the same two quantities. However, these boundaries can be improved and the actual range of $\tilde{\gamma}_E^{\alpha}(F)$ is considerably smaller.

Proposition 3.28. *Let $\alpha \in (0, \frac{1}{2})$. Then, $-1 + 2\alpha < \tilde{\gamma}_E^{\alpha}(F) < 1 - 2\alpha$ holds for all $F \in \mathcal{L}^1$ and both bounds cannot be improved.*

Proof. We reuse the idea from the proof of Proposition 3.26. Plugging equations (3.17) and (3.18) into the outer parts of inequalities (3.19) and (3.20) yields

$$\alpha(e_X(1-\alpha) - \mu_X) < (1-\alpha)(\mu_X - e_X(\alpha)),$$
$$\alpha(\mu_X - e_X(\alpha)) < (1-\alpha)(e_X(1-\alpha) - \mu_X).$$

This can be rearranged into

$$\frac{e_X(1-\alpha) - \mu_X}{\mu_X - e_X(\alpha)} < \frac{1-\alpha}{\alpha},$$
$$\frac{e_X(1-\alpha) - \mu_X}{\mu_X - e_X(\alpha)} > \frac{\alpha}{1-\alpha}.$$

By applying the increasing transformation $(0, \infty) \to (-1, 1),\ t \mapsto (1 + t^{-1})^{-1} - (1+t)^{-1}$ as in the proof of Lemma 3.9, the first inequality yields the upper bound for $\tilde{\gamma}_E^{\alpha}$ and the second yields the lower bound.

Let now $X \sim \text{Bin}(1, p)$ for some $p \in (0, 1)$. Some calculations yield

$$\tilde{\gamma}_E^\alpha(X) = (2\alpha - 1)(2p - 1).$$

Hence, $\tilde{\gamma}_E^\alpha(X) \to 1 - 2\alpha$ for $p \to 0$ as well as $\tilde{\gamma}_E^\alpha(X) \to 2\alpha - 1$ for $p \to 1$. Both inequalities are sharp since this discrete distribution can be approximated arbitrarily closely by distributions in $\mathcal{P}_I^1$. ▨

Based on this result we redefine our expectile-based skewness measure as

$$\gamma_E^\alpha = \frac{1}{1 - 2\alpha} \tilde{\gamma}_E^\alpha, \quad \alpha \in (0, 1/2).$$

Then, $-1 < \gamma_E^\alpha(F) < 1$ holds, and both inequalities are sharp for all $\alpha \in (0, 1/2)$ and $F \in \mathcal{L}^1$. Note that this factor is the same as the one used to obtain the limiting expectile-based dispersion measure τ_{EL} in (3.16). This factor again allows us to determine the limiting value for $\alpha \nearrow \frac{1}{2}$. Assuming $F \in \mathcal{P}_I^1$, Proposition 2.22f) yields that the corresponding expectile function e_F is twice differentiable. Then, we can rewrite γ_E^α as a ratio of first- and second-order difference quotients

$$\gamma_E^\alpha(F) = \frac{1}{2} \frac{\frac{e_F(\frac{1}{2}+\beta) - 2e_F(\frac{1}{2}) + e_F(\frac{1}{2}-\beta)}{\beta^2}}{\frac{e_F(\frac{1}{2}+\beta) - e_F(\frac{1}{2}-\beta)}{\beta}},$$

where $\beta = \frac{1}{2} - \alpha$. Splitting the central difference in the denominator into a forward and a backward difference yields the limit

$$\lim_{\alpha \nearrow \frac{1}{2}} \gamma_E^\alpha(F) = \lim_{\beta \searrow 0} \gamma_E^\alpha(F) = \frac{e_F''(\frac{1}{2})}{4e_F'(\frac{1}{2})}.$$

We have already seen that $e_F'(\frac{1}{2}) = 2\tau_{EL}(F) = 2\mathbb{E}[|X - \mu_X|]$. For the calculation of $e_X''(1/2)$, we denote numerator and denominator of $e_X'(\alpha)$ in Proposition 2.22f) by $u(\alpha)$ and $v(\alpha)$, respectively. Then, $\lim_{\alpha \nearrow \frac{1}{2}} u(\alpha) = \tau_{EL}(F)$ and $\lim_{\alpha \nearrow \frac{1}{2}} v(\alpha) = \frac{1}{2}$ as well as

$$u'(\alpha) = e_F'(\alpha)(2F(e_F(\alpha)) - 1) \to e_F'(1/2)(2F(\mu_F) - 1), \tag{3.21}$$

$$v'(\alpha) = (1 - 2F(e_F(\alpha))) + (1 - 2\alpha) f(e_F(\alpha)) e_F'(\alpha) \to 1 - 2F(\mu_F), \tag{3.22}$$

for $\alpha \nearrow 1/2$. By combining these results, it follows

$$e_X''(1/2) = \lim_{\alpha \nearrow \frac{1}{2}} \frac{u'(\alpha)v(\alpha) - u(\alpha)v'(\alpha)}{(v(\alpha))^2} = 8\tau_{EL}(F)(2F(\mu_F) - 1), \tag{3.23}$$

which overall yields

$$\gamma_{EL}(F) = \lim_{\alpha \nearrow \frac{1}{2}} \gamma_E^\alpha(F) = 2F(\mu_F) - 1.$$

Using γ_{EL} to measure skewness is not a new idea. It has already been introduced as such by Tajuddin (1999) and the quantity $F(\mu)$ is the theoretical counterpart of the test statistic of the sign test for symmetry with estimated center (see Gastwirth, 1971). Similarly to γ_{IQ} in (3.5), the measure γ_{EL} exploits the idea that the difference between mean μ_F and median $F^{-1}(\frac{1}{2})$ indicates the skewness of the underlying distribution. Since a substitution of the mean by the median in $\gamma_{EL}(F)$ always results in the value 0 for $F \in \mathcal{P}_I$, a positive value of $\mu_F - F^{-1}(\frac{1}{2})$ yields a positive value of $\gamma_{EL}(F)$ and thus right-skewness and vice versa. A similar principle is used by the skewness measure γ_{Mode} based on the mode, given in Theorem 3.18.

Since the measure γ_{EL} has a much simpler structure than its 'parent' γ_E^α, it is also much simpler to prove that it fulfils the properties (S1) and (S2).

Theorem 3.29. *The mapping $\gamma_{EL} : \mathcal{L}^1 \to \mathbb{R}, F \mapsto 2F(\mu_F) - 1$ is a measure of skewness.*

Proof. Property (S1) follows directly from Lemma 3.5a) and the linearity of the expectation. For property (S2), let $F, G \in \mathcal{L}^1$ with $F \leq_c G$, so R_{FG} is convex. Then, the following equivalences hold:

$$\gamma_{EL}(F) \leq \gamma_{EL}(G) \Leftrightarrow F(\mu_F) \leq G(\mu_G) \Leftrightarrow R_{FG}(\mathbb{E}[X]) \leq \mathbb{E}[Y].$$

The assertion follows by Jensen's inequality and $Y \overset{\mathcal{D}}{=} G^{-1}(F(X)) = R_{FG}(X)$. ▨

Just like for the expectile-based dispersion measure, we can use the limiting value for $\alpha \nearrow \frac{1}{2}$ as a way to make the expectile-based measure γ_E^α more accessible. For that, consider the following result.

Proposition 3.30. *For $F \in \mathcal{P}_I^3$, $\lim_{\alpha \nearrow \frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}\alpha} \gamma_E^\alpha(F) = 0$ holds.*

Proof. By Proposition 2.22f), it follows from $F \in \mathcal{P}_I^3$ that e_F is four times differentiable. First, we differentiate $\gamma_E^\alpha(F)$ with respect to α, which yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\alpha}\gamma_E^\alpha(F) = {} & 2\frac{e_F'(\alpha)(e_F(1-\alpha) - e_F(\frac{1}{2})) - e_F'(1-\alpha)(e_F(\frac{1}{2}) - e_F(\alpha))}{(1-2\alpha)(e_F(1-\alpha) - e_F(\alpha))^2} \\ & + 2\frac{e_F(1-\alpha) - 2e_F(\frac{1}{2}) + e_F(\alpha)}{(1-2\alpha)^2(e_F(1-\alpha) - e_F(\alpha))}\end{aligned}$$

for $\alpha \in (0, \frac{1}{2})$. Using the notation $\beta = \frac{1}{2} - \alpha$, this can once again be rewritten as a composition of difference quotients. Now, using Taylor expansions for each of them such that the remainders are of order $O(\beta^3)$ in the numerators as well as in the denominators yields after some computations

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\gamma_E^\alpha(F) = \frac{1}{4}\frac{-e_F'(\frac{1}{2})e_F''(\frac{1}{2})\beta + O(\beta^3)}{(e_F'(\frac{1}{2}))^2\beta^2 + O(\beta^3)} + \frac{1}{4}\frac{e_F''(\frac{1}{2}) + \frac{1}{12}e_F^{(4)}(\frac{1}{2})\beta^2 + O(\beta^3)}{e_F'(\frac{1}{2})\beta + O(\beta^3)}$$

$$= \frac{1}{4} \frac{O(\beta^3)}{e_F'(\frac{1}{2})\beta^2 + O(\beta^3)},$$

where we used for the second equality that $e_F'(\frac{1}{2}) > 0$ by Proposition 2.22c). Then, taking the limit $\beta \searrow 0$ yields the asserted result $\lim_{\alpha \nearrow 1/2} \frac{\mathrm{d}}{\mathrm{d}\alpha} \gamma_E^\alpha(F) = \lim_{\beta \searrow 0} \frac{\mathrm{d}}{\mathrm{d}\alpha} \gamma_E^\alpha(F) = 0$. ∎

The fact that $\frac{\mathrm{d}}{\mathrm{d}\alpha}\gamma_E^\alpha(F)$ converges to zero as α tends to $\frac{1}{2}$ means that $\gamma_E^\alpha(F)$ flattens out towards $\gamma_{EL}(F)$. Thus, at least for values of α close to $\frac{1}{2}$, $\gamma_{EL}(F)$ is close to and thereby representative for a range of values of $\gamma_E^\alpha(F)$ without the need of a specific choice of the parameter α.

In order to show that γ_E^α also is a measure of skewness, we define two further orders of skewness to bridge the gap between the expectile-based measure and $\leq_c$. For the remainder of this chapter, $\tilde{F} : \mathbb{R} \to [0,1], t \mapsto F(\tau_{EL}(F) \cdot t + \mu_F)$ denotes the version of F that is standardized with respect to the mean and the MAD, which is denoted by τ_{EL}. $\tilde{G}$ is defined analogously.

Definition 3.31. Let $F, G \in \mathcal{L}^1$.

a) G is *more skewed with respect to the mean and the MAD* than F, denoted by $F \leq_\mu^{MAD} G$, if $S^-(\tilde{F}|_{(-\infty,0)} - \tilde{G}|_{(-\infty,0)}) = 1$, $S^-(\tilde{F}|_{(0,\infty)} - \tilde{G}|_{(0,\infty)}) = 1$ and $\tilde{F}(0) \leq \tilde{G}(0)$.

b) F precedes G in the *s-order*, denoted by $F \leq_s G$, if

$$\int_{-\infty}^t \tilde{F}(x)\,\mathrm{d}x \geq \int_{-\infty}^t \tilde{G}(x)\,\mathrm{d}x \quad \forall t \leq 0,$$

and

$$\int_t^\infty (1 - \tilde{F}(x))\,\mathrm{d}x \leq \int_t^\infty (1 - \tilde{G}(x))\,\mathrm{d}x \quad \forall t \geq 0.$$

The first order $\leq_\mu^{MAD}$ is closely related to the intersection characterization given in Proposition 2.17b) for $k = 2$, only with the MAD instead of the standard deviation. The additional condition that the two intersections are on either side of zero follows from the standardization with the MAD, see the proof of the first implication in Theorem 3.32a).

The s-order $\leq_s$ is not to be confused with the order of kurtosis for symmetric distributions, which is introduced in Definition 4.14a) using the same notation. The order $\leq_s$ as defined in Definition 3.31 can be reformulated in terms of the convex order. For standardized cdf's $\tilde{F}$ and $\tilde{G}$ with $\tilde{X} \sim \tilde{F}$ and $\tilde{Y} \sim \tilde{G}$, $\tilde{X} \leq_{cx} \tilde{Y}$ is equivalent to

$$\int_t^\infty (1 - \tilde{F}(x))\,\mathrm{d}x = \pi_{\tilde{X}}(t) \leq \pi_{\tilde{Y}}(t) = \int_t^\infty (1 - \tilde{G}(x))\,\mathrm{d}x \quad \forall t \in \mathbb{R}.$$

This implies that $F \leq_s G$ is equivalent to

$$\tilde{X}_+ \leq_{cx} \tilde{Y}_+ \quad \text{and} \quad \tilde{Y}_- \leq_{cx} \tilde{X}_-$$

(see Arab et al., 2022, p. 5). Thus, $F \leq_s G$ means that the probability mass on the right side of G is more stretched out than for F and that the probability mass on the left side of F is more stretched out than for G. This intuitively corresponds to G being more skewed to the right than F. A rigorous argument for both $\leq_\mu^{MAD}$ and $\leq_s$ to be considered as skewness orders is given in the first part of the following result.

Theorem 3.32. *Let $F, G \in \mathcal{L}^1$. Then:*

a) $F \leq_c G \Longrightarrow F \leq_\mu^{MAD} G \Longrightarrow F \leq_s G.$

b) $F \leq_s G \Longleftrightarrow \tilde{F} \leq_e \tilde{G}.$

Proof. a) We start with the first implication. Using the representations

$$\mathbb{E}[X] = \int_0^\infty (1 - F(t))\,\mathrm{d}t - \int_{-\infty}^0 F(t)\,\mathrm{d}t$$

and

$$\mathbb{E}[|X|] = \int_0^\infty (1 - F(t) - F(-t))\,\mathrm{d}t$$

we obtain

$$\mathbb{E}[X] - \mathbb{E}[Y] = \int_{-\infty}^\infty (G - F)(t)\,\mathrm{d}t, \tag{3.24}$$

$$\mathbb{E}[|X|] - \mathbb{E}[|Y|] = \int_0^\infty (G - F)(t)\,\mathrm{d}t - \int_{-\infty}^0 (G - F)(t)\,\mathrm{d}t.$$

By applying this to the standardized distributions, we obtain

$$\int_{-\infty}^\infty (\tilde{G} - \tilde{F})(t)\,\mathrm{d}t = 0, \tag{3.25}$$

$$\int_0^\infty (\tilde{G} - \tilde{F})(t)\,\mathrm{d}t = \int_{-\infty}^0 (\tilde{G} - \tilde{F})(t)\,\mathrm{d}t. \tag{3.26}$$

Due to Corollary 2.16, $F \leq_c G$ implies that $\tilde{F}$ and $\tilde{G}$ cross each other at most twice. Since $\mu_{\tilde{F}} = \mu_{\tilde{G}}$, it follows from (3.24) (and is well-known) that $\tilde{F}$ and $\tilde{G}$ are either identical or cross each other at least once.

Now, assume that $\tilde{F} \neq \tilde{G}$. Then, $R_{\tilde{F}\tilde{G}}$ is not linear and Jensen's inequality implies $R_{\tilde{F}\tilde{G}}(0) < 0$, resulting in $\tilde{F}(0) < \tilde{G}(0)$ (see van Zwet, 1964, p. 10).

Assume that the function $\tilde{G} - \tilde{F}$ has exactly one root x_1, where $x_1 \leq 0$. Put $x_0 = -\infty, x_2 = 0, x_3 = \infty$, and

$$A_i = \int_{x_{i-1}}^{x_i} (\tilde{G} - \tilde{F})(t)\,\mathrm{d}t, \quad i = 1, 2, 3.$$

From (3.25) and (3.26), we obtain

$$A_1 + A_2 + A_3 = 0, \qquad A_1 + A_2 = A_3.$$

Hence, $A_3 = 0$, which implies that $\tilde{F}(x) = \tilde{G}(x)$ for $x \geq 0$, a contradiction to $\tilde{F}(0) < \tilde{G}(0)$. Since an analogous reasoning excludes a single root $x_1 > 0$, it follows that $\tilde{F}$ and $\tilde{G}$ cross each other exactly twice, with $\tilde{G} - \tilde{F}$ changing sign from negative to positive to negative, and $(\tilde{G} - \tilde{F})(0) > 0$.

The proof of the second implication is a modification of the proof of Theorem 13.2. by Arab et al. (2022, p. 6). We denote the cdf of $\tilde{X}_+$ and $\tilde{Y}_+$ by $\tilde{F}_+$ and $\tilde{G}_+$, respectively, and use the same notation for the negative parts. Then, $F \leq_\mu^{MAD} G$ implies that $\tilde{G}_+ - \tilde{F}_+$ changes sign exactly once, from positive to negative. Now, using Theorem 3.A.44 by Shaked and Shanthikumar (2006, p. 133), we infer $\tilde{X}_+ \leq_{cx} \tilde{Y}_+$. Analogously, $F \leq_\mu^{MAD} G$ also implies that $\tilde{G}_- - \tilde{F}_-$ changes sign exactly once, from negative to positive. The same result as before then yields $\tilde{Y}_- \leq_{cx} \tilde{X}_-$.

b) See Theorem 14 by Arab et al. (2022, p. 6).

▨

As mentioned in the proof, the implication $F \leq_c G \Rightarrow F \leq_s G$ was already shown by Arab et al. (2022, p. 6). Part b) is proved in a similar way as Theorem 3.23. It states that the location order $\leq_e$ is equivalent to the skewness order $\leq_s$ for suitably standardized distributions. The standardization is crucial to understand that statement. If central location and dispersion are fixed in the sense of a single measure, an increase with respect to $\leq_e$ means that the probability mass on the left side is condensed close to the centre of the distribution and that the probability mass on the right side is stretched out away from the centre. This corresponds to an increase in (right-)skewness.

The approach to measure skewness with $\leq_s$ can also be used to construct another family of expectile-based skewness measures. Theorem 3.32 implies that these new measures as well as the measures γ_E^α are indeed skewness measures according to Definition 3.1.

Theorem 3.33. *a) For all $\alpha \in (0, \frac{1}{2})$, the mapping*

$$\gamma_E^\alpha : \mathcal{L}^1 \to \mathbb{R}, F \mapsto \frac{1}{1-2\alpha} \frac{e_F(1-\alpha) - 2\mu_F + e_F(\alpha)}{e_F(1-\alpha) - e_F(\alpha)}$$

is a measure of skewness.

b) For all $t > 0$, the mapping

$$\gamma_{EA}^t : \mathcal{L}^1 \to \mathbb{R}, F \mapsto \tilde{\gamma}_{EA}^{\tau_{EL}(F)\cdot t}(F),$$

where

$$\tilde{\gamma}^t_{EA} : \mathcal{L}^1 \to \mathbb{R}, F \mapsto \frac{1}{t} \int_{\mu_F - t}^{\mu_F + t} F(x)\,\mathrm{d}x - 1$$

is a measure of skewness.

Proof. a) (S1) follows in the same way as for γ^α_Q because of the parallels between Proposition 2.22a), b), e) and Lemma 3.5d). This can be used in the proof of (S2) because $\gamma^\alpha_E(F) \le \gamma^\alpha_E(G)$ is equivalent to $\gamma^\alpha_E(\tilde{F}) \le \gamma^\alpha_E(\tilde{G})$. By Theorem 3.32, $F \le_c G$ implies $\tilde{F} \le_e \tilde{G}$, i.e. $e_{\tilde{F}}(\alpha) \le e_{\tilde{G}}(\alpha)$ for all $\alpha \in (0,1)$. A straightforward computation shows that $\gamma^\alpha_E(\tilde{F}) \le \gamma^\alpha_E(\tilde{G})$ is equivalent to

$$e_{\tilde{G}}(1-\alpha)e_{\tilde{F}}(\alpha) \le e_{\tilde{F}}(1-\alpha)e_{\tilde{G}}(\alpha) \quad \forall \alpha \in (0, \tfrac{1}{2})$$

(by using the increasing transformation from the proof of Lemma 3.9 and noting $\mu_{\tilde{F}} = \mu_{\tilde{G}} = 0$). This, in turn, is equivalent to

$$e_{\tilde{G}}(1-\alpha)\,|e_{\tilde{F}}(\alpha)| \ge e_{\tilde{F}}(1-\alpha)\,|e_{\tilde{G}}(\alpha)| \quad \forall \alpha \in (0, \tfrac{1}{2}). \tag{3.27}$$

Because of $e_{\tilde{G}}(1-\alpha) \ge e_{\tilde{F}}(1-\alpha)$ and $|e_{\tilde{F}}(\alpha)| \ge |e_{\tilde{G}}(\alpha)|$ for $\alpha \in (0, \frac{1}{2})$, inequality (3.27) holds.

b) For (S1), let $a > 0, b \in \mathbb{R}$. Then,

$$\begin{aligned}\tilde{\gamma}^t_{EA}(aX+b) &= \frac{1}{t} \int_{a\mu_X + b - t}^{a\mu_X + b + t} F\left(\frac{x-b}{a}\right) \mathrm{d}x - 1 \\ &= \frac{a}{t} \int_{\mu_X - \frac{t}{a}}^{\mu_X + \frac{t}{a}} F(x)\,\mathrm{d}x - 1 = \tilde{\gamma}^{t/a}_{EA}(X).\end{aligned}$$

Now we can assume without restriction that $\mu_X = 0$ and we obtain

$$\tilde{\gamma}^t_{EA}(-X) = \frac{1}{t} \int_{-t}^{t} 1 - F(x)\,\mathrm{d}x - 1 = 1 - \frac{1}{t} \int_{-t}^{t} F(x)\,\mathrm{d}x = -\tilde{\gamma}^t_{EA}(X).$$

Thus, for $a, b \in \mathbb{R}$,

$$\gamma^t_{EA}(aX+b) = \tilde{\gamma}^{\tau_{EL}(aX+b)\cdot t}_{EA}(aX+b) = \operatorname{sgn}(a) \cdot \tilde{\gamma}^{|a|\tau_{EL}(X)\cdot t/|a|}_{EA}(X) = \operatorname{sgn}(a) \cdot \gamma^t_{EA}(X).$$

For (S2), assume $F \le_c G$, which implies $F \le_s G$. Because (S1) holds, we can consider $\tilde{F}$ and $\tilde{G}$ instead of F and G. By definition, $\tilde{F} \le_s \tilde{G}$ entails

$$\int_{-\infty}^{t} (\tilde{F} - \tilde{G})(x)\,\mathrm{d}x \ge 0 \;\forall t \le 0 \quad \text{and} \quad \int_{t}^{\infty} (\tilde{F} - \tilde{G})(x)\,\mathrm{d}x \ge 0 \;\forall t \ge 0.$$

Thus,

$$\int_{-t}^{t} (\tilde{F} - \tilde{G})(x)\,\mathrm{d}x \leq 0 \quad \forall t > 0,$$

which is equivalent to $\gamma_{EA}^t(F) \leq \gamma_{EA}^t(G)$ for all $t > 0$.

▨

It is easy to see that both $\tilde{\gamma}_{EA}^t$ and γ_{EA}^t are standardized to the interval $[-1, 1]$ because $\int_{\mu_F - t}^{\mu_F + t} F(x)\,\mathrm{d}x \in [0, 2t]$. γ_{EA}^t measures skewness in a similar way to how it is quantified by the order $\leq_s$, which is obvious from the proof of (S2) in Theorem 3.33b). A skewness measure of a similar structure has been proposed in the literature by Arnold and Groeneveld (1993, p. 19), who considered

$$\int_0^\alpha F^{-1}(\tfrac{1}{2} + u) + F^{-1}(\tfrac{1}{2} - u)\,\mathrm{d}u$$

for $\alpha \in (0, \frac{1}{2})$. While this quantity is centred around the median, γ_{EA}^t is centred around the mean.

It can be shown that γ_{EA}^t and γ_{EL} belong to a common family of expectile-based skewness measures. To see this, let $U_t \sim \mathcal{U}(\mu_F - t \cdot \tau_{EL}(F), \mu_F - t \cdot \tau_{EL}(F))$ and note that

$$\gamma_{EA}^t(F) = 2\mathbb{E}[F(U_t)] - 1. \tag{3.28}$$

For $t \searrow 0$, the entire probability mass of U_t is concentrated in μ_F and we obtain $U_0 = \mu_F$ almost surely. The dominated convergence theorem implies

$$\lim_{t \searrow 0} \gamma_{EA}^t(F) = \gamma_{EL}(F)$$

for all $F \in \mathcal{L}^1$. The random variable U_t in (3.28) can also reasonably be replaced by any other random variable that has a unimodal density and is symmetric around μ_F, e.g. a normal distribution with mean μ_F.

Finally, we examine the connection between γ_E^α and γ_{EA}^t more closely. Both represent an entire family of skewness measures and the former is highly similar to γ_Q^α. Thus, the question arises whether these families characterize symmetry as the quantile-based measures are shown to in Proposition 3.11. This is answered by the following result.

Theorem 3.34. *Let $F \in \mathcal{L}^1$.*

a) $\gamma_{EA}^t(F) \gtreqless 0$ *holds for all* $t > 0$*, if and only if* $\gamma_E^\alpha(F) \gtreqless 0$ *holds for all* $\alpha \in (0, \frac{1}{2})$.

b) The following three statements are all equivalent:

(i) $F \in \mathcal{S}$,

(ii) $\gamma_E^\alpha(F) = 0$ *for all* $\alpha \in (0, \frac{1}{2})$,

(iii) $\gamma_{EA}^t(F) = 0$ *for all* $t > 0$.

Proof. a) We only prove the equivalence between the '$\geq$'-inequalities; the reasoning for the reverse inequalities is entirely analogous.

In order to connect the two families of skewness measures, we use the so-called Omega ratio, which has been introduced by Keating and Shadwick (2002) as

$$\Omega_X(t) = \frac{\mathbb{E}[(X-t)_+]}{\mathbb{E}[(X-t)_-]}, \quad t \in \mathbb{R}.$$

Then, the first order condition (2.13), which identifies the p-expectile uniquely, can be written as

$$\Omega_X(e_X(p)) = \frac{1-p}{p}, \quad p \in (0,1). \tag{3.29}$$

This gives the following one-to-one relation between expectiles and Omega ratios:

$$e_X(p) = \Omega_X^{-1}\left(\frac{1-p}{p}\right) \ \forall p \in (0,1), \quad \Omega_X(t) = \frac{1-e_X^{-1}(t)}{e_X^{-1}(t)} \ \forall t \in \mathbb{R}$$

(see Rémillard, 2013, pp. 128–129).

$\gamma_E^\alpha(F) \geq 0$ for all $\alpha \in (0, \frac{1}{2})$ is obviously equivalent to

$$e_X(1-\alpha) - \mu_X \geq \mu_X - e_X(\alpha) \quad \forall \alpha \in (0, \tfrac{1}{2}),$$

which, in turn, is equivalent to

$$e_{X-\mu_X}(\alpha) \geq e_{-(X-\mu_X)}(\alpha) \quad \forall \alpha \in (\tfrac{1}{2}, 1)$$

(see Proposition 2.22a), e)). For $\alpha \in (\frac{1}{2}, 1)$, choose $\beta = \frac{1-\alpha}{\alpha} \in (0,1)$. Using (3.29), the condition $e_{-(X-\mu_X)}(\alpha) \leq e_{X-\mu_X}(\alpha)$ is equivalent to

$$\Omega_{-(X-\mu_X)}(x) = \beta, \, \Omega_{X-\mu_X}(y) = \beta \ \Rightarrow \ x \leq y. \tag{3.30}$$

Since $\Omega_{-(X-\mu_X)}(0) = \Omega_{X-\mu_X}(0) = 1$ and since $\Omega_{-(X-\mu_X)}$ and $\Omega_{X-\mu_X}$ are strictly decreasing, (3.30) holds for all $\beta \in (0,1)$, if and only if

$$\Omega_{-(X-\mu_X)}(t) \leq \Omega_{X-\mu_X}(t) \quad \forall t > 0. \tag{3.31}$$

Using $\Omega_{X-\mu_X}(t) = \Omega_X(\mu_X + t)$ and $\Omega_{-(X-\mu_X)}(t) = \frac{1}{\Omega_X(\mu_X - t)}$, we finally obtain that (3.31), and, hence, the initial statement that $\gamma_E^\alpha(F) \geq 0$ for all $\alpha \in (0, \frac{1}{2})$, is equivalent to

$$\Omega_X(\mu_X + t) \cdot \Omega_X(\mu_X - t) \geq 1 \quad \forall t > 0. \tag{3.32}$$

The Omega ratio, in turn, is closely related to the stop-loss transform: from

$$\mathbb{E}[(X-t)_-] = t - \mathbb{E}[X] + \mathbb{E}[(X-t)_+]$$

we immediately get

$$\Omega_X(t) = \frac{\pi_X(t)}{t - \mu_X + \pi_X(t)}. \tag{3.33}$$

Plugging (3.33) into condition (3.32) gives

$$\pi_X(\mu_X + t)\cdot \pi_X(\mu_X - t) \geq (t + \pi_X(\mu_X + t))\cdot(-t + \pi_X(\mu_X - t)) \quad \forall t > 0,$$

which is equivalent to

$$\tfrac{1}{t}(\pi_X(\mu_X + t) - \pi_X(\mu_X - t)) + 1 \geq 0 \quad \forall t > 0. \tag{3.34}$$

It remains to be shown that the left side of inequality (3.34) is equal to $\tilde{\gamma}^t_{EA}(X)$ because $\tilde{\gamma}^t_{EA}(X) \geq 0$ for all $t > 0$ is equivalent to $\gamma^t_{EA}(X) \geq 0$ for all $t > 0$. For this, let $t > 0$ and note that

$$\begin{aligned}
\tilde{\gamma}^t_{EA}(X) &= \frac{1}{t}\int_{\mu_X - t}^{\mu_X + t} F(x)\,\mathrm{d}x - 1 \\
&= 1 - \frac{1}{t}\int_{\mu_X - t}^{\mu_X + t} (1 - F(x))\,\mathrm{d}x \\
&= \frac{1}{t}\left(\int_{\mu_X + t}^{\infty} (1 - F(x))\,\mathrm{d}x - \int_{\mu_X - t}^{\infty} (1 - F(x))\,\mathrm{d}x\right) + 1.
\end{aligned}$$

b) The equivalence between (ii) and (iii) follows by combining the two statements from part a). For the equivalence between (i) and (ii), note that $F \in \mathcal{S}$ is equivalent to $X - \mu_X \stackrel{\mathcal{D}}{=} \mu_X - X$ because the centre of a symmetric distribution coincides with its mean (see Proposition 3.4a) and Theorem 3.6a)). Note that a cdf $F \in \mathcal{L}^1$ is uniquely determined by its expected value and its stop-loss-transform (see Müller and Stoyan, 2002, p. 20). Thus, F is uniquely determined by its expectile function, which implies that $F \in \mathcal{S}$ is equivalent to $e_{X-\mu_X}(p) = e_{\mu_X - X}(p)$ for all $p \in (0,1)$. Using Proposition 2.22a), e), this is equivalent to

$$e_X(1-\alpha) - \mu_X = \mu_X - e_X(\alpha) \quad \forall \alpha \in (0, \tfrac{1}{2}),$$

which, in turn, is equivalent to (ii).

Part b) of this result states that the two families $\{\gamma^\alpha_E : \alpha \in (0,\frac{1}{2})\}$ and $\{\gamma^t_{EA} : t > 0\}$ of skewness measures both characterize symmetry. Part a) states that the two families additionally interpret skewness both to the left and to the right in the same way.

Chapter 4

Quantifying Kurtosis Irrespective of Symmetry

4.1. What is Kurtosis?: Existing Literature and Preliminary Remarks

There has been much discussion in the literature concerning the question of what kurtosis describes exactly. In particular, a number of articles have been published both advocating its interpretation as 'peakedness' of a distributions and opposing it. See Crack (2019) and Westfall (2014) for examples of either position and Fiori and Zenga (2009) for a more neutral historical review. The heuristic notion of peakedness (as opposed to flat-toppedness) describes how sharp the peak of the density of a unimodal distribution is.

Balanda and MacGillivray (1988, p. 116) provide a critical review of the literature concerning kurtosis and, based on that, aptly describe an increase in kurtosis as 'the location- and scale-free movement of probability mass from the shoulders of a distribution into its center and tails'. This heuristic, as is usually the case for kurtosis, is applied solely to unimodal symmetric distributions. In accordance with this observation, many of the publications, which are critical of the association of kurtosis with peakedness, justify this by stating that kurtosis is not only characterized by peakedness but also by fat tails.

While the author generally concurs with the heuristic of kurtosis given by Balanda and MacGillivray (1988), there is a major issue in the literature concerning kurtosis that needs to be addressed: the notion of kurtosis is almost exclusively limited to symmetric distributions. Most publications about kurtosis just assume the considered distributions to be symmetric without giving any justification (see, e.g., van Zwet, 1964, Oja, 1981 or Groeneveld and Meeden, 1984). A number of articles also claim that 'kurtosis is essentially a property of symmetric distributions' (Tracy and Doane, 2005, p. 272) or nearly symmetric distributions (see McAlevey and Stent, 2018, p. 122 or Crack, 2019, p. 64). However, all of the latter cited articles do so by citing either Pearson (1902, p. 275) or MacGillivray and Balanda (1988), both of which merely state that certain interpretations of kurtosis are simpler if symmetric distributions are used. Overall, there seems to be no convincing reason to only apply the notion of kurtosis to symmetric distributions. Instead, it seems to be rather difficult to untangle the notions of skewness and kurtosis in asymmetric distributions (see also Balanda and MacGillivray, 1990, p. 20). If only symmetric distributions are considered, their skewness is fixed and their location and spread can be standardized, leaving kurtosis as their major distinguishing characteristic. A prime example of this would be Students t-distribution for varying degrees of freedom with the normal distribution being the lower limit in terms of kurtosis.

Since shape categorization, which is often done using the skewness-kurtosis-plane, considers combinations of skewness and kurtosis, the notion of kurtosis needs to be applied to all distributions, irrespective of symmetry. It is, however, not clear, how the notions of skewness and kurtosis can be untangled in order to consider 'pure' kurtosis. Contrary to (central) location and dispersion, distributions cannot simply be standardized with respect to skewness. Two major attempts were made in the literature to define a kurtosis order for asymmetric distributions, one of which was made by MacGillivray and Balanda (1988) by coining the term anti-skewness. For cdf's F and G, they define $F \leq_a G$ to hold, if R_{FG} is concave on $[\inf(\text{supp}(F)), F^{-1}(\frac{1}{2})]$ and convex on $[F^{-1}(\frac{1}{2}), \sup(\text{supp}(F))]$. Although the a in the index of the order stands for anti-skewness, $\leq_a$ is defined as a generalization of the kurtosis order $\leq_s$, defined by van Zwet (1964) and reused, among others, by Oja (1981). The order $\leq_s$ was only ever defined as a kurtosis order for symmetric distributions and, on this restricted class of distributions, its definition coincides with that of $\leq_a$. In most publications on kurtosis, $\leq_s$ is the strongest considered order. One major problem with $\leq_a$ is that it artificially centres the kurtosis comparison of the two considered distributions around their medians. While this is appropriate for symmetric distributions as well as pairs of distributions, the asymmetries of which cancel each other out in a certain way, it is too restrictive in general. This shortcoming is made more obvious in Section 4.2.2, where $\leq_a$ is compared to other kurtosis orders that are more adaptive to asymmetries. In a nutshell, $\leq_a$ seeks to welcome asymmetric distributions into the quantification of kurtosis without opening up its framework for subsequent irregularities.

The other major approach to define a kurtosis order for asymmetric distributions utilizes the so-called spread function. This approach, along with a number of weaker orders and further references, is discussed in detail by Balanda and MacGillivray (1990). For a cdf F, they define the spread function by

$$S_F(\alpha) = F^{-1}(\tfrac{1}{2} + \alpha) - F^{-1}(\tfrac{1}{2} - \alpha), \quad \alpha \in [0, \tfrac{1}{2}),$$

which can be interpreted as one half of a symmetrized quantile function of F. Heuristically, the distribution is again artificially centred around the median by folding it around the median and averaging out the two overlaying halves of the distributions. If the resulting half of a distribution is then mirrored at the median, a symmetric distribution is obtained, which can be ordered with respect to kurtosis using $\leq_s$. This methodology is equivalent to defining the 'symmetrized' kurtosis order $\leq_S$ by

$$F \leq_S G \quad \Leftrightarrow \quad S_G \circ S_F^{-1} \text{ is convex.}$$

This definition of a kurtosis order is fairly easy to use and theoretically applicable to all univariate distributions. It does, however, have significant downsides, especially if it is intended to be used as a foundational order that establishes what is meant by the notion of kurtosis. This was in part noted by Balanda and MacGillivray (1990, p. 29) themselves. First, a significant amount of information is lost in just combining the two 'sides' (with respect to the median) of the distribution. Contrary to the strongest and most basic orders for the lower order characteristics of location, dispersion and skewness, the order $\leq_S$ does not compare two cdf's in a pointwise manner. These basic orders require the function Δ_{FG} to be 0-convex, 1-convex or 2-convex on its entire domain. In contrast, the order $\leq_S$ theoretically allows arbitrarily large deviations from the desired concavity or convexity on one side, if they are compensated by the other side. This kind of behaviour is not desirable for these basic orders. The second downside becomes apparent, if distributions are considered that are not supported by the entire real numbers and that are significantly skewed. In that case, e.g. for the exponential distribution, the support ends close to the median on one side and might even be infinite on the other side. The idea of folding this kind of distribution around the median does not make much sense and the symmetrized version is not representative of the original distribution.

Overall, both major approaches to define a kurtosis order for asymmetric distributions in the literature are unsatisfactory in general. Both of them artificially centre the comparison with respect to kurtosis around the medians of the considered distributions. One could of course modify these orders to be centred around any other measure of central location, but this does not solve the underlying problem.

To the author, the most intuitive approach is to follow the pattern that can be observed for the basic orders of location, dispersion and skewness. This pattern was laid out in great detail by Oja (1981). There, the strongest and therefore most foundational order is always the order of the corresponding convex characteristic as it is generally defined in Definition 2.8b). Since Oja required all distributions to be sufficiently regular, he equivalently defined the order of the k-th convex characteristic, $k \in \mathbb{N}_0$, via the k-convexity of the modified RIDF Δ_{FG} (see Proposition 2.9). A hierarchical structure of the characteristics is represented by the order of the convexity of said function. With location, dispersion and skewness, the first three of these convex characteristics are discussed in Chapter 3. In keeping with the work of Oja (1981), the fundamental orders used there are $\leq_0$ (which is equivalent to $\leq_{st}$), $\leq_1$ (which is

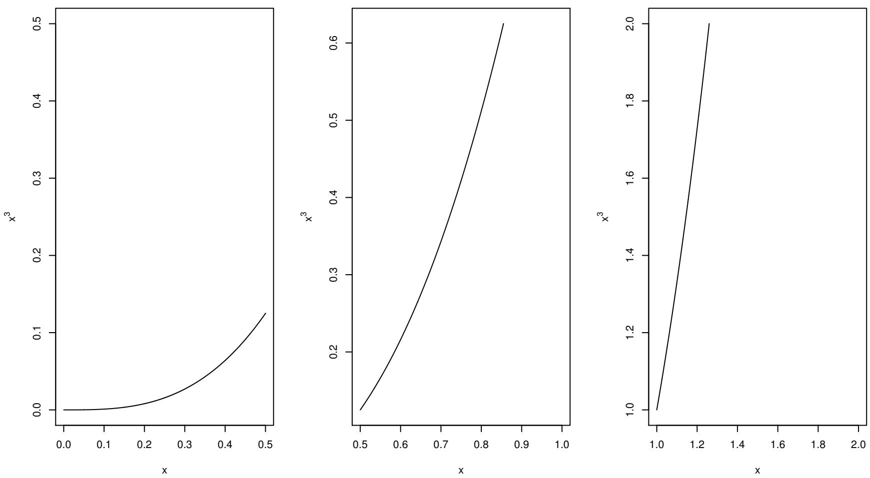

Figure 4.1.: Graph of the function $\phi(x) = x^3$ for three different ranges of x-values on a linear scale to illustrate the perceived changes in curvature as opposed to the constant theoretical change in curvature.

equivalent to $\leq_{disp}$) and $\leq_2$ (which is equivalent to $\leq_c$). This makes $\leq_3$, which is characterized by the 3-convexity of the modified RIDF Δ_{FG}, the canonical choice for the foundational kurtosis order. This order has the additional advantage of being naturally applicable to all distributions, including asymmetric ones. Oja (1981), however, instead restricts his attention to symmetric distributions and defined the previously mentioned $\leq_s$ as his foundational kurtosis order. The order $\leq_3$ is only briefly mentioned in the closing remarks of the paper with lack of transitivity stated as its only major drawback.

While pointing out the fact that $\leq_3$ is not generally transitive as its most notable deficiency is a flawed argument (as shown in Sections 4.2.1 and 4.2.2), the order does have downsides. For the orders of the three previous convex characteristics, it is sufficient to draw the Q-Q-plot of two given distributions to see whether they are ordered with respect to the corresponding characteristic. This is simply due to the fact that non-negativity, monotonicity and convexity of a function are all properties that are easy to detect graphically. However, third order convexity is a considerably less intuitive property. While it is easy to see whether a function has positive or negative curvature, it is a lot more difficult to quantify curvature and therefore also to grasp gradients in curvature. As an example, consider the function $\phi(x) = x^3$ with constant third derivative and, therefore, with constant increase in curvature. If the scale is kept linear, the increase in curvature is obvious if the gradient of the function is small, but almost non-detectable if the gradient of the function is higher (see Figure 4.1). Hence, the visually most striking feature of a prototypical 3-convex function is that it also is concave-convex.

Overall, there are a number of arguments for using $\leq_3$ as a basis of the notion of kurtosis for symmetric and asymmetric distributions. Since the concave-convex order $\leq_s$ is standardly used in the literature, we analyze both orders in more detail in Section 4.2. A particular focus is put on their transitivity, especially for $\leq_3$. For $\leq_s$, a more suitable generalization to asymmetric distributions than $\leq_a$ is proposed. Subsequently, a number of different approaches to measuring kurtosis are presented and their compatibility with the kurtosis orders is discussed.

Throughout Chapter 4, we assume all cdf's to be three times differentiable and to have interval support, i.e. we assume them to fulfil $F \in \mathcal{P}_I^3$. Note that this assumption is not necessary in order to define the kurtosis orders we consider in the following. However, the first three derivatives of the involved cdf's are needed to utilize the characterization of $\leq_3$ given in Corollary 2.12 for $k = 3$, which is essential for numerous results. Furthermore, we assume the densities of all cdf's to be strictly positive on the interiors of their supports because they appear in denominators multiple times throughout the chapter.

4.2. Kurtosis Orders on Asymmetric Distributions

4.2.1. The Order of the Third Convex Characteristic

The order of the third convex characteristic $\leq_3$ is formally defined in Definition 2.8b). We start out by listing a number of equivalent characterizations from Section 2.2 for easier reference throughout this chapter.

Corollary 4.1. *Let $F, G \in \mathcal{P}_I^3$. Then, all of the following statements are equivalent:*

(i) $F \leq_3 G$,

(ii) $\det\left(\tilde{\Xi}^3_{F^{-1},G^{-1}}(p_0, p_1, p_2, p_3)\right) \geq 0$ *for all* $0 < p_0 < p_1 < p_2 < p_3 < 1$,

(iii) the function Δ_{FG} is 3-convex,

(iv) the function R_{FG} is 3-convex,

(v) $\Delta'''_{FG}(t) \geq 0$ *for all* $t \in D_F$,

(vi) $R'''_{FG}(t) \geq 0$ *for all* $t \in D_F$.

Characterizations (ii), (iii) and (iv) are, by definition, based on the determinants of matrices and are therefore all fairly similar. Condition (iv) can be written in the most compact way and it is fulfilled, if $\det\left(\Xi^3_{R_{FG}}(t_0, t_1, t_2, t_3)\right) \geq 0$ holds for all $t_0, t_1, t_2, t_3 \in D_F$ with $t_0 < t_1 < t_2 < t_3$, where

$$\Xi^3_{R_{FG}}(t_0, t_1, t_2, t_3) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ t_0 & t_1 & t_2 & t_3 \\ t_0^2 & t_1^2 & t_2^2 & t_3^2 \\ R_{FG}(t_0) & R_{FG}(t_1) & R_{FG}(t_2) & R_{FG}(t_3) \end{pmatrix}.$$

This inequality is equivalent to

$$
\begin{aligned}
&- R_{FG}(t_0)(t_2-t_1)(t_3-t_1)(t_3-t_2) + R_{FG}(t_1)(t_2-t_0)(t_3-t_0)(t_3-t_2)\\
&- R_{FG}(t_2)(t_1-t_0)(t_3-t_0)(t_3-t_1) + R_{FG}(t_3)(t_1-t_0)(t_2-t_0)(t_2-t_1) \geq 0.
\end{aligned}
$$

Putting $F(t_0)=p_0, F(t_1)=p_1, F(t_2)=p_2$ and $F(t_3)=p_3$ shows that

$$
\begin{aligned}
&- G^{-1}(p_0)\Big(F^{-1}(p_2)-F^{-1}(p_1)\Big)\Big(F^{-1}(p_3)-F^{-1}(p_1)\Big)\Big(F^{-1}(p_3)-F^{-1}(p_2)\Big)\\
&+ G^{-1}(p_1)\Big(F^{-1}(p_2)-F^{-1}(p_0)\Big)\Big(F^{-1}(p_3)-F^{-1}(p_0)\Big)\Big(F^{-1}(p_3)-F^{-1}(p_2)\Big)\\
&- G^{-1}(p_2)\Big(F^{-1}(p_1)-F^{-1}(p_0)\Big)\Big(F^{-1}(p_3)-F^{-1}(p_0)\Big)\Big(F^{-1}(p_3)-F^{-1}(p_1)\Big)\\
&+ G^{-1}(p_3)\Big(F^{-1}(p_1)-F^{-1}(p_0)\Big)\Big(F^{-1}(p_2)-F^{-1}(p_0)\Big)\Big(F^{-1}(p_2)-F^{-1}(p_1)\Big) \geq 0 \qquad (4.1)
\end{aligned}
$$

for all $0 < p_0 < p_1 < p_2 < p_3 < 1$ is equivalent to $F \leq_3 G$. Inequality (4.1) gives the impression that it can not be rewritten in a symmetric way, involving only G^{-1} and F^{-1} on either side of the inequality. If this is true, then there does not exist a family of scalar measures that characterizes the order $\leq_3$. To proof this claim, we first give a simple example showing that the ordering $\leq_3$ is generally not transitive, a fact that was already mentioned by Oja (1981, p. 168). For the example and the subsequent considerations, we mainly use the derivative-based characterization of $\leq_3$ given in Corollary 4.1(vi). It is easier to use than the result (4.1) of the matrix-based characterizations, which are considered more in the context of quantile-based kurtosis measures in Section 4.3.2.

Example 4.2. Define by

$$
\begin{aligned}
F &: [0,1] \to [0,1], \quad t \mapsto t^3,\\
G &: [0,1] \to [0,1], \quad t \mapsto t,\\
H &: [0,1] \to [0,1], \quad t \mapsto 1-\sqrt[3]{1-t}
\end{aligned}
$$

three infinitely often differentiable cdf's on the unit interval. Considering $H^{-1}(t) = (t-1)^3 + 1,\ t \in [0,1]$, we obtain the relative inverse distribution functions

$$
\begin{aligned}
R_{FG} &: [0,1] \to [0,1], \quad t \mapsto G^{-1}(F(t)) = F(t) = t^3,\\
R_{GH} &: [0,1] \to [0,1], \quad t \mapsto H^{-1}(G(t)) = H^{-1}(t) = (t-1)^3+1,\\
R_{FH} &: [0,1] \to [0,1], \quad t \mapsto H^{-1}(F(t)) = H^{-1}(t^3) = (t^3-1)^3+1
\end{aligned}
$$

(see Figure 4.2). Furthermore,

$$
\begin{aligned}
R'_{FG}(t) &= 3t^2 &, \quad R''_{FG}(t) &= 6t &, \quad R'''_{FG}(t) &= 6 \geq 0 \quad \forall t \in [0,1],\\
R'_{GH}(t) &= 3(t-1)^2, & R''_{GH}(t) &= 6(t-1), & R'''_{GH}(t) &= 6 \geq 0 \quad \forall t \in [0,1]
\end{aligned}
$$

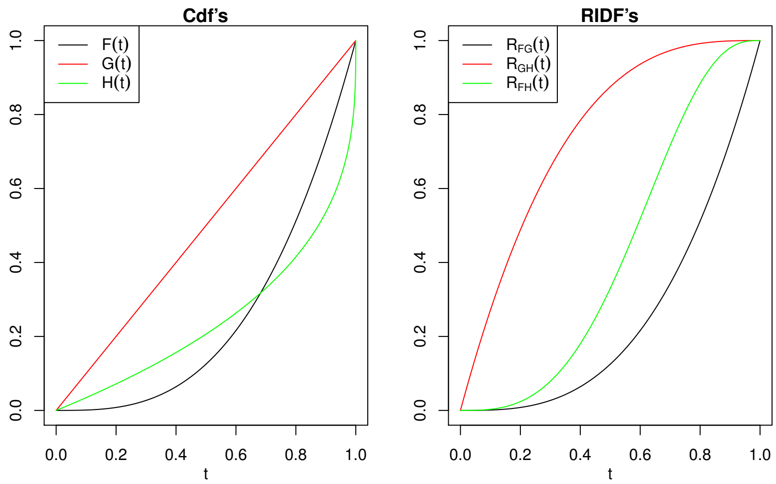

Figure 4.2.: Graphs of the three cdf's (left panel) and the three RIDF's from Example 4.2.

yields $F \leq_3 G$ as well as $G \leq_3 H$. However, since

$$\begin{aligned} R'_{FH}(t) &= 3(t^3-1)^2 \cdot 3t^2 = 9(t^8 - 2t^5 + t^2), \\ R''_{FH}(t) &= 18(4t^7 - 5t^4 + t), \\ R'''_{FH}(t) &= 18(28t^6 - 20t^3 + 1), \end{aligned}$$

substituting t^3 and applying the quadratic formula yields that

$$R'''_{FH}(t) < 0 \text{ for } t \in \left(\sqrt[3]{\tfrac{5-3\sqrt{2}}{14}}, \sqrt[3]{\tfrac{5+3\sqrt{2}}{14}} \right) \approx (0.378, 0.871) \subseteq [0,1].$$

This contradicts $F \leq_3 H$ and, thus, the transitivity of $\leq_3$.

This negative result yields directly the following corollary, since an order based on the comparison between $\kappa(F)$ and $\kappa(G)$, where $\kappa : \mathcal{P}_I^3 \to \mathbb{R}$ is a mapping and potentially a kurtosis measure, would always be transitive.

Corollary 4.3. *There does not exist a family* $\{\kappa_\iota : \mathcal{P}_I^3 \to \mathbb{R} \,|\, \iota \in I\}$ *of mappings such that*

$$\kappa_\iota(F) \leq \kappa_\iota(G) \quad \forall \iota \in I$$

is equivalent to $F \leq_3 G$.

Note that for the orders $\leq_0$, $\leq_1$ and $\leq_2$, there exist families of mappings that characterize the order in the way stated above, e.g. the corresponding density-based measures (see Corollary 3.14). This is just one example that underlines the importance of transitivity as a property of a fundamental order of a distributional characteristic. For that reason and since lack of transitivity has been identified as a major drawback of $\leq_3$ by Oja (1981), we examine it more closely in the following. Specifically, we look for conditions or sets of distributions that allow transitivity of $\leq_3$. As a starting point, all pairs of distributions that are ordered with respect to $\leq_3$ are divided into two mutually exclusive categories.

Remark 4.4. Let $F \leq_3 G$. Since $R'''_{FG}(t) \geq 0$ for all $t \in D_F$, R''_{FG} is non-decreasing. Now, exactly one of the following statements holds:

(i) F and G are skewness-comparable with respect to $\leq_2$, i.e., $F \leq_2 G$ or $G \leq_2 F$,

(ii) F and G are not skewness comparable. In that case, R_{FG} has an inflection point at a $t_{FG} \in D_F = \text{int}(\text{supp}(F))$ with $R''_{FG}(t) \leq 0$ $\forall t \leq t_{FG}$ and $R''_{FG}(t) \geq 0$ $\forall t \geq t_{FG}$. More specifically, there exist values $t_\ell, t_r \in D_F$ with $t_\ell < t_{FG} < t_r$ such that $R''_{FG}(t_\ell) < 0$ and $R''_{FG}(t_r) > 0$.

The inflection point at t_{FG} in (ii) is, in general, not unique since R_{FG} can be linear on a given non-degenerate interval. However, any inflection point of R_{FG} can be uniquely identified by the value $p_{FG} = F(t_{FG}) \in (0, 1)$.

Note that (i) can be viewed as a limiting case of (ii) with $t_{FG} = \inf D_F$ or $t_{FG} = \sup D_F$, yielding $p_{FG} = 0$ or $p_{FG} = 1$, respectively. So in order to obtain the most general setting, we allow $t_{FG} \in \overline{D_F} = \text{supp}(F)$.

Definition 4.5. Let F and G be two cdf's satisfying $F \leq_3 G$. A value $p_{FG} \in [0, 1]$ is said to be an *inflection value of F and G*, if $R''_{FG}(t) \leq 0$ for all $t \leq F^{-1}(p_{FG})$ and $R''_{FG}(t) \geq 0$ for all $t \geq F^{-1}(p_{FG})$. The set of all inflection values of F and G is denoted by Π_{FG}.

With this definition, a pair F, G of cdf's can fall into case (i) in Remark 4.4 and still have an inflection value $p_{FG} \in (0, 1)$. This holds, if and only if there exists a $t_0 \in D_F$ such that R_{FG} is linear on $(\inf D_F, t_0)$ or $(t_0, \sup D_F)$.

As stated in Remark 4.4, any pair F, G satisfying $F \leq_3 G$ has at least one inflection value. Requiring $R'''_{FG}(t) > 0$ for all $t \in D_F$ is sufficient for the inflection value p_{FG} to be unique.

With this in mind, we analyze more closely why $\leq_3$ is not transitive. Let F, G and H satisfy $F \leq_3 G$ and $G \leq_3 H$. Then,

$$R_{FH}(t) = H^{-1}(F(t)) = H^{-1}(G(G^{-1}(F(t)))) = R_{GH}(R_{FG}(t)),$$

and, consequently,

$$\begin{aligned}
R'_{FH}(t) &= R'_{GH}(R_{FG}(t)) \cdot R'_{FG}(t), \\
R''_{FH}(t) &= R''_{GH}(R_{FG}(t)) \cdot (R'_{FG}(t))^2 + R'_{GH}(R_{FG}(t)) \cdot R''_{FG}(t), && (4.2) \\
R'''_{FH}(t) &= R'''_{GH}(R_{FG}(t)) \cdot (R'_{FG}(t))^3 + R'_{GH}(R_{FG}(t)) \cdot R'''_{FG}(t) \\
&\quad + 3R''_{GH}(R_{FG}(t)) \cdot R'_{FG}(t) \cdot R''_{FG}(t) && (4.3)
\end{aligned}$$

holds for all $t \in D_F$. Note that any RIDF is increasing as a composition of two increasing functions. Hence, the first two summands on the right side of equation (4.3) are non-negative and

$$R''_{GH}(G^{-1}(p)) \cdot R''_{FG}(F^{-1}(p)) \geq 0 \quad \forall p \in (0,1)$$

is a sufficient condition for $F \leq_3 H$. By assumption, the sets Π_{FG} and Π_{GH} are both non-empty. If the intersection of these two sets is also non-empty, i.e., if there exists a $p_0 \in [0,1]$ such that $p_0 \in \Pi_{FG}$ and $p_0 \in \Pi_{GH}$, the signs of $R''_{FG}(F^{-1}(p))$ and $R''_{GH}(G^{-1}(p))$ coincide for all $p \in (0,1)$ since they are both non-positive for $p < p_0$ and both non-negative for $p > p_0$. Otherwise, if the intersection of Π_{FG} and Π_{GH} is empty, choose a representative from each set such that their difference is minimal. Assuming without restriction that $p_{FG} < p_{GH}$, where $p_{FG} \in \Pi_{FG}$ and $p_{GH} \in \Pi_{GH}$, and it follows that

$$R''_{GH}(G^{-1}(p)) \cdot R''_{FG}(F^{-1}(p)) < 0 \quad \forall p \in (p_{FG}, p_{GH}).$$

We summarize our results thus far in the following proposition.

Proposition 4.6. *Let $p_0 \in [0,1]$ and let $\mathcal{F}_0$ be a set of cdf's such that any pair $F, G \in \mathcal{F}_0$ with $F \leq_3 G$ has p_0 as an inflection value. Then, the order $\leq_3$ is transitive on $\mathcal{F}_0$.*

We now study the structure of the sets mentioned in Proposition 4.6 or suitable subsets thereof. First, we assume that F and G with $F \leq_3 G$ have an inflection value $p_{FG} \in (0,1)$ (so F and G could e.g. satisfy case (ii) from Remark 4.4). The fact that $p_{FG} = F(t_{FG}) \in (0,1)$ is an inflection value of the pair F, G is equivalent to $R''_{FG}(t_{FG}) = 0$. Denoting by f and g the density functions of F and G, respectively, we get

$$R'_{FG}(t) = \frac{f(t)}{g(R_{FG}(t))}, \tag{4.4}$$

$$R''_{FG}(t) = \frac{f'(t) \cdot (g(R_{FG}(t)))^2 - f^2(t) \cdot g'(R_{FG}(t))}{(g(R_{FG}(t)))^3} \tag{4.5}$$

for all $t \in D_F$. Hence, p_{FG} is an inflection value of F and G, if and only if

$$\begin{aligned} & f'(t_{FG}) \cdot (g(R_{FG}(t_{FG})))^2 = (f(t_{FG}))^2 \cdot g'(R_{FG}(t_{FG})) \\ \iff \quad & \frac{f'(t_{FG})}{(f(t_{FG}))^2} = \frac{g'(R_{FG}(t_{FG}))}{(g(R_{FG}(t_{FG})))^2} \\ \iff \quad & \frac{f'(F^{-1}(p_{FG}))}{(f(F^{-1}(p_{FG})))^2} = \frac{g'(G^{-1}(p_{FG}))}{(g(G^{-1}(p_{FG})))^2}. \end{aligned} \tag{4.6}$$

Hence, any pair that is ordered with respect to $\leq_3$ out of a given set of cdf's has the same inflection value $p_0 \in (0,1)$, if and only if the term $\frac{f'(F^{-1}(p_0))}{(f(F^{-1}(p_0)))^2}$ coincides for all cdf's F in that set. The following result is obtained by combining this observation with Proposition 4.6.

Proposition 4.7. *Let $p_0 \in (0,1)$ and let $\mathcal{F}_0$ be a set of cdf's such that the value of the term $\frac{f'(F^{-1}(p_0))}{(f(F^{-1}(p_0)))^2}$ coincides for all $F \in \mathcal{F}_0$. Then, all pairs $F, G \in \mathcal{F}_0$ with $F \leq_3 G$ have p_0 as an inflection value and, by Proposition 4.6, the order $\leq_3$ is transitive on $\mathcal{F}_0$.*

The situation in which $p_{FG} \in \{0,1\}$ is the sole inflection value of F and G with $F \leq_3 G$ remains to be considered. For these edge cases, (4.4) and (4.5) are not valid because the densities f and g are not uniquely defined at the edges of their respective supports. A necessary condition for $p_{FG} = 0$ is $F \leq_2 G$ and a necessary condition for $p_{FG} = 1$ is $G \leq_2 F$. Thus, $\leq_3$ is transitive on any subset of $\mathcal{T} \subseteq \mathcal{P}_I^3$ such that the implication $F \leq_3 G \Rightarrow F \leq_2 G$ holds for all $F, G \in \mathcal{T}$. The same is true for the implication $F \leq_3 G \Rightarrow G \leq_2 F$. However, since these sets are much more difficult to handle than the sets described in Proposition 4.7, they are not pursued any further.

For $p \in (0,1)$, the quantities $\gamma_D^p(F) = \frac{f'(F^{-1}(p))}{(f(F^{-1}(p)))^2}$ have already been defined in Section 3.1.3. They satisfy the crucial condition (S2) for skewness measures for all $p \in (0,1)$, but $p = \frac{1}{2}$ is the only choice for which γ_D^p is a skewness measure (see Corollary 3.13). By defining the set

$$\mathcal{T}_{D,p}^t = \{F \in \mathcal{P}_I^3 : \gamma_D^p(F) = t\}$$

for all $p \in (0,1)$ and all $t \in \mathbb{R}$ and defining $\mathcal{T}_D^t = \mathcal{T}_{D,\frac{1}{2}}^t$, we can rephrase Proposition 4.7 in the following way.

Theorem 4.8. *For any $t \in \mathbb{R}$ and any $p \in (0,1)$, the kurtosis order $\leq_3$ is transitive on the set $\mathcal{T}_{D,p}^t$.*

The fact that $\leq_3$ is transitive, if a suitable skewness measure is constant, suggests that the non-transitivity of $\leq_3$ on the set of all cdf's is because pairs of cdf's with differing degrees of skewness lack comparability with respect to kurtosis. Although this statement can only be applied to the sets $\mathcal{T}_D^t, t \in \mathbb{R}$, it is also true to a certain extent for $\mathcal{T}_{D,p}^t, t \in \mathbb{R}, p \in (0,1) \setminus \{\frac{1}{2}\}$, since the corresponding mappings γ_D^p also satisfy (S2). Because (S2) is the crucial property for determining whether a mapping measures skewness 'correctly', the mappings $\gamma_D^p, p \in (0,1) \setminus \{\frac{1}{2}\}$, can be thought of as asymmetric or non-central skewness measures (see remarks after Corollary 3.13). Concerning this interpretation, it is also notable that, for all $p \in (0,1)$, p is an inflection value of all pairs $F, G \in \mathcal{T}_{D,p}^t$ with $F \leq_3 G$. Thus, the inflection values for cdf's in $\mathcal{T}_{D,p}^t$ lie in the centre of the unit interval, if and only if $p = \frac{1}{2}$.

As opposed to the first two convex characteristics, location and dispersion, a distribution cannot be standardized with respect to skewness by an arithmetic operation like addition for location and scalar multiplication for dispersion. Thus, in order to obtain a transitive kurtosis order without interference caused by skewness, attention has to be restricted to sets of constant skewness. Note that, for all $p \in (0,1)$, the sets $\mathcal{T}_{D,p}^t, t \in \mathbb{R}$, constitute an (uncountable) partition of the entire underlying set $\mathcal{P}_I^3$ of distributions. $\bigcup_{t \in \mathbb{R}} \mathcal{T}_{D,p}^t = \mathcal{P}_I^3$ holds because all distributions are assigned a value by the measure γ_D^p, and $\mathcal{T}_{D,p}^t \cap \mathcal{T}_{D,p}^s = \emptyset$ holds for all $s, t \in \mathbb{R}, s \neq t$, because one distribution cannot be assigned multiple values by that measure. Thus, each $F \in \mathcal{P}_I^3$ lies within a subset of $\mathcal{P}_I^3$ on which $\leq_3$ is transitive.

These observations beg the question whether there exist other skewness measures that induce transitivity sets analogous to Theorem 4.8. To that end, note that a simple sufficient condition for the term $\frac{f'(F^{-1}(p_0))}{(f(F^{-1}(p_0)))^2}$ to coincide is to require $f'(F^{-1}(p_0)) = 0$ for all cdf's F in the given set. Hence, for each $p_0 \in (0,1)$, $\leq_3$ is transitive on the set of all cdf's, the density of which has a stationary point at the p_0-quantile. One well known point, at which this commonly occurs, is the mode of a distribution.

For the following considerations in which the mode is utilized, we assume that all distributions are unimodal. Recall that the mode of a cdf F is denoted by M_F. If the mode lies in the interior of the support, the differentiability assumptions on F directly yield $f'(M_F) = 0$ since M_F maximizes f. It follows that, for any $p \in (0,1)$ (or for any $\tilde{p} \in (-1,1)$, where $\tilde{p} = 1-2p$), $\frac{f'(F^{-1}(p))}{(f(F^{-1}(p)))^2} = 0$ holds for all cdf's F in the set

$$\mathcal{T}^{\tilde{p}}_{Mode} = \{F : M_F = F^{-1}(p)\} = \{F : 1 - 2F(M_F) = \tilde{p}\}.$$

This observation in combination with Proposition 4.7 yields the following result.

Theorem 4.9. *For any $\tilde{p} \in (-1,1)$, the kurtosis order $\leq_3$ is transitive on the set $\mathcal{T}^{\tilde{p}}_{Mode}$.*

For any $\tilde{p} \in (-1,1)$ and any pair of cdf's $F, G \in \mathcal{T}^{\tilde{p}}_{Mode}$ with $F \leq_3 G$, the corresponding inflection value is given by $p = \frac{\tilde{p}+1}{2}$. Since it was shown by Arnold and Groeneveld (1995, p. 35) and noted in Theorem 3.18 that $\gamma_{Mode}(F) = 1 - 2F(M_F), F \in \mathcal{P}^3_I$, is a measure of skewness, the transitivity of $\leq_3$ on the sets $\mathcal{T}^{\tilde{p}}_{Mode}$ has a similar interpretation to before: for $\leq_3$ to be transitive, the skewness of the involved distributions needs to be constant in some sense.

For distributions with modes at the edges of their interval supports, the above transitivity property does not hold, i.e., $\leq_3$ is not generally transitive on $\mathcal{T}^{-1}_{Mode}$ and $\mathcal{T}^{1}_{Mode}$. The crucial result in Proposition 4.7 does not hold in these edge cases. For the set $\mathcal{T}^{1}_{Mode}$, a specific counterexample can be constructed using Weibull distributions and their conditions for being ordered with respect to $\leq_3$, which are given in (4.10), (4.11) and (4.12). Note that all Weibull distributions are supported by $[0,\infty)$ and their mode is zero for values of the shape parameter in $(0,1]$.

Thus, while $\leq_3$ is transitive on sets where γ_{Mode} is constant on a value in $(-1,1)$, the same is not true for the values -1 an 1, i.e., if the distributions are too skewed in either direction. Particularly, the sets $\mathcal{T}^{\tilde{p}}_{Mode}, \tilde{p} \in (-1,1)$, do not provide a partition of the set of all (sufficiently regular) probability distributions on the real numbers as the ones obtained for the sets $\mathcal{T}^{t}_{D,p}$.

Note that the notion of a mode can be generalized without losing the transitivity of $\leq_3$ on the corresponding set of cdf's. Specifically, Theorem 4.9 still holds, if f only attains a local maximum at M_F, no longer assuming F to be unimodal. However, Arnold and Groeneveld (1995) only proved γ_{Mode} to be a skewness measure under the assumption of unimodality.

The transitivity sets found thus far and their relationships are summarized in the following remark.

Remark 4.10. a) Let $F \in \mathcal{S}$. According to Proposition 3.4c), the fact that γ_D and

γ_{Mode} are skewness measures in the sense of Definition 3.1 directly implies $\gamma_D(F) = \gamma_{Mode}(F) = 0$. It follows $\mathcal{S} \subseteq \mathcal{T}^0_{Mode}$ and $\mathcal{S} \subseteq \mathcal{T}^0_D$.

Now let $F \in \mathcal{T}^0_{Mode}$. This means that $2F(M_F) - 1 = \gamma_{Mode}(F) = 0$ or, equivalently, $M_F = F^{-1}(\frac{1}{2})$. Since the median of any continuous distribution lies in the interior of its support and it maximizes the continuous density f of F, we obtain $f'(F^{-1}(\frac{1}{2})) = 0$, and therefore $\gamma_D(F) = \frac{f'(F^{-1}(\frac{1}{2}))}{(f(F^{-1}(\frac{1}{2})))^2} = 0$. Overall, it follows $\mathcal{S} \subseteq \mathcal{T}^0_{Mode} \subseteq \mathcal{T}^0_D$.

b) Part a) can be generalized to asymmetric distributions in the following way. Let $\tilde{p} \in (-1, 1)$ and $F \in \mathcal{T}^{\tilde{p}}_{Mode}$. It follows that $M_F = F^{-1}(p)$, where $p = \frac{\tilde{p}+1}{2} \in (0, 1)$. Since that lies within the interior of the support of F, we obtain $f'(F^{-1}(p)) = 0$ and therefore $\gamma^p_D(F) = \frac{f'(F^{-1}(p))}{(f(F^{-1}(p)))^2} = 0$. Thus, the inclusion $\mathcal{T}^{\tilde{p}}_{Mode} \subseteq \mathcal{T}^0_{D,p}$ holds for all $p \in (0, 1)$ with $\tilde{p} = 2p - 1$.

Since $\leq_3$ is transitive on $\mathcal{T}^0_D$, it is also transitive on the set of all symmetric cdf's. Oja (1981), as the only paper with a significant mention of the order $\leq_3$ in the literature, dismissed it and instead focussed on the previously mentioned concave-convex order. The only given reason is that $\leq_3$ is not transitive. Oja, however, restricted all of his considerations concerning kurtosis to symmetric distributions and therefore also only proved the transitivity of the concave-convex order on this class. Since $\leq_3$ is also transitive on symmetric distributions, Oja's argument for dismissing it is refuted. This observation therefore strengthens the order of the third convex characteristic in its role as the fundamental kurtosis order.

Equivalence With Respect to $\leq_3$

The concept of equivalence with respect to a stochastic order is generally introduced in (2.5). Applied to the order of the third convex characteristic, $F =_3 G$, if $R'''_{FG} \geq 0$ and $R'''_{GF} \geq 0$. To better understand what that means, we rewrite the second condition. Since $R_{GF} = R^{-1}_{FG}$, the first three derivatives of R_{GF} are given by

$$\begin{aligned} R'_{GF}(t) &= \frac{1}{R'_{FG}(R_{GF}(t))}, \\ R''_{GF}(t) &= -\frac{R''_{FG}(R_{GF}(t))}{(R'_{FG}(R_{GF}(t)))^3}, \\ R'''_{GF}(t) &= 3\frac{(R''_{FG}(R_{GF}(t)))^2}{(R'_{FG}(R_{GF}(t)))^5} - \frac{R'''_{FG}(R_{GF}(t))}{(R'_{FG}(R_{GF}(t)))^4} \\ &= \frac{3(R''_{FG}(R_{GF}(t)))^2 - R'''_{FG}(R_{GF}(t)) \cdot R'_{FG}(R_{GF}(t))}{(R'_{FG}(R_{GF}(t)))^5}. \end{aligned} \tag{4.7}$$

It follows that

$$G \leq_3 F \Leftrightarrow R'''_{GF}(t) \geq 0 \quad \forall t \in D_G$$

$$\Leftrightarrow \frac{3(R''_{FG}(t))^2 - R'''_{FG}(t) \cdot R'_{FG}(t)}{(R'_{FG}(t))^5} \geq 0 \quad \forall t \in D_F$$
$$\Leftrightarrow R'''_{FG}(t) \leq 3\frac{(R''_{FG}(t))^2}{R'_{FG}(t)} \quad \forall t \in D_F.$$

The fact that $G \leq_3 F$ is not equivalent to $R'''_{FG} \leq 0$ and therefore $F =_3 G$ is not equivalent to $R'''_{FG} \equiv 0$ is notable as it systematically differs from what can be observed with the orders $\leq_0$, $\leq_1$ and $\leq_2$. Equivalence with respect to any of these orders occurs if and only if the corresponding derivative of the RIDF is constantly zero (see Proposition 3.3). However, $F =_3 G$ is equivalent to R_{FG} satisfying

$$0 \leq R'''_{FG}(t) \leq 3\frac{(R''_{FG}(t))^2}{R'_{FG}(t)}$$

for all $t \in D_F$. Obviously, an analogous condition applies to R_{GF}. Hence, we have the following result.

Proposition 4.11. *$F =_3 G$ holds, if and only if R_{FG} satisfies the differential inequality*

$$0 \leq \varphi'''(t) \leq 3\frac{(\varphi''(t))^2}{\varphi'(t)} \quad \forall t \in D_F.$$

Recalling that $F =_2 G$, if and only if there exist $a > 0, b \in \mathbb{R}$ such that $F(\cdot) = G(a \cdot +b)$, one might guess that $F =_3 G$ holds for quadratic transforms. The following example shows that this is indeed true.

Example 4.12. Let $R_{FG}(t) = t^p, 0 < t < 1$, for some $p > 0, p \neq 1$. This RIDF arises, for example, for $F(t) = t, G(t) = t^{1/p}$ or for $F(t) = t^p, G(t) = t$. For $p \notin \{1, 2\}$, $F \leq_3 G$ is equivalent to

$$0 \leq R'''_{FG}(t) = p(p-1)(p-2)t^{p-3} \; \forall t \in (0,1) \Leftrightarrow p \notin (1,2).$$

Because of $R'''_{FG} \equiv 0$ for $p \in \{1, 2\}$, $F \leq_3 G$ is generally equivalent to $p \notin (1, 2)$. Conversely, for $p \notin \{1, 2\}$, $G \leq_3 F$ is equivalent to

$$p(p-1)(p-2)t^{p-3} = R'''_{FG}(t) \leq 3\frac{(R_{FG}(t)'')^2}{R'_{FG}(t)} = 3p(p-1)^2 t^{p-3} \; \forall t \in (0,1) \Leftrightarrow \; p \notin (\tfrac{1}{2}, 1).$$

Since the inequality is obviously satisfied for $p \in \{1, 2\}$, $G \leq_3 F$ is equivalent to $p \notin (\frac{1}{2}, 1)$ in general. Overall, $F =_3 G$ is satisfied, if and only if

$$p \in (0, 1/2] \cup \{1\} \cup [2, \infty).$$

In particular, $F(t) = t, t \in (0, 1)$, and $G(t) = t^2, t \in (0, 1)$, are equivalent with respect to $\leq_3$.

Note that $R'''_{FG} \equiv 0$ and therefore also $F =_3 G$ holds, if R_{FG} is any polynomial of degree

≤ 2. This result can also easily be obtained using Corollary 2.16. While $F =_2 G$ is equivalent to R_{FG} being a polynomial of degree ≤ 1, the fact that R_{FG} is a polynomial of degree ≤ 2 is only a sufficient, but not a necessary condition for $F =_3 G$.

The result of Example 4.12 can also be obtained by checking $R'''_{FG}(t) \geq 0$ and $R'''_{GF}(t) \geq 0$ for $t \in (0,1)$, but the advantage of using Proposition 4.11 is that only the derivatives of one of the two functions must be known. This is illustrated by the following example.

Example 4.13. Let F be the cdf of a symmetric beta distribution with parameter $p = q > 0$ and let $G(t) = t, t \in (0,1)$. For $p \in (0,\infty) \setminus \{1,2\}$, the derivatives of the corresponding RIDF are given by

$$\begin{aligned} R'_{FG}(t) &= \frac{t^{p-1}(1-t)^{p-1}}{B(p,p)}, \\ R''_{FG}(t) &= \frac{(p-1)t^{p-2}(1-t)^{p-2}(1-2t)}{B(p,p)}, \\ R'''_{FG}(t) &= \frac{(p-1)t^{p-3}(1-t)^{p-3}\left((4p-6)t^2-(4p-6)t+p-2\right)}{B(p,p)}, \end{aligned} \tag{4.8}$$

where $B(\cdot,\cdot)$ denotes the beta function. The quadratic formula can be applied to obtain that the non-negativity of (4.8) for all $t \in (0,1)$ is equivalent to $p \leq 1$. Since $p = 1$ implies $R'''_{FG} \equiv 0$ and $p = 2$ implies $R'''_{FG} \equiv -\frac{1}{3}$, $F \leq_3 G$ is overall equivalent to $p \leq 1$.

For the reverse statement $G \leq_3 F$, again let $p \in (0,\infty) \setminus \{1,2\}$ and consider

$$3\frac{(R''_{FG}(t))^2}{R'_{FG}(t)} = \frac{3(p-1)^2 t^{p-3}(1-t)^{p-3}(1-2t)^2}{B(p,p)}, \quad t \in (0,1).$$

By combining this with (4.8), the right inequality in Proposition 4.11 boils down to

$$(p-1)\Big((8p-6)(t^2-t)+2p-1\Big) \geq 0 \quad \forall t \in (0,1),$$

which is true for $p \geq 1$. Furthermore, the same inequality is easily shown to be true for $p \in \{1,2\}$. Overall, $G \leq_3 F$ holds, if and only if $p \geq 1$. Hence, the only symmetric beta distribution that is equivalent to the uniform distribution with respect to $\leq_3$ is the uniform distribution itself.

4.2.2. Concave-Convex Orders

We start out by recalling two definitions from the literature (see van Zwet, 1964, p. 65 and MacGillivray and Balanda, 1988, p. 326).

Definition 4.14. a) Let $F, G \in \mathcal{S}$. Then, F is said to be *less kurtotic than G in the convex sense*, denoted by $F \leq_s G$, if R_{FG} is convex on $D_F \cap [F^{-1}(\frac{1}{2}), \infty)$.

b) F is said to be *less kurtotic than G in the anti-skewness sense*, denoted by $F \leq_a G$, if R_{FG} is concave on $D_F \cap (-\infty, F^{-1}(\frac{1}{2})]$ and convex on $D_F \cap [F^{-1}(\frac{1}{2}), \infty)$.

Since the centre of symmetry of a symmetric distribution is equal to its median, it is easy to see that the order $\leq_s$ is equivalent to the anti-skewness order $\leq_a$ on the set $\mathcal{S}$.

The following holds for the kurtosis order $\leq_a$: since it requires the RIDF to switch from negative to positive curvature at $F^{-1}(\frac{1}{2})$, it immediately follows that the RIDF has an inflection point with corresponding inflection value $\frac{1}{2}$. Thus, the order is a priori only able to compare two cdf's in terms of their kurtosis if they have $\frac{1}{2}$ as an inflection value. While this is always the case for symmetric cdf's (see Remark 4.10), there is no reason to assume that this is a prerequisite for two cdf's to be ordered with respect to kurtosis. Particularly, the more fundamental order $\leq_3$ is able to order pairs of cdf's irrespective of their inflection value. It follows that $\leq_3$ does generally not imply the anti-skewness order $\leq_a$. A valid counterexample would be the cdf's F and G from Example 4.2 as they are two cdf's that can be ordered with respect to $\leq_3$ but do not have $\frac{1}{2}$ as an inflection value. Instead, their inflection value is $p_{FG} = 0$. Oja (1981) already noted that $\leq_3$ does indeed imply $\leq_a$ as an extension of $\leq_s$, if we restrict ourselves to symmetric cdf's. In fact, this class of cdf's can even be extended to $\mathcal{T}^0_{Mode}$, $\mathcal{T}^0_D$ or any other class of cdf's such that any pair within has inflection value $\frac{1}{2}$. Formally, $F \leq_3 G$ with $\frac{1}{2} \in \Pi_{FG}$ implies $F \leq_a G$.

As a possible solution, we define our own concave-convex kurtosis order that is flexible in terms of the inflection value.

Definition 4.15. F is said to be *less kurtotic in the concave-convex sense* than G, denoted by $F \leq_{gs} G$, if there exists a $p_{FG} \in [0,1]$ such that R_{FG} is concave on $D_F \cap (-\infty, F^{-1}(p_{FG}))$ and convex on $D_F \cap (F^{-1}(p_{FG}), \infty)$.

This more general definition has a much more straight forward relationship to the order of the third convex characteristic.

Proposition 4.16. *$F \leq_3 G$ implies $F \leq_{gs} G$ whereas the reverse implication does not hold in general.*

Proof. $F \leq_3 G$ is equivalent to $R'''_{FG}(t) \geq 0$ for all $t \in D_F$ and, therefore, to R''_{FG} being increasing on D_F. It follows that either $R''_{FG} \leq 0$, $R''_{FG} \geq 0$, or R''_{FG} changes its sign once from negative to positive at some point $t_{FG} \in D_F$. All the cases imply $F \leq_{gs} G$, in the first case with inflection value $p_{FG} = 1$, in the second with $p_{FG} = 0$ and in the third with $p_{FG} = F(t_{FG}) \in (0,1)$.

In order to see that the reverse implication is not true in general, consider the following counterexample:

$$\begin{aligned} F(t) &= -\tfrac{3}{14}t^5 + \tfrac{5}{7}t^3 + \tfrac{1}{2}, \quad && t \in [-1,1] \quad \text{and} \\ G(t) &= t, && t \in [0,1] \end{aligned}$$

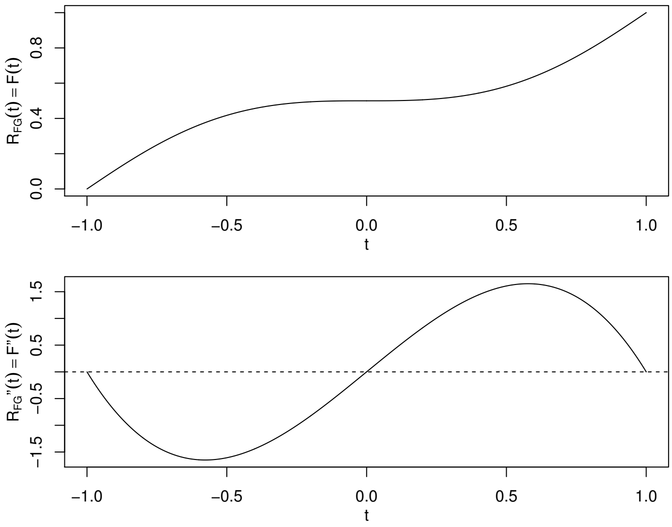

Figure 4.3.: Graph of the function $R_{FG} = F$ and its second derivative from the counterexample disproving that $F \leq_{gs} G$ implies $F \leq_3 G$ (given in the proof of Proposition 4.16).

(see Figure 4.3). Note that F is indeed a cdf as

$$F(-1) = 0, \quad F(1) = 1, \quad F'(t) = \tfrac{15}{14}\left(2t^2 - t^4\right) \geq 0 \; \forall t \in [-1,1].$$

Furthermore, because of $R_{FG} = F$, we obtain

$$R''_{FG}(t) = \tfrac{30}{7}(t - t^3) \begin{cases} \leq 0 & \text{for } -1 \leq t \leq 0, \\ \geq 0 & \text{for } 0 \leq t \leq 1, \end{cases}$$

$$R'''_{FG}(t) = \tfrac{30}{7}(1 - 3t^2) \begin{cases} \leq 0 & \text{for } t \leq -\frac{1}{\sqrt{3}}, t \geq \frac{1}{\sqrt{3}}, \\ \geq 0 & \text{for } -\frac{1}{\sqrt{3}} \leq t \leq \frac{1}{\sqrt{3}}. \end{cases}$$

It follows that $F \leq_{gs} G$ holds (with $p_{FG} = \frac{1}{2}$) whereas $F \not\leq_3 G$. ▨

Note that the cdf F in the above counterexample does not fit the formal assumptions made at the beginning of Chapter 4 because its derivative satisfies $f(0) = 0$. Thus, f is not strictly positive on D_F. However, this can easily be solved by slightly increasing the value $f(0)$ and slightly decreasing the absolute value of the slope of f throughout. This does not change the nature of the example, but complicates the necessary computations.

The counterexample given in the proof of Proposition 4.16 nicely captures the difference between the two orders $\leq_3$ and $\leq_{gs}$. The RIDF R_{FG} in the example has three distinct inflection points at -1, 0 and 1, which would not be possible with $F \leq_3 G$ holding. Between the first two, R''_{FG} has a local minimum and between the latter two, it has a local maximum. Once again, this would not be possible with $F \leq_3 G$ holding because then, R''_{FG} would be

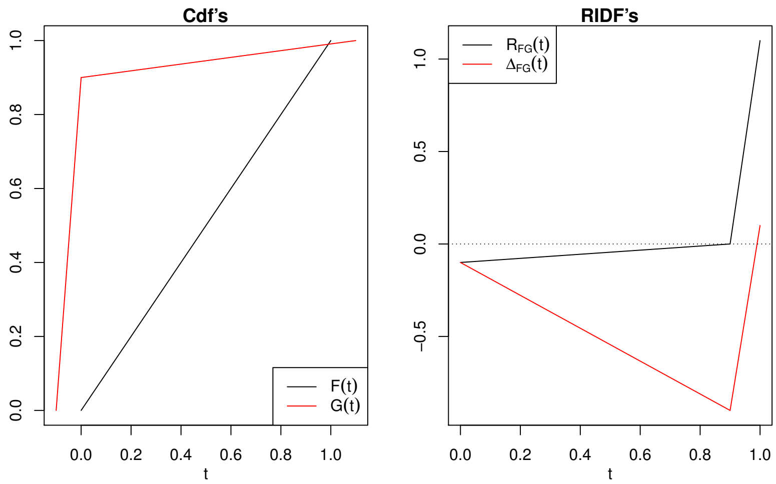

Figure 4.4.: Graphs of the cdf's F and G from Example 4.17 in the left panel and of the corresponding RIDF's in the right panel with $\varepsilon = \frac{1}{10}$.

required to be increasing. In a nutshell, if we move from $F \leq_3 G$ to $F \leq_{gs} G$, the property of being increasing gets replaced by the property of switching from being non-positive to being non-negative at a certain point.

As a more conceivable analogue, one can define corresponding orders for dispersion instead of kurtosis, since dispersion is the only symmetric lower order convex characteristic. The sole difference is that the two aforementioned properties are applied to the function $\Delta_{FG} = R_{FG} - \mathrm{id}$ instead of R''_{FG}. The analogue to $F \leq_3 G$ would be to require Δ_{FG} to be non-decreasing, which is equivalent to $F \leq_1 G$ or $F \leq_{disp} G$, i.e. the dispersive order. The analogue to $F \leq_{gs} G$, however, would be to require Δ_{FG} to be non-positive up to a certain point and non-negative from that point on. This latter analogue is very similar to the so-called 'more dangerous'-order $\leq_D$ (see Müller and Stoyan, 2002, p. 23). Specifically, $F \leq_D G$ is defined by the existence of a $t_0 \in \mathbb{R}$ such that $\Delta_{FG}(t) \leq 0$ for $t \leq t_0$ and $\Delta_{FG} \geq 0$ for $t \geq t_0$, and the additional requirement $\mathbb{E}[X] \leq \mathbb{E}[Y]$. The 'more dangerous'-order is equivalent to the convex order $\leq_{cx}$, if the additional requirement is strengthened to $\mathbb{E}[X] = \mathbb{E}[Y]$ (see Müller and Stoyan, 2002, pp. 17, 23). Recall from Example 2.21b) that $\leq_{cx}$ orders distributions with respect to dispersion. The following example illustrates that we do not obtain a meaningful dispersion order without the additional condition concerning the expectations.

Example 4.17. Let $\varepsilon \in (0, \frac{1}{2})$ and define

$$
\begin{aligned}
F(t) &= t, \quad t \in [0,1], \\
G(t) &= \left(\frac{1-\varepsilon}{\varepsilon} t + 1 - \varepsilon\right) \mathbb{1}_{[-\varepsilon,0]}(t) + \left(\frac{\varepsilon}{1+\varepsilon} t + 1 - \varepsilon\right) \mathbb{1}_{(0,1+\varepsilon]}(t), \quad t \in [-\varepsilon, 1+\varepsilon]
\end{aligned}
$$

(see Figure 4.4). The corresponding RIDF is given by

$$R_{FG}(t) = G^{-1}(t) = \begin{cases} \frac{\varepsilon}{1-\varepsilon}t - \varepsilon, & \text{if } t \in (0, 1-\varepsilon], \\ \frac{1+\varepsilon}{\varepsilon}t - \frac{1-\varepsilon^2}{\varepsilon}, & \text{if } t \in (1-\varepsilon, 1), \end{cases}$$

$$\Delta_{FG}(t) = G^{-1}(t) - t = \begin{cases} \frac{2\varepsilon-1}{1-\varepsilon}t - \varepsilon, & \text{if } t \in (0, 1-\varepsilon], \\ \frac{1}{\varepsilon}t - \frac{1-\varepsilon^2}{\varepsilon}, & \text{if } t \in (1-\varepsilon, 1). \end{cases}$$

First, we have $\Delta_{FG}(t) < 0$ for $t \in (0, 1-\varepsilon]$ since $\lim_{t \searrow 0} \Delta_{FG}(t) = -\varepsilon$ and since the function is decreasing on $(0, 1-\varepsilon]$ because of $\Delta'_{FG}(t) = \frac{2\varepsilon-1}{1-\varepsilon} < 0$. Furthermore, Δ_{FG} is continuous in $1-\varepsilon$ because of $\Delta_{FG}(1-\varepsilon) = 2\varepsilon - 1 - \varepsilon = \varepsilon - 1 = \frac{\varepsilon^2-\varepsilon}{\varepsilon} = \lim_{t \searrow 1-\varepsilon} \Delta_{FG}(t)$ and from there, it linearly approaches its limiting value $\lim_{t \nearrow 1} \Delta_{FG}(t) = \frac{1}{\varepsilon} - \frac{1-\varepsilon^2}{\varepsilon} = \varepsilon$. Overall, since $\Delta_{FG}(1-\varepsilon^2) = 0$, we have $\Delta_{FG}(t) < 0$ for $t \in (0, 1-\varepsilon^2)$ and $\Delta_{FG}(t) > 0$ for $t \in (1-\varepsilon^2, 1)$. It follows that F is less dispersed than G with respect to the dispersion analogue of the kurtosis order $\leq_{gs}$.

However, for $\varepsilon \searrow 0$, G converges towards the cdf of a degenerate distribution with all of the probability mass concentrated in 0. Since that limiting distribution exhibits no dispersion at all, there should exist an $\varepsilon_0 \in (0, \frac{1}{2})$ such that G is less dispersed than F, or at least that F is not less dispersed than G. This is supported by the behaviour of popular measures of dispersion like the standard deviation and the interquartile range. Their values for the cdf G drop below their values of F at some point if $\varepsilon \searrow 0$.

The fact that its analogue dispersion order is not meaningful also casts doubt upon the suitability of $\leq_{gs}$ as an order of kurtosis. Oja (1981, p. 158) defined another similar dispersion order to the one considered in Example 4.17, which has no additional requirements on the expectations. Specifically, he defined $F \leq_1^* G$, if there exists a $t_0 \in D_F$ such that $\Delta_{FG}(t) \leq \mathbb{E}[Y] - \mathbb{E}[X]$ for $t \leq t_0$ and $\Delta_{FG}(t) \geq \mathbb{E}[Y] - \mathbb{E}[X]$ for $t \geq t_0$. Just like the 'more dangerous'-order, it turns into the order considered in Example 4.17 as a special case, if both involved distribution have the same expected value. This dispersion order is shown to be weaker than $\leq_1$ but it is still preserved by, e.g, the standard deviation (see Oja, 1981, p. 159). The former result follows from Proposition 2.17b) being applied to the centralized cdf's $F(\cdot - \mathbb{E}[X])$ and $G(\cdot - \mathbb{E}[Y])$. Additionally, if $F \leq_1^* G$ holds, it is also guaranteed that Δ_{FG} assumes the value $\mathbb{E}[Y] - \mathbb{E}[X]$ at some point. The same cannot be said about the value 0, which is used in the place of $\mathbb{E}[Y] - \mathbb{E}[X]$ in Example 4.17.

Overall, it is fairly clear that $\leq_{gs}$ is not a suitable kurtosis order for analogous reasons, namely that 0 is an somewhat arbitrary threshold for the second derivative of the RIDF. While the results of Oja (1981) demonstrate that 0 is a suitable threshold for symmetric distributions, it is not if there is a notable difference in skewness between the two involved distributions. Consider the case in which $F \leq_2 G$ holds, so if G is more skewed to the right than F: there can still exist a notable difference in shape beyond skewness between F and G, however, this could not be detected by $\leq_{gs}$ as, in this case, $F =_{gs} G$ holds. The solution is to change the threshold, so that even a strictly convex RIDF can be under the threshold for its second derivative up to some point and over it from that point on.

Definition 4.18. Let $t_0 \in \mathbb{R}$. Then, F is said to be *less kurtotic than G in the concave-convex sense with threshold* t_0, denoted by $F \leq_{gs}^{t_0} G$, if there exists a $p_{FG}^{t_0} \in [0,1]$ such that $R''_{FG}(t) \leq t_0$ holds for all $t \in D_F \cap (-\infty, F^{-1}(p_{FG}^{t_0}))$ and $R''_{FG}(t) \geq t_0$ holds for all $t \in D_F \cap (F^{-1}(p_{FG}^{t_0}), \infty)$.

Note that $F \leq_{gs}^{0} G \Leftrightarrow F \leq_{gs} G$. While the order $\leq_{gs}^{t_0}$ is theoretically defined for all $t_0 \in \mathbb{R}$, its imposed requirement is only meaningful, if $t_0 \in \operatorname{int}(R''_{FG}(D_F))$. Otherwise, it is obvious that either $R''_{FG}(t) \leq t_0$ or $R''_{FG}(t) \geq t_0$ holds for all $t \in D_F$. Hence, all thresholds $t_0 \in \operatorname{int}(R''_{FG}(D_F))$ are said to be *reasonable*. The only exception to this rule is the case that the set of reasonable thresholds is empty, which is equivalent to R''_{FG} being constant. In this rather uninteresting case with respect to kurtosis, the sole value of R''_{FG} is the only candidate for a reasonable threshold.

Because a cdf cannot be standardized with respect to skewness, a reasonable threshold cannot be obtained in the same way as for $\leq_1^*$. However, another specific choice for a reasonable threshold t_0 is considered in the following example.

Example 4.19. Define the specific threshold

$$t_{0,D} = R''_{FG}(F^{-1}(\tfrac{1}{2})) = \frac{\tau_D(G)}{\tau_D(F)^2} \left(\gamma_D(G) - \gamma_D(F)\right),$$

where τ_D and γ_D are defined as in Theorem 3.12. Obviously, this threshold is meaningful in the sense that $t_{0,D} \in R''_{FG}(D_F)$. The only case in which $t_{0,D}$ is not reasonable is if R''_{FG} is constant on $D_F \cap (-\infty, F^{-1}(\frac{1}{2})]$ of $D_F \cap [F^{-1}(\frac{1}{2}), \infty)$, but not on its entire domain D_F. Furthermore, note that the corresponding concave-convex kurtosis order $\leq_{gs}^{t_{0,D}}$ is equivalent to the original concave-convex order $\leq_{gs}$ if we restrict ourselves to equally skewed distributions with respect to the skewness measure γ_D. This verifies our previously established intuition that non-zero thresholds are necessary, if there is a significant difference in skewness between the two distributions to be compared in terms of kurtosis.

Analogously to how no intersection of the third derivative with the zero-function (i.e. $F \leq_3 G$) implies one intersection of the second derivative of Δ_{FG} with a suitable constant function (i.e. $F \leq_{gs}^{t_0} G$), we can further infer that the first derivative intersects a suitable linear function twice (with according sign changes). However, this second implication is not pursued any further here since it requires even more constants to be determined. It is desirable for a fundamental order to be as independent as possible from specific constants in order to be as general as possible and to not favour some measures over others. For these reasons, we limit ourselves to the kurtosis orders $\leq_3$ and $\leq_{gs}^{t_0}, t_0 \in \mathbb{R}$. The relationship between these two orders is given in the following.

Theorem 4.20. *The following three statements are equivalent:*

(i) $F \leq_3 G$*,*

(ii) $F \leq_{gs}^{t_0} G$ *for all* $t_0 \in \mathbb{R}$,

(iii) $F \leq_{gs}^{t_0} G$ *for all* $t_0 \in \operatorname{int}(R''_{FG}(D_F))$.

Proof. The implication (i)⇒(iii) is shown analogously to the proof of Proposition 4.16. For the reverse implication, let $t_1 \in D_F$. If t_1 lies within an interval, on which R''_{FG} is constant, $R'''_{FG}(t_1) = 0$ follows. Otherwise, it follows that $t_0 = R''_{FG}(t_1) \in \operatorname{int}(R''_{FG}(D_F))$. Now

$$R'''_{FG}(t_1) = \lim_{\varepsilon \searrow 0} \frac{R''_{FG}(t_1 + \varepsilon) - R''_{FG}(t_1 - \varepsilon)}{2\varepsilon} \geq 0$$

holds because of $R''_{FG}(t_1 + \varepsilon) \geq t_0$ and $R''_{FG}(t_1 - \varepsilon) \leq t_0$ by assumption. The assertion follows since t_1 was chosen arbitrarily.

The equivalence between (ii) and (iii) holds because either $R''_{FG}(t) \leq t_0$ or $R''_{FG} \geq t_0$ is true by construction for all unreasonable thresholds $t_0 \notin \operatorname{int}(R''_{FG}(D_F))$. ▨

Corollary 4.21. *Let* $t_0 \in \mathbb{R}$. *Then,* $F \leq_3 G$ *implies* $F \leq_{gs}^{t_0} G$ *whereas the reverse implication does not hold in general.*

In the following, we examine the concave-convex kurtosis orders with respect to transitivity. First, the following result states that $\leq_{gs}^{t_0}$ is generally not transitive and therefore not superior to $\leq_3$ in that respect.

Proposition 4.22. *For all* $t_0 \in \mathbb{R}$, *the kurtosis order* $\leq_{gs}^{t_0}$ *is not generally transitive.*

Proof. A counterexample for the transitivity of $\leq_{gs}^{t_0}$ can be obtained for all $t_0 \in \mathbb{R}$ by again reusing Example 4.2 and slightly modifying the cdf H. For $c > 0$, let

$$H : [0, c] \to [0, 1], \quad t \mapsto 1 - \sqrt[3]{\frac{c - t}{c}}.$$

This implies that the functions R_{GH} and R_{FH} as well as all of their derivatives are multiplied by the factor c. So, additionally to $F \leq_3 G$, $R'''_{GH}(t) = 6c \geq 0$ holds for all $t \in [0, 1]$ and, thus, $G \leq_3 H$. Corollary 4.21 states that both $F \leq_{gs}^{t_0} G$ and $G \leq_{gs}^{t_0} H$ hold for all $t_0 \in \mathbb{R}$. In contrast, we have

$$R''_{FH}(t) = 18c(4t^7 - 5t^4 + t) \begin{cases} < 0 & \text{for } t \in (2^{-\frac{2}{3}}, 1), \\ = 0 & \text{for } t \in \{0, 2^{-\frac{2}{3}}, 1\}, \\ > 0 & \text{for } t \in (0, 2^{-\frac{2}{3}}). \end{cases}$$

It follows that, for any $t_0 > 0$, there exists a $c > 0$ such that R''_{FH} first takes values smaller than t_0, then larger, and finally smaller again. For any $t_0 < 0$, there exists a $c > 0$ such that R''_{FH} first takes values larger than t_0, then smaller and finally larger again. For $t_0 = 0$, we obtain $R''_{FH}(t) \geq 0$ for $t \leq 2^{-\frac{2}{3}}$ and $R''_{FH}(t) \leq 0$ for $t \geq 2^{-\frac{2}{3}}$. All three cases pose a contradiction to $F \leq_{gs}^{t_0} G$. ▨

Remark 4.23. a) For symmetric cdf's F and G, R_{FG} always has an inflection point at $F^{-1}(\frac{1}{2})$. Thus, $\leq_{gs}$ is equivalent to $\leq_s$ on $\mathcal{S}$ and therefore also transitive on $\mathcal{S}$ (see Oja, 1981, p. 165). The situation for $\leq_{gs}^{t_0}, t_0 \neq 0$, is different because the critical switch from $R''_{FG}(t) \leq t_0$ to $R''_{FG}(t) \geq t_0$ cannot occur at $F^{-1}(\frac{1}{2})$ due to the point symmetry of R_{FG}.

b) The specific order $\leq_{gs}$ (or, equivalently, $\leq_{gs}^0$) can be altered slightly to become transitive on the more general sets $\mathcal{T}_{Mode}^{\tilde{p}}$, $\tilde{p} \in (-1,1)$, and $\mathcal{T}_{D,p}^t$, $t \in \mathbb{R}, p \in (0,1)$. For two cdf's F and G, we say that $F <_{gss} G$ holds, if there exists a $p_{FG} \in [0,1]$ such that R''_{FG} is strictly negative on $D_F \cap (-\infty, F^{-1}(p_{FG}))$ and strictly positive on $D_F \cap (F^{-1}(p_{FG}), \infty)$. Note that $<_{gss}$ is not equivalent to $<_{gs}$ since the latter is defined by

$$F <_{gs} G \Leftrightarrow F \leq_{gs} G \text{ and } F \neq_{gs} G \Leftrightarrow F \leq_{gs} G \text{ and } G \not\leq_{gs} F,$$

as generally stipulated by 2.4. To see that $<_{gss}$ is transitive, let $p \in (0,1)$ and $F, G, H \in \mathcal{T}_{D,p}^t$ with $F <_{gss} G$ and $G <_{gss} H$. By the line of reasoning used to prove Proposition 4.7 and Theorem 4.8, $R''_{FG}(F^{-1}(p)) = 0 = R''_{GH}(G^{-1}(p))$ then holds. Since, by definition of $<_{gss}$, there exists at most one $t \in D_F$ and one $s \in D_G$ such that $R''_{FG}(t) = 0$ and $R''_{GH}(s) = 0$, $t = F^{-1}(p)$ and $s = G^{-1}(p)$ follows. Considering (4.2) for $t = F^{-1}(p)$ along with the fact that R_{GH} is increasing, this yields $R''_{FH}(F^{-1}(q)) < 0$ for $q < p$ and $R''_{FH}(F^{-1}(q)) > 0$ for $q > p$. Overall, $F <_{gss} H$ follows.

The transitivity of $<_{gss}$ on the sets $\mathcal{T}_{Mode}^{\tilde{p}}$, $p \in (-1,1)$, now follows from $\mathcal{T}_{Mode}^{\tilde{p}} \subseteq \mathcal{T}_{D,p}^0$, where $p = \frac{\tilde{p}+1}{2}$.

The order $\leq_{gs}$ cannot be shown to be transitive on the given sets in the same way as $<_{gss}$. This is because if we only assume $F \leq_{gs} G$, $R''_{FG}(F^{-1}(p)) = 0$ for any $p \in (0,1)$ is not sufficient to infer that p is an inflection value. Because the concavity and the convexity of R_{FG} on either side of the actual inflection value is not assumed to be strict, the RIDF could be convex on both sides of $F^{-1}(p)$ or concave on both sides.

Further problems arise if the transitivity of $\leq_{gs}^{t_0}$ or $<_{gss}^{t_0}$ on any of the above transitivity sets is considered for $t_0 \neq 0$. (Naturally, $F <_{gss}^{t_0} G$ is defined by $R''_{FG} < t_0$ holding up to a certain point and $R''_{FG} > t_0$ from that point onward.) Note that the notion of an inflection value was chosen in such a way that the second derivative of the RIDF in question is zero at the point associated with that inflection value. This is the natural choice as it associates the inflection value with actual inflection points of the RIDF. Through (4.2) this only results in implications concerning the points, at which R''_{FH} is equal to zero. However, it remains unclear, if and at which points R''_{FH} is equal to (or smaller/larger than) the threshold t_0. In order to circumvent this problem, one could propose to generalize the notion of a inflection value and, consequently, of the transitivity sets. Particularly, one could associate it with the second derivative of the RIDF in question passing the value t_0. Considering (4.2), the behaviour of R_{FH} with respect to the order $\leq_{gs}^{t_0}$ would then also dependent on certain values of the first derivatives R'_{FG} and

R'_{GH}, if $t_0 \neq 0$. In the opinion of the author, this would further complicate the transitivity structure of $\leq_{gs}^{t_0}$ to a point that renders it essentially useless. For this reason, we will not examine said structure further at this point.

Equivalence With Respect to Concave-Convex Orders

Let $t_0 \in \mathbb{R}$. Because of (4.7), $G \leq_{gs}^{t_0} F$ is equivalent to

$$\begin{aligned}
&\exists t_{GF} \in D_G : R''_{GF}(t) \leq t_0 \ \forall t \leq t_{GF} \text{ and } R''_{GF}(t) \geq t_0 \ \forall t \geq t_{GF} \\
\Leftrightarrow\ &\exists t_{GF} \in D_G : R''_{FG}(R_{GF}(t)) \geq -t_0 \cdot (R'_{FG}(R_{GF}(t)))^3 \ \forall t \leq t_{GF} \text{ and} \\
&\qquad\qquad\quad R''_{FG}(R_{GF}(t)) \leq -t_0 \cdot (R'_{FG}(R_{GF}(t)))^3 \ \forall t \geq t_{GF} \\
\Leftrightarrow\ &\exists \tilde{t}_{GF} \in D_F : R''_{FG}(t) \geq -t_0 \cdot (R'_{FG}(t))^3 \ \forall t \leq \tilde{t}_{GF} \text{ and} \\
&\qquad\qquad\quad R''_{FG}(t) \leq -t_0 \cdot (R'_{FG}(t))^3 \ \forall t \geq \tilde{t}_{GF}.
\end{aligned}$$

In the case $t_0 = 0$, this leads to a fairly simple structure for $F =_{gs}^{t_0} G$ (which, then, is equivalent to $F =_{gs} G$), namely

$$\begin{aligned}
\exists t_{FG}, \tilde{t}_{GF} \in D_F : R''_{FG}(t) \leq 0 \ \forall t \in D_F : t \leq t_{FG} \vee t \geq \tilde{t}_{GF} \text{ and}& \\
R''_{FG}(t) \geq 0 \ \forall t \in D_F : t \geq t_{FG} \vee t \leq \tilde{t}_{GF}& \qquad (4.9)
\end{aligned}$$

If $t_{FG} = \tilde{t}_{GF}$, it directly follows that $R''_{FG} \equiv 0$. $t_{FG} < \tilde{t}_{GF}$ yields $R''_{FG}(t) \geq 0$ for $t \in (t_{FG}, \tilde{t}_{GF})$ and $R''_{FG}(t) = 0$ otherwise. Similarly, $\tilde{t}_{GF} < t_{FG}$ yields $R''_{FG}(t) \leq 0$ for $t \in (\tilde{t}_{GF}, t_{FG})$ and $R''_{FG}(t) = 0$ otherwise. In any case, F and G are skewness-comparable with respect to $\leq_2$. Conversely, any skewness-comparable pair F, G of cdf's satisfies (4.9) by choice of $t_{FG} = \inf D_F, \tilde{t}_{GF} = \sup D_F$ if $F \leq_2 G$ and $t_{FG} = \sup D_F, \tilde{t}_{GF} = \inf D_F$ if $G \leq_2 F$. Overall, we have

$$F =_{gs} G \Leftrightarrow F \leq_2 G \text{ or } G \leq_2 F.$$

Note, however, that $t_0 = 0$ is only a meaningful threshold if $0 \in R''_{FG}(D_F)$. If the set $R''_{FG}(D_F)$ has positive Lebesgue-measure, the threshold should indeed lie in its interior so that R''_{FG} is strictly smaller than t_0 at one point and strictly larger than t_0 at another point. Since, for $t_0 = 0$, this contradicts $F <_2 G$ as well as $G <_2 F$, $F =_{gs} G$ is practically equivalent to $F =_2 G$ and, therefore, to R_{FG} being linear. Oja (1981, p. 165) already showed this to be true for $F =_s G$, which is equivalent to $F =_{gs} G$ for $F, G \in \mathcal{S}$.

The situation becomes more complicated in the case $t_0 \neq 0$ in which $F =_{gs}^{t_0} G$ is equivalent to

$$\begin{aligned}
\exists t_{FG}, \tilde{t}_{GF} \in \operatorname{supp}(F) : \frac{R''_{FG}(t)}{t_0} \leq 1 \ \forall t \leq t_{FG}, \frac{R''_{FG}(t)}{t_0} \geq 1 \ \forall t \geq t_{FG},& \\
\frac{R''_{FG}(t)}{t_0} \geq -(R'_{FG}(t))^3 \ \forall t \leq \tilde{t}_{GF} \text{ and } \frac{R''_{FG}(t)}{t_0} \leq -(R'_{FG}(t))^3 \ \forall t \geq \tilde{t}_{GF}.&
\end{aligned}$$

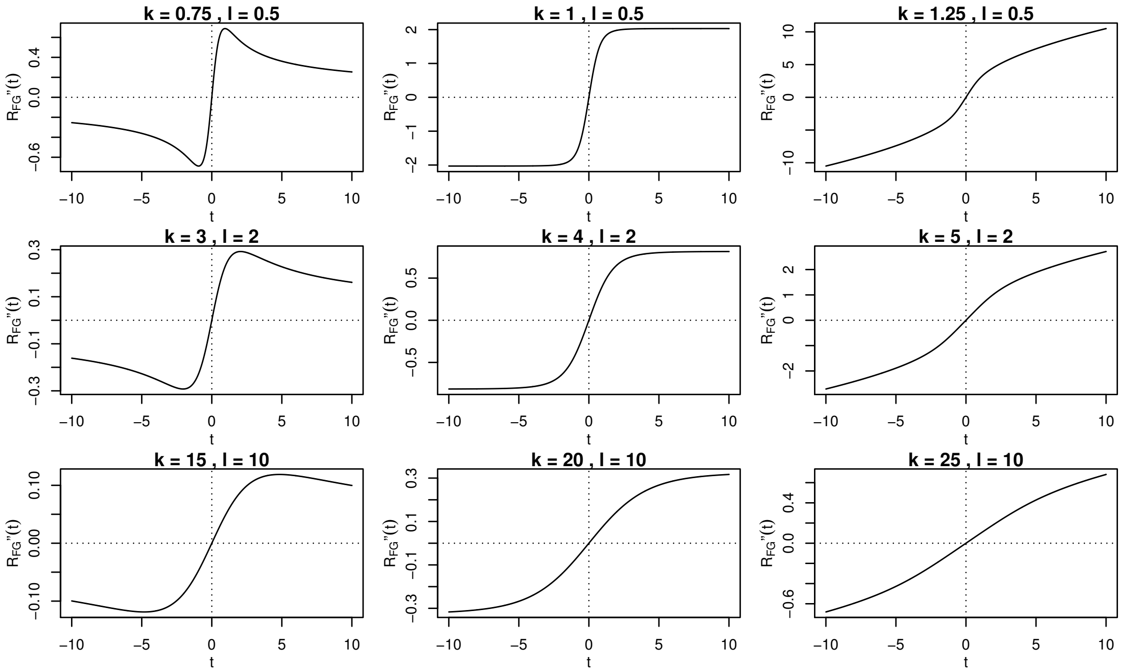

Figure 4.5.: Graphs of R''_{FG} with F and G being the cdf's of t-distributions with k and ℓ degrees of freedom, respectively.

Similarly to the equivalence with respect to $\leq_3$, this is difficult to verify based only on one of the two RIDF's (or its second derivative).

4.2.3. Application to Specific Distributions

Student's t-Distribution

The most popular example of a distribution class with changing kurtosis as its major characteristic probably is the class of Student's t-distributions. All of the distributions are symmetric around zero and exhibit only minor differences in spread. It is popular knowledge that the kurtosis of the distributions declines for increasing degrees of freedom. This behaviour is exemplified by the moment-based kurtosis measure (i.e. the standardized fourth moment), which is equal to $\frac{6}{k-4}$, if $k > 4$ holds for the number k of degrees of freedom.

Two t-distributions with different numbers of degrees of freedom can only be compared numerically with respect to the kurtosis orders $\leq_3$ and $\leq_{gs}$ as there is no explicit representation of its quantile function and, therefore, of the corresponding RIDF. Let $X \sim t_k$ with cdf F and let $Y \sim t_\ell$ with cdf G. The second derivative of the corresponding RIDF R_{FG} is plotted in Figure 4.5 for a number of different parameter values $0 < \ell < k$.

Overall, the steepness of the curves declines with increasing degrees of freedom. While $R''_{FG}(t)$ is increasing for values of t around zero for all considered combinations of k and ℓ, the behaviour for larger absolute values of t seems to depend on the ratio between k and ℓ. For

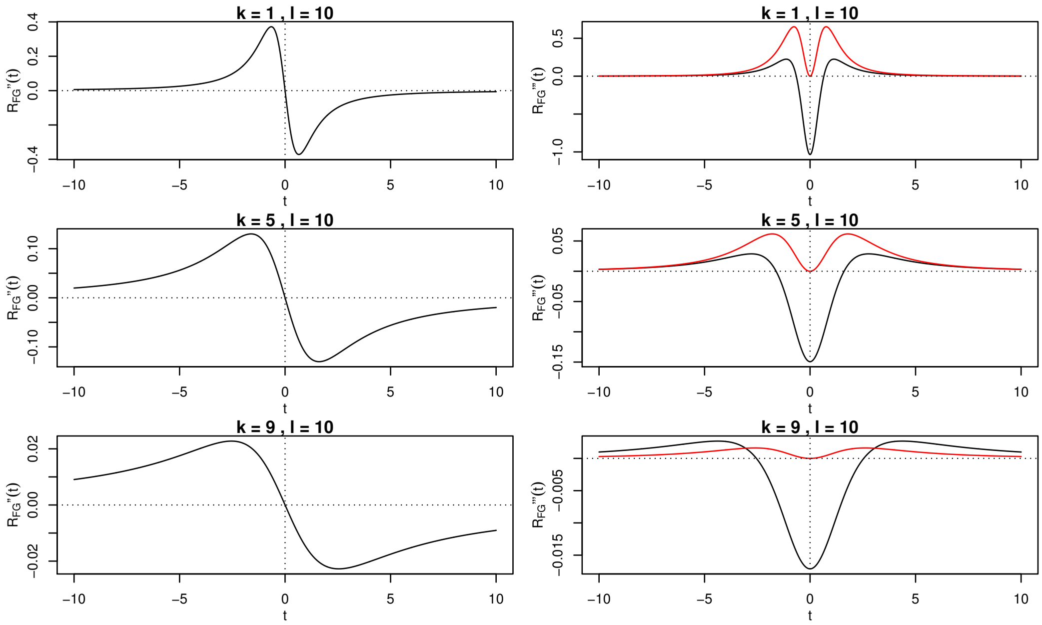

Figure 4.6.: Graphs of R''_{FG} (left panels) and R'''_{FG} (black graphs in the right panels) with F and G being the cdf's of t-distributions with k and ℓ degrees of freedom, respectively. The red graphs in the right panels are of the function $t \mapsto 3\frac{(R''_{FG}(t))^2}{R'_{FG}(t)}$.

$k < 2\ell$, the curve has a negative slope for $|t|$ large enough while the curve is strictly increasing for $k \geq 2\ell$. However, for $k = 2\ell$, the curve flattens for large values of $|t|$, seeming to converge to a horizontal line. We conclude that, while $F \leq_{gs} G$ holds for all $0 < \ell < k$, $F \leq_3 G$ only holds if $2\ell \leq k$.

The case $0 < k < \ell$ is obviously symmetric. Nonetheless, we consider the second and third derivative of R_{FG} for selected parameter values in order to see what can be inferred about the validity of $G \leq_{gs} F$ or $G \leq_3 F$. Here, we use the results from the end of Section 4.2.1 for $\leq_3$ and from the end of Section 4.2.2 for $\leq_{gs}$. The corresponding graphs can be found in Figure 4.6.

Varying the parameter values essentially only changes the steepness of the curves, the basic shape of the graphs stays the same. Since $t_0 = 0$ is a reasonable threshold for the concave-convex order $\leq_{gs}^{t_0}$, the considerations from the end of Section 4.2.2 yield $G \leq_{gs} F$ for all $0 < k < \ell$ because the second derivatives are all positive for negative values of t and negative for positive values of t. Graphical results with respect to the order $\leq_3$ are more difficult to obtain as $G \leq_3 F$ is not simply equivalent to $R'''_{FG} \leq 0$, but to $R'''_{FG}(t) \leq 3\frac{(R''_{FG}(t))^2}{R'_{FG}(t)}$ (see Proposition 4.11). Therefore, the graph of the right side of that inequality (red) is plotted alongside the third derivative (black) in the right panels of Figure 4.6. The resulting plots confirm our observations from Figure 4.5 as the red graph drops below the black graph for $\ell < 2k$ and large values of $|t|$. For $\ell = 2k$, the red graph seems to converge towards the black graph for $|t| \to \infty$.

Weibull Distribution

We now move on to asymmetric distributions, which are a particular focus of the approaches to quantifying kurtosis in this thesis. A rare example of a skewed distribution family with explicit representations of both the cdf and the quantile function is the Weibull distribution. We fix the scale parameter at value 1, so that the cdf and the quantile function are given by

$$F(t) = 1 - \exp(-t^k),\ t > 0, \quad \text{and} \quad F^{-1}(p) = (-\log(1-p))^{1/k},\ p \in (0,1).$$

Now, let $k, \ell > 0$ and let $X \sim \mathrm{Weib}(k)$ with cdf F and let $Y \sim \mathrm{Weib}(\ell)$ with cdf G. The corresponding RIDF is then given by

$$R_{FG}(t) = (-\log(1-(1-\exp(-t^k))))^{1/\ell} = t^{k/\ell}$$

for $t > 0$. It follows directly from Example 4.12 that $F \leq_3 G$ is equivalent to $\frac{k}{\ell} \notin (1,2)$ and that $G \leq_3 F$ is equivalent to $\frac{k}{\ell} \notin (\frac{1}{2}, 1)$. Now let $t_0 \in \mathrm{int}(R''_{FG}(D_F))$. Due to Corollary 4.21, $F \leq_{gs}^{t_0} G$ and $G \leq_{gs}^{t_0} F$ hold under the same conditions, respectively. Since R''_{FG} is decreasing for $\frac{k}{\ell} \in (1,2)$, $F \not\leq_{gs}^{t_0} G$ follows directly. Analogously, $R_{FG}(t) = t^{\ell/k}, t > 0$, implies that $G \not\leq_{gs}^{t_0} F$ holds for $\frac{k}{\ell} \in (\frac{1}{2}, 1)$. Our results concerning the Weibull distribution can be summarized by the following distinction by cases:

$$\text{(i)} \quad \text{If } \frac{k}{\ell} \in \left(\frac{1}{2}, 1\right), \qquad \text{then } F \leq_3 G \text{ and } F \leq_{gs}^{t_0} G, \tag{4.10}$$

$$\text{(ii)} \quad \text{If } \frac{k}{\ell} \in (1,2), \qquad \text{then } G \leq_3 F \text{ and } G \leq_{gs}^{t_0} F, \tag{4.11}$$

$$\text{(iii)} \quad \text{If } \frac{k}{\ell} \in \left(0, \frac{1}{2}\right] \cup \{1\} \cup [2,\infty), \text{ then } F =_3 G \text{ and } F =_{gs}^{t_0} G, \tag{4.12}$$

where t_0 is always chosen as a reasonable threshold. Cases (i) and (ii) remain true if the non-strict order $\leq_3$ and $\leq_{gs}^{t_0}$ are replaced by their respective strict versions $<_3$ and $<_{gs}^{t_0}$.

This distinction by cases allows us to construct another counterexample for the transitivity of both $\leq_3$ and $\leq_{gs}^{t_0}$. For all $k > 0$, denote the cdf of $X \sim \mathrm{Weib}(k)$ by F_k. Then, e.g.,

$$F_k \leq_3 F_{1.5k} =_3 F_{0.7k}$$

holds. Now $\leq_3$ being transitive is contradicted by $F_k \not\leq_3 F_{0.7k}$.

Gamma Distribution

We consider the family of gamma distributions with the scale (or rate) parameter fixed to 1 with the density function

$$f(t) = \frac{t^{k-1}e^{-t}}{\Gamma(k)}, \quad t > 0,$$

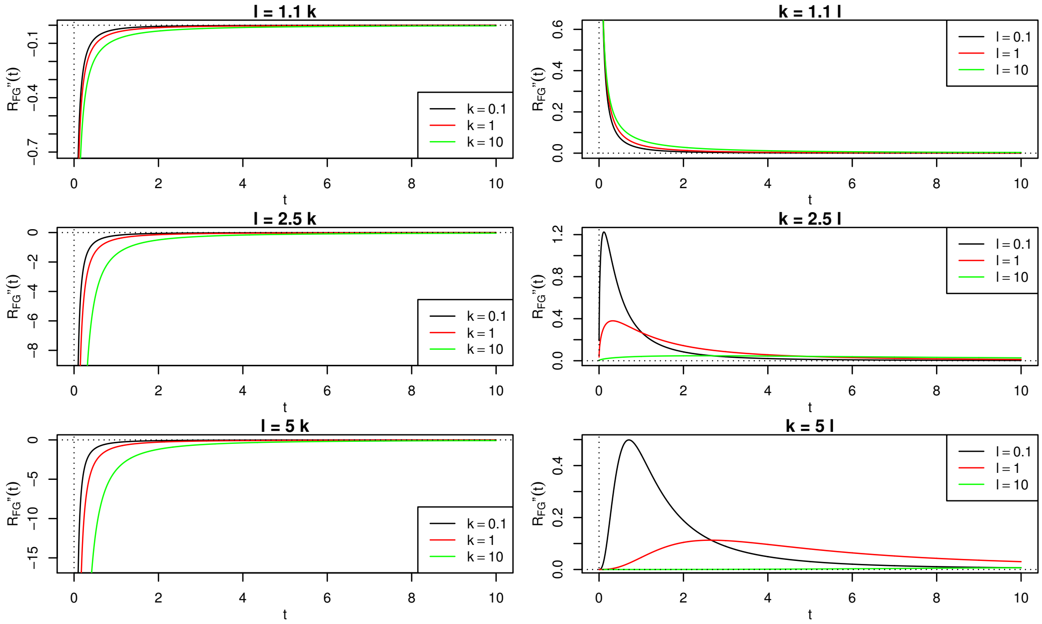

Figure 4.7.: Graphs of R''_{FG} with F and G being the cdf's of gamma distributions with shape parameter values k and ℓ, respectively.

where $k > 0$ denotes the shape parameter and $\Gamma(\cdot)$ denotes the gamma function. Now let $X \sim \Gamma(k)$ with cdf F and $Y \sim \Gamma(\ell)$ with cdf G, where $k, \ell > 0$. As for the t-distribution, the RIDF R_{FG} does not have a explicit representation in this case and we therefore rely solely on graphical considerations. For any parameter choice $0 < k < \ell$, the second derivative R''_{FG} is a strictly negative, increasing and concave function with its gradient tending towards 0 for increasing t (see left panels of Figure 4.7). It follows directly that $F \leq_3 G$ holds as well as $F \leq_{gs}^{t_0} G$ for all reasonable thresholds t_0.

The behaviour of R''_{FG} in the case $0 < \ell < k$ is dependent upon a further distinction (see right panels of Figure 4.7). If $k < 2\ell$, then the second derivative is decreasing and convex; if $k > 2\ell$, then it starts out by increasing, reaches a maximum and decreases towards 0 for large values of t. In both cases, the function is strictly positive. Additionally, both cases directly contradict both $F \leq_3 G$ and $F \leq_{gs}^{t_0} G$ for any $t_0 \in \operatorname{int}(R''_{FG}(D_F))$.

Overall, if t_0 is chosen to be a reasonable threshold, $F \leq_3 G$ is equivalent to both $F \leq_{gs}^{t_0} G$ and $k < \ell$. Furthermore, if we exclude the case $F = G$, it is also equivalent to both $F \nleq_2 G$ and $G \leq_2 F$ (see van Zwet, 1964, pp. 60–61).

Sinh-Arsinh Distribution

The family of sinh-arsinh distributions was introduced by Jones and Pewsey (2009). It is dependent upon four parameters, which can be associated with location, dispersion, skewness and kurtosis (although Jones and Pewsey state throughout their work that the fourth parameter is associated with tailweight). Here, we consider a simplified two-parameter family by fixing the location and spread parameters to zero and one, respectively. A random variable X is said to be sinh-arsinh-distributed with skewness parameter $\nu \in \mathbb{R}$ and kurtosis parameter $\tau > 0$, denoted by $X \sim \mathrm{SAS}(\nu, \tau)$, if the random variable

$$Z = S_{\nu,\tau}(X) = \sinh(\tau \cdot \operatorname{arsinh}(X) - \nu)$$

is standard normal. Skewness to the right increases with increasing ν and kurtosis decreases with increasing τ. More specifically, $F \leq_2 G$ if $\nu_F \leq \nu_G, \tau_F = \tau_G$ and $F \leq_{gs} G$ if $\nu_F = \nu_G = 0, \tau_F \leq \tau_G$ (see Jones and Pewsey, 2009, pp. 763, 765, 766). One can directly infer the corresponding cdf $F = \Phi \circ S_{\nu,\tau}$ and quantile function $F^{-1} = S_{\nu,\tau}^{-1} \circ \Phi^{-1} = S_{-\frac{\nu}{\tau},\frac{1}{\tau}} \circ \Phi^{-1}$ of X. Here, Φ denotes the cdf of the standard normal distribution and the generally valid identity $S_{\nu,\tau}^{-1} = S_{-\frac{\nu}{\tau},\frac{1}{\tau}}$ for the inverse transformation was used. The corresponding density also has a (slightly more complex) explicit representation (see Jones and Pewsey, 2009, p. 762).

There exist multiple other distribution families with four parameters that are associated with the first four convex characteristics of location, dispersion, skewness and kurtosis (or, alternatively, peakedness or tailweight). Examples include the skew-normal distributions by Azzalini and subsequent skew-t distributions (Azzalini, 1985; Azzalini and Capitanio, 2003) and Tukey's g-and-h or g-and-k distributions (Hoaglin, 1985; MacGillivray and Cannon, 1997; Tukey, 1977). However, these other examples do not have similarly explicit representations of their cdf's, quantile functions, densities and RIDF's. Furthermore, while the skew-t distributions do include the standard normal distribution, it only appears as a limiting case and not as a standard case as for the sinh-arsinh distributions. Finally, the sinh-arsinh transformation can also be applied to (symmetric) base distributions other than the standard normal. For example, Rosco et al. (2011) applied it to Student's t-distribution.

Let $X \sim \mathrm{SAS}(\nu_F, \tau_F)$ with cdf F and $Y \sim \mathrm{SAS}(\nu_G, \tau_G)$ with cdf G. Then, the corresponding RIDF is given by

$$\begin{aligned}
R_{FG}(t) = G^{-1}(F(t)) &= (S_{\nu_G,\tau_G}^{-1} \circ \Phi^{-1} \circ \Phi \circ S_{\nu_F,\tau_F})(t) = S_{-\frac{\nu_G}{\tau_G},\frac{1}{\tau_G}}(S_{\nu_F,\tau_F}(t)) \\
&= \sinh\left(\frac{\operatorname{arsinh}(\sinh(\tau_F \cdot \operatorname{arsinh}(t) - \nu_F)) + \nu_G}{\tau_G}\right) \\
&= \sinh\left(\frac{\tau_F}{\tau_G} \cdot \operatorname{arsinh}(t) - \frac{\nu_F - \nu_G}{\tau_G}\right) \\
&= S_{\frac{\nu_F - \nu_G}{\tau_G},\frac{\tau_F}{\tau_G}}(t).
\end{aligned}$$

Note that the fulfilment of $F \leq_{gs}^{t_0} G$ and $F \leq_3 G$ is solely dependent on this RIDF. Hence, the ordering of F and G in terms of kurtosis is only dependent upon two parameters instead of four. Particularly, it is independent of the concrete values of the skewness parameters and instead only depends on their difference. Define $\tilde{\nu} = \frac{\nu_F - \nu_G}{\tau_G}$ and $\tilde{\tau} = \frac{\tau_F}{\tau_G}$ as well as $C_{\nu,\tau}(t) = \cosh(\tau \cdot \operatorname{arsinh}(t) - \nu)$ for $\nu, t \in \mathbb{R}, \tau > 0$ and note that $C_{\nu,\tau}^2 - S_{\nu,\tau}^2 \equiv 1$ for all $\nu \in \mathbb{R}, \tau > 0$. The following result gives equivalent conditions on when sinh-arsinh distributions are ordered with respect to the kurtosis orders $\leq_3$ and $\leq_{gs}$.

Theorem 4.24. *Let $F \neq G$.*

a) $F \leq_3 G$ is equivalent to $\tau_F \geq 2\tau_G$

Now, let $t_0 \in \operatorname{int}(R''_{FG}(D_F))$

b) If $t_0 \neq 0$, then $F \leq_{gs}^{t_0} G$ is also equivalent to $\tau_F \geq 2\tau_G$.

c) If $t_0 = 0$, $F \leq_{gs}^{t_0} G$ is equivalent to $\tau_F > \tau_G$.

Proof. In order to prove both a) and b), we assume $t_0 \in \operatorname{int}(R''_{FG}(D_F)) \setminus \{0\}$ and we prove the chain of implications $F \leq_{gs}^{t_0} G \Rightarrow \tau_F \geq 2\tau_G \Rightarrow F \leq_3 G$. Since $F \leq_3 G \Rightarrow F \leq_{gs}^{t_0} G$ holds due to Corollary 4.21, all three statements are then equivalent, which corresponds to parts a) and b). Note that the first three derivatives of the RIDF are given by

$$\begin{aligned}
R'_{FG}(t) &= \tilde{\tau}(1+t^2)^{-1/2} C_{\tilde{\nu},\tilde{\tau}}(t),\\
R''_{FG}(t) &= \tilde{\tau}(1+t^2)^{-3/2} \left[\tilde{\tau}\sqrt{1+t^2} S_{\tilde{\nu},\tilde{\tau}}(t) - tC_{\tilde{\nu},\tilde{\tau}}(t)\right],\\
R'''_{FG}(t) &= \tilde{\tau}(1+t^2)^{-5/2} \left[-3\tilde{\tau}t\sqrt{1+t^2} S_{\tilde{\nu},\tilde{\tau}}(t) + \left((\tilde{\tau}^2+2)t^2 + \tilde{\tau}^2 - 1\right) C_{\tilde{\nu},\tilde{\tau}}(t)\right].
\end{aligned}$$

First, we show the implication $F \leq_{gs}^{t_0} G \Rightarrow \tau_F \geq 2\tau_G$ by contradiction. In order to obtain the asymptotic behaviour of $S_{\tilde{\nu},\tilde{\tau}}(t)$, we rewrite it as follows

$$\begin{aligned}
S_{\tilde{\nu},\tilde{\tau}}(t) &= \sinh(\tilde{\tau} \cdot \log(t + \sqrt{1+t^2}) - \tilde{\nu})\\
&= \tfrac{1}{2}\left[\exp\left(\tilde{\tau} \cdot \log(t+\sqrt{1+t^2}) - \tilde{\nu}\right) - \exp\left(-\tilde{\tau} \cdot \log(t+\sqrt{1+t^2}) + \tilde{\nu}\right)\right]\\
&= \frac{1}{2}\left[\frac{(t+\sqrt{1+t^2})^{\tilde{\tau}}}{e^{\tilde{\nu}}} - \frac{e^{\tilde{\nu}}}{(t+\sqrt{1+t^2})^{\tilde{\tau}}}\right].
\end{aligned}$$

Since $\tilde{\tau} > 0$, the second summand converges to zero as $t \to \infty$. The first summand is obviously positive and diverges; asymptotically it behaves like $(2t)^{\tilde{\tau}}/e^{\tilde{\nu}}$. Overall, $S_{\tilde{\nu},\tilde{\tau}}(t) \sim 2^{\tilde{\tau}-1} e^{-\tilde{\nu}} |t|^{\tilde{\tau}}$ for $t \to \infty$. For the asymptotic behaviour as $t \to -\infty$, note that $t + \sqrt{1+t^2}$ behaves for $t \to -\infty$ as

$$\sqrt{1+t^2} - t = \frac{(\sqrt{1+t^2}-t)(\sqrt{1+t^2}+t)}{\sqrt{1+t^2}+t} = \frac{1}{\sqrt{1+t^2}+t} \sim (2t)^{-1}$$

behaves for $t \to \infty$. With similar reasoning as before, we obtain $S_{\tilde{\nu},\tilde{\tau}}(t) \sim -2^{\tilde{\tau}-1} e^{\tilde{\nu}} |t|^{\tilde{\tau}}$ for $t \to -\infty$. The relationship between the hyperbolic functions now gives $C_{\tilde{\nu},\tilde{\tau}}(t) \sim S_{\tilde{\nu},\tilde{\tau}}(t)$ for $t \to \infty$ and $C_{\tilde{\nu},\tilde{\tau}}(t) \sim -S_{\tilde{\nu},\tilde{\tau}}(t)$ for $t \to -\infty$. Overall, we infer

$$R''_{FG}(t) \sim \tilde{\tau}|t|^{-3}\left[\tilde{\tau}|t| S_{\tilde{\nu},\tilde{\tau}}(t) - tC_{\tilde{\nu},\tilde{\tau}}(t)\right] \sim \tilde{\tau}(\tilde{\tau}-1)\frac{S_{\tilde{\nu},\tilde{\tau}}(t)}{|t|^2}$$

$$\sim \begin{cases} \tilde{\tau}(\tilde{\tau}-1)2^{\tilde{\tau}-1}e^{-\tilde{\nu}}|t|^{\tilde{\tau}-2} & \text{for } t \to \infty, \\ -\tilde{\tau}(\tilde{\tau}-1)2^{\tilde{\tau}-1}e^{\tilde{\nu}}|t|^{\tilde{\tau}-2} & \text{for } t \to -\infty, \end{cases} \tag{4.13}$$

if $\tilde{\tau} \neq 1$. In the case $\tilde{\tau} = 1$, the asymptotically leading summands of $\sqrt{1+t^2}S_{\tilde{\nu},\tilde{\tau}}(t)$ and $tC_{\tilde{\nu},\tilde{\tau}}(t)$ cancel out and, therefore, a closer investigation is required. Specifically,

$$\begin{aligned}
&\tilde{\tau}\sqrt{1+t^2}S_{\tilde{\nu},\tilde{\tau}}(t) - tC_{\tilde{\nu},\tilde{\tau}}(t) \\
&= \frac{\sqrt{1+t^2}}{2}\left[\frac{(t+\sqrt{1+t^2})^{\tilde{\tau}}}{e^{\tilde{\nu}}} - \frac{e^{\tilde{\nu}}}{(t+\sqrt{1+t^2})^{\tilde{\tau}}}\right] - \frac{t}{2}\left[\frac{(t+\sqrt{1+t^2})^{\tilde{\tau}}}{e^{\tilde{\nu}}} + \frac{e^{\tilde{\nu}}}{(t+\sqrt{1+t^2})^{\tilde{\tau}}}\right] \\
&= \frac{1}{2}\left[\frac{\sqrt{1+t^2}^2 - t^2}{e^{\tilde{\nu}}} - \frac{(t+\sqrt{1+t^2})e^{\tilde{\nu}}}{t+\sqrt{1+t^2}}\right] \\
&= \frac{1-e^{2\tilde{\nu}}}{2e^{\tilde{\nu}}}
\end{aligned}$$

yields

$$R''_{FG}(t) = \frac{1-e^{2\tilde{\nu}}}{2e^{\tilde{\nu}}}(1+t^2)^{-3/2} \sim \frac{1-e^{2\tilde{\nu}}}{2e^{\tilde{\nu}}}|t|^{-3} \tag{4.14}$$

for $\tilde{\tau} = 1$. Now, assuming $\tilde{\tau} < 2$, it follows that $R''_{FG}(t) \overset{|t|\to\infty}{\to} 0$. If $t_0 > 0$, $R''_{FG}(t_u) < t_0$ follows for t_u large enough. However, since t_0 lies in the interior of the image of R''_{FG}, there also exists a $t_\ell < t_u$ such that $R''_{FG}(t_\ell) > t_0$. This contradicts $F \leq_{gs}^{t_0} G$. If $t_0 < 0$, $R''_{FG}(s_\ell) > t_0$ follows for s_ℓ small enough and, by assumption, there also exists an $s_u > s_\ell$ such that $R''_{FG}(s_u) < t_0$, thus also contradicting $F \leq_{gs}^{t_0} G$.

We now prove the implication $\tau_F \geq 2\tau_G \Rightarrow F \leq_3 G$ and therefore assume $\tilde{\tau} \geq 2$. $F \leq_3 G$ is equivalent to

$$[(\tilde{\tau}^2+2)t^2 + \tilde{\tau}^2 - 1]C_{\tilde{\nu},\tilde{\tau}}(t) \geq 3\tilde{\tau}t\sqrt{1+t^2}S_{\tilde{\nu},\tilde{\tau}}(t) \tag{4.15}$$

holding for all $t \in \mathbb{R}$. Because of $C_{\tilde{\nu},\tilde{\tau}}(t) \geq 0$ and $(\tilde{\tau}^2+2)t^2 + \tilde{\tau}^2 - 1 \geq 6t^2 + 3 > 0$, the left hand side of inequality (4.15) is positive for all t. Hence, substituting both sides of the inequality with their squares gives a sufficient condition. We obtain

$$\begin{aligned}
&\Big[(\tilde{\tau}^2+2)t^2 + (\tilde{\tau}^2-1)\Big]^2 C^2_{\tilde{\nu},\tilde{\tau}}(t) \geq 9\tilde{\tau}^2t^2(1+t^2)(C^2_{\tilde{\nu},\tilde{\tau}}(t)-1) \qquad \forall t \in \mathbb{R} \\
\Leftrightarrow\ &\Big[\Big((\tilde{\tau}^2+2)^2 - 9\tilde{\tau}^2\Big)t^4 + \Big(2(\tilde{\tau}^2+2)(\tilde{\tau}^2-1) - 9\tilde{\tau}^2\Big)t^2 + (\tilde{\tau}^2-1)^2\Big] C^2_{\tilde{\nu},\tilde{\tau}}(t) \\
&\qquad + 9\tilde{\tau}^2t^2(1+t^2) \geq 0\ \forall t \in \mathbb{R}.
\end{aligned}$$

The second summand on the left hand side is obviously non-negative. It is now sufficient to show that all coefficients of the polynomial, with which $C^2_{\tilde{\nu},\tilde{\tau}}(t)$ is multiplied, are non-negative. For the constant $(\tilde{\tau}^2-1)^2$, this is obvious. The coefficient of t^2 is equal to $2\tilde{\tau}^4 - 7\tilde{\tau}^2 - 4 = (\tilde{\tau}^2-4)(2\tilde{\tau}^2+1)$, which is non-negative since $\tilde{\tau} \geq 2$ was assumed. The same is true for the coefficient of t^4, which is equal to $\tilde{\tau}^4 - 5\tilde{\tau}^2 + 4 = (\tilde{\tau}^2-4)(\tilde{\tau}^2-1)$.

It remains to prove part c), so let now $t_0 = 0$. Note that the sign of $R''_{FG}(t)$ corresponds to the sign of $h(t) = \tilde{\tau}\sqrt{1+t^2}S_{\tilde{\nu},\tilde{\tau}}(t) - tC_{\tilde{\nu},\tilde{\tau}}(t), t \in \mathbb{R}$. Using (4.13), (4.14) as well as $h(t) = \frac{(1+t^2)^{3/2}}{\tilde{\tau}} R''_{FG}(t), t \in \mathbb{R}$, we obtain that

$$h(t) \sim \begin{cases} (\tilde{\tau}-1)2^{\tilde{\tau}-1}e^{-\tilde{\nu}}|t|^{\tilde{\tau}+1} & \text{for } t \to \infty, \\ -(\tilde{\tau}-1)2^{\tilde{\tau}-1}e^{\tilde{\nu}}|t|^{\tilde{\tau}+1} & \text{for } t \to -\infty, \end{cases} \tag{4.16}$$

for $\tilde{\tau} \neq 1$ and

$$h(t) = \frac{1-e^{2\tilde{\nu}}}{2e^{\tilde{\nu}}}, \quad t \in \mathbb{R},$$

for $\tilde{\tau} = 1$. From the latter, we infer that either $R''_{FG} \geq 0$ (in the case $\tilde{\nu} < 0$) or $R''_{FG} \leq 0$ (in the case $\tilde{\nu} > 0$) holds. (Note that the case $\tilde{\nu} = 0$ is excluded due to the assumption $F \neq G$.) While this yields $F \leq^0_{gs} G$ for $\tilde{\tau} = 1$, the threshold $t_0 = 0$ does not satisfy $t_0 \in \operatorname{int}(R''_{FG}(D_F))$, which is assumed in the result. Continuing under the assumption $\tilde{\tau} \neq 1$, (4.16) yields

$$\lim_{t\to\pm\infty} h(t) = \begin{cases} \pm\infty & \text{for } \tilde{\tau} > 1, \\ \mp\infty & \text{for } \tilde{\tau} < 1. \end{cases} \tag{4.17}$$

Considering $S'_{\tilde{\nu},\tilde{\tau}}(t) = \tilde{\tau}(1+t^2)^{-1/2}C_{\tilde{\nu},\tilde{\tau}}(t)$ and $C'_{\tilde{\nu},\tilde{\tau}}(t) = \tilde{\tau}(1+t^2)^{-1/2}S_{\tilde{\nu},\tilde{\tau}}(t)$, the derivative of h is given by

$$\begin{aligned} h'(t) &= \tilde{\tau}t(1+t^2)^{-1/2}S_{\tilde{\nu},\tilde{\tau}}(t) + \tilde{\tau}(1+t^2)^{1/2}S'_{\tilde{\nu},\tilde{\tau}}(t) - C_{\tilde{\nu},\tilde{\tau}}(t) - tC'_{\tilde{\nu},\tilde{\tau}}(t) \\ &= \tilde{\tau}t(1+t^2)^{-1/2}S_{\tilde{\nu},\tilde{\tau}}(t) + \tilde{\tau}^2C_{\tilde{\nu},\tilde{\tau}}(t) - C_{\tilde{\nu},\tilde{\tau}}(t) - \tilde{\tau}t(1+t^2)^{-1/2}S_{\tilde{\nu},\tilde{\tau}}(t) \\ &= (\tilde{\tau}^2-1)C_{\tilde{\nu},\tilde{\tau}}(t). \end{aligned}$$

for $t \in \mathbb{R}$. Because of $C_{\tilde{\nu},\tilde{\tau}}(t) > 0$, we have $h' > 0$ for $\tilde{\tau} > 1$ and $h' < 0$ for $\tilde{\tau} < 1$. Combined with (4.17), it follows that h has exactly one root, at which its sign changes from '$-$' to '$+$' if $\tilde{\tau} > 1$ and from '$+$' to '$-$' if $\tilde{\tau} < 1$. Since the sign of R''_{FG} coincides with the sign of h, it follows directly that $F \leq^0_{gs} G$ holds for $\tilde{\tau} > 1$ and that the same does not hold for $\tilde{\tau} < 1$. ▨

Remark 4.25. It follows from $F \neq G$ that $\operatorname{int}(R''_{FG}(D_F)) \neq \emptyset$, so that parts b) and c) of Theorem 4.24 are not statements concerning the empty set. We prove this by contradiction and therefore assume $\operatorname{int}(R''_{FG}(D_F)) = \emptyset$. Since R''_{FG} is continuous, this occurs if and only if R''_{FG} is constant. Defining the function $h(t) = \tilde{\tau}\sqrt{1+t^2}S_{\tilde{\nu},\tilde{\tau}}(t) - tC_{\tilde{\nu},\tilde{\tau}}(t), t \in \mathbb{R}$, as in the proof of Theorem 4.24, this is equivalent to the existence of a constant $c \in \mathbb{R}$ such that $h(t) = c(1+t^2)^{3/2}, t \in \mathbb{R}$. The case $c = 0$ is equivalent to $F = G$ as h is not constant for $\tilde{\tau} \neq 1$ and non-zero for $\tilde{\tau} = 1$ and $\tilde{\nu} \neq 0$. In the case $c \neq 0$, we either obtain $\lim_{t\to\pm\infty} h(t) = \infty$ (for $c > 0$) or $\lim_{t\to\pm\infty} h(t) = -\infty$ (for $c < 0$), which contradicts (4.17) in combination with the fact that h is constant for $\tilde{\tau} = 1$.

Since the orders used in the equivalent conditions in Theorem 4.24 are transitive, the

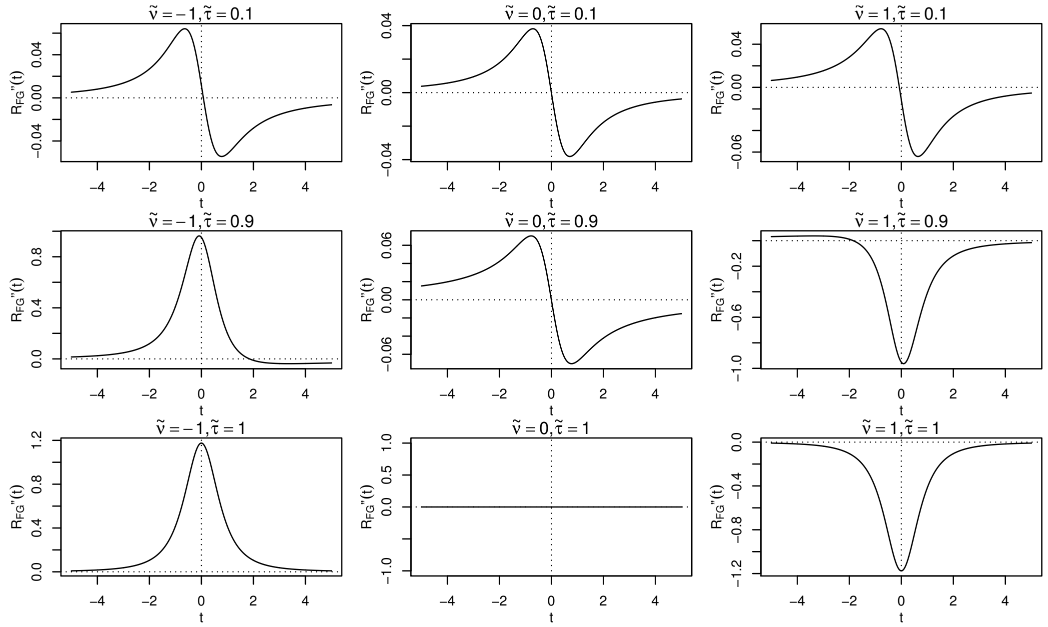

Figure 4.8.: Graphs of R''_{FG} with F and G being the cdf's of $X \sim \mathrm{SAS}(\nu_F, \tau_F)$ and $Y \sim \mathrm{SAS}(\nu_G, \tau_G)$, respectively.

following result is directly implied.

Corollary 4.26. *Let $t_0 \in \operatorname{int}(R''_{FG}(D_F))$. Then, the orders $\leq_3$ and $\leq_{gs}^{t_0}$ are both transitive on the set $\{F \in \mathcal{P}_I^3 : \exists \nu \in \mathbb{R}, \tau > 0 : F = \mathrm{SAS}(\nu, \tau)\}$.*

For a number of choices of $\tilde{\nu}$ and $\tilde{\tau}$, the function R''_{FG} is plotted in Figures 4.8, 4.9 and 4.10. Additionally, a number of properties are summarized in Table 4.1. It is obvious from (4.13) and (4.16) that R''_{FG} asymptotically always behaves like a monomial, where the exponent is linearly increasing in $\tilde{\tau}$ (except for the case $\tilde{\tau} = 1$). The exponent reaches the value 0 for $\tilde{\tau} = 2$, which corresponds to the fact that $F \leq_3 G$ is equivalent to $\tilde{\tau} \geq 2$. Furthermore, the function has exactly one root for $\tilde{\tau} \neq 1$ with the direction of the sign change switching for $\tilde{\tau} = 1$. The graph of R''_{FG} is point symmetric around the origin for $\tilde{\nu} = 0$. For $\tilde{\nu} < 0$, the side with the positive values of R''_{FG} is scaled up and the other side is scaled down. Additionally, the sole root of the function shifts to the side with the scaled-down values. The reverse is true for $\tilde{\nu} > 0$ with the extent of the rescaling and the shift exponentially depending on the absolute value of $\tilde{\nu}$.

In the symmetric case of $\tilde{\nu} = 0$, a number of special cases stand out, which are also singled out in Table 4.1. First, for $\tilde{\tau} = 1$, $R''_{FG} \equiv 0$ obviously holds since $F = G$ and, therefore, R_{FG} is the identity function (see lower central panel of Figure 4.8). Then, for $\tilde{\tau} = 2$, the rather simple form $R_{FG}(t) = 2t\sqrt{t^2+1}$ is obtained, yielding the second derivative $R''_{FG}(t) = \frac{4t^3+6t}{(t^2+1)^{3/2}}$, which converges to 4 as $t \to \infty$ and to -4 as $t \to -\infty$ (see lower central panel of Figure 4.9). Finally, for $\tilde{\tau} = 3$, the RIDF is given by $R_{FG}(t) = 4t^3 + 3t$, which leads to the linear second derivative $R''_{FG}(t) = 24t$ (see central panel of Figure 4.10).

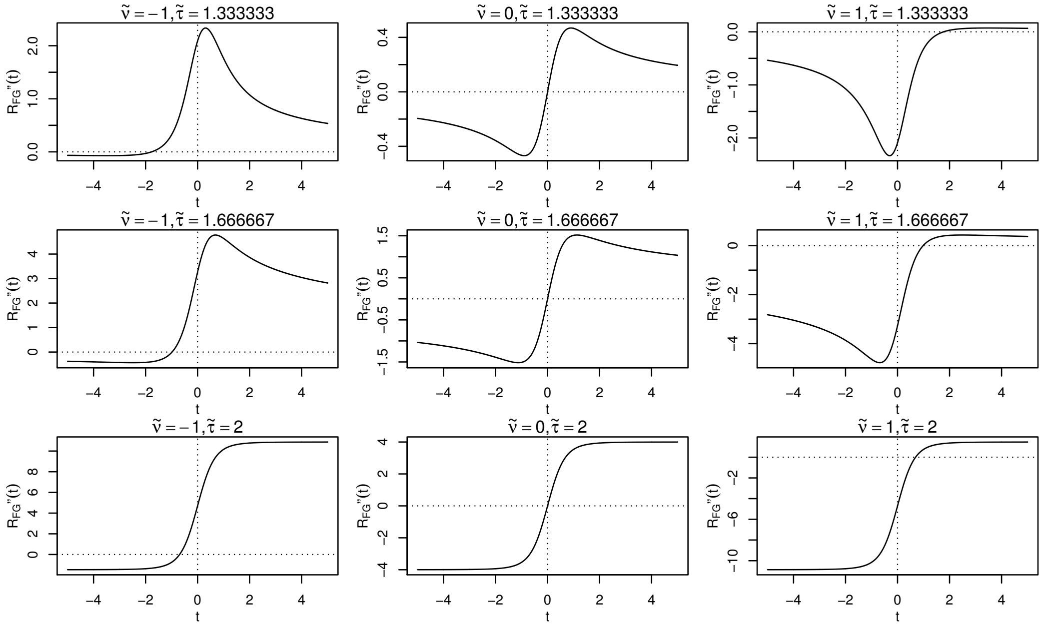

Figure 4.9.: Graphs of R''_{FG} with F and G being the cdf's of $X \sim \mathrm{SAS}(\nu_F, \tau_F)$ and $Y \sim \mathrm{SAS}(\nu_G, \tau_G)$, respectively.

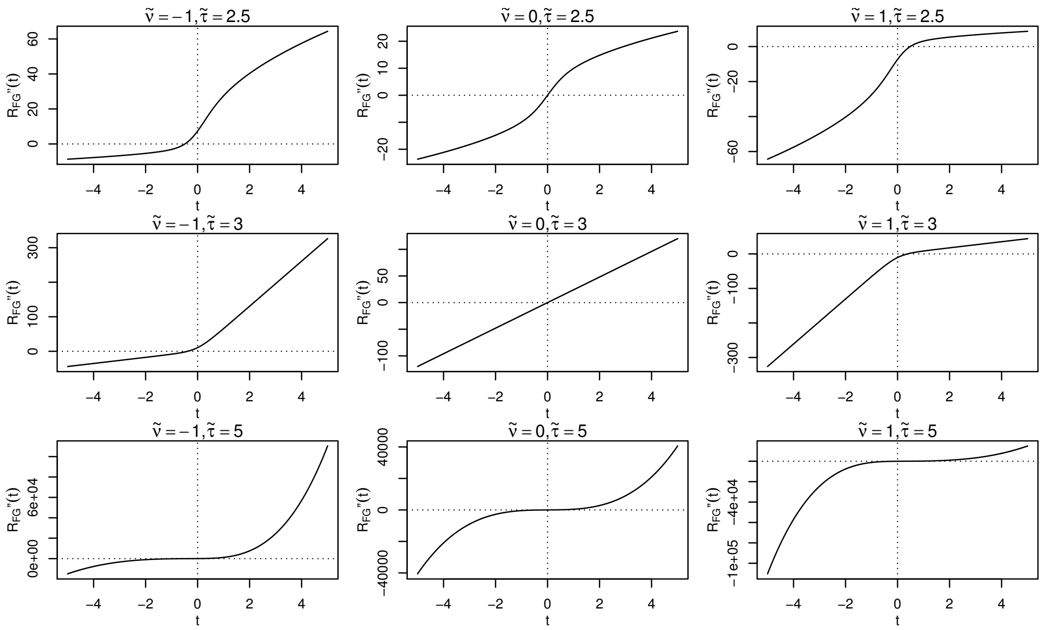

Figure 4.10.: Graphs of R''_{FG} with F and G being the cdf's of $X \sim \mathrm{SAS}(\nu_F, \tau_F)$ and $Y \sim \mathrm{SAS}(\nu_G, \tau_G)$, respectively.

	$R''_{FG}(t)$				
Values of $\tilde{\tau}$	Sign change?	Monotonicity?	$\lim_{t\to\pm\infty}$	$F \leq_3 G$?	$F \leq_{gs}^{t_0} G$?
$\tilde{\tau} \in (0,1)$	'+' to '−'	No	0	No	No
$\tilde{\tau} = 1$	No	No*	0	No**	No**
$\tilde{\tau} \in (1,2)$	'−' to '+'	No	0	No	Iff $t_0 = 0$
$\tilde{\tau} = 2$	'−' to '+'	Increasing	± 4	Yes	Yes
$\tilde{\tau} \in (2,3)$	'−' to '+'	Increasing	$\pm\infty$ sub-linear growth	Yes	Yes
$\tilde{\tau} = 3$	'−' to '+'	Increasing	$\pm\infty$ linear growth	Yes	Yes
$\tilde{\tau} \in (3,\infty)$	'−' to '+'	Increasing	$\pm\infty$ super-linear growth	Yes	Yes

Table 4.1.: Behaviour of the second derivative of the RIDF and kurtosis orders for cdf F of $X \sim \text{SAS}(\nu_F, \tau_F)$ and cdf G of $Y \sim \text{SAS}(\nu_G, \tau_G)$. *: Constant if $\tilde{\nu} = 0$. **: Yes if $\tilde{\nu} = 0$.

Heuristically, Theorem 4.24 states that, within the family of sinh-arsinh distributions, comparisons in terms of kurtosis are skewness-invariant. This is due to the fact that equivalent characterizations for both major kurtosis orders are independent of both ν_F and ν_G, which are skewness parameters by construction and also in the sense of $\leq_2$ for $\tau_F = \tau_G$ (see Jones and Pewsey, 2009, p. 763). Moreover, the characterizations from Theorem 4.24 not only stay the same for equally skewed asymmetric distributions, but also for pairs of distributions with arbitrarily big differences in skewness (within the bounds of the distribution family). Also note that these results can be generalized to families of sinh-arsinh distributions that arise from symmetric base distributions other than the normal one since the RIDF's only depend on the transformations and not on the specific base distribution. This observation serves as a strong argument for considering the notion of kurtosis irrespective of skewness, in particular using the two kurtosis orders used in Theorem 4.24.

The skewness-invariance of the sinh-arsinh distribution in terms of kurtosis was noted by Jones et al. (2011, pp. 91–92). Specifically, they showed that quantile-based kurtosis measures that are constructed from symmetric differences of the form $F^{-1}(1-\alpha) - F^{-1}(\alpha), \alpha \in (0, \frac{1}{2})$, are invariant under changes of the skewness parameter ν. Theorem 4.24 generalizes this skewness-invariance from a specific family of kurtosis measures to the underlying kurtosis orders. Quantile-based kurtosis measures are discussed along with other types of measures in Section 4.3. However, the focus there is laid on the transitivity sets established in Sections 4.2.1 and 4.2.2.

4.2.4. Graphical Interpretation of the Kurtosis Orders

In conclusion of our discussion of the different kurtosis orders, we look at how they detect gradients in kurtosis between two distributions and how this is compatible with heuristic

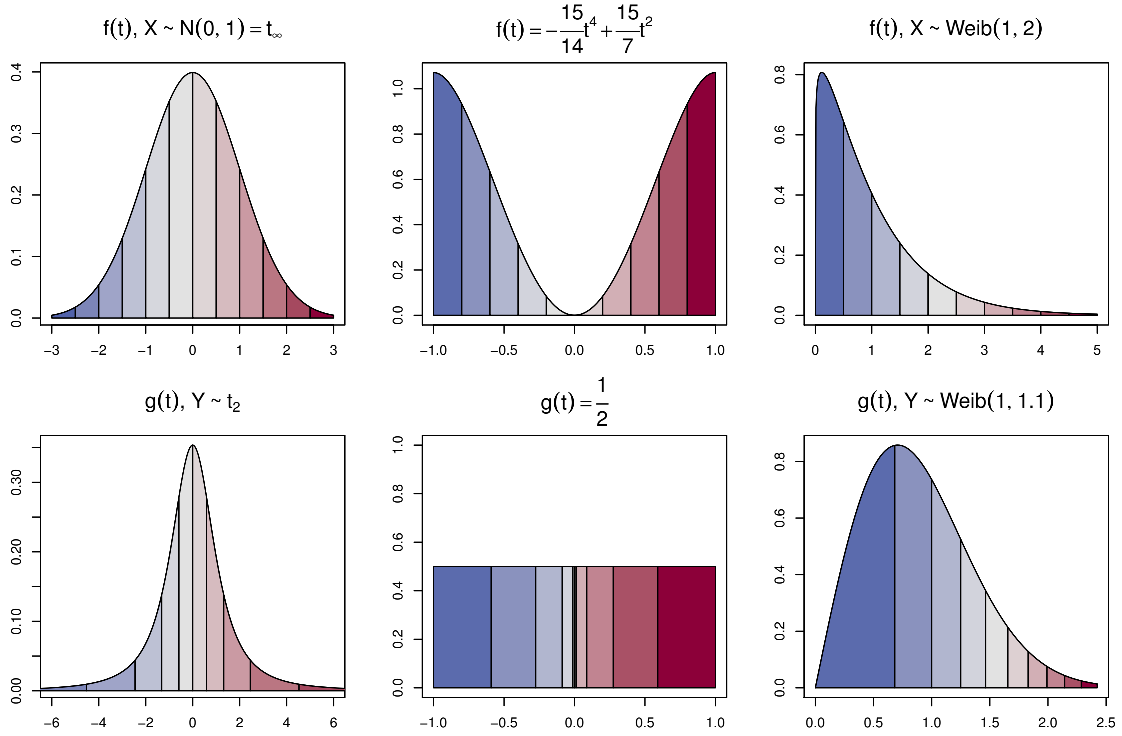

Figure 4.11.: Illustration of three pairs of distributions that are compared with respect to kurtosis in Section 4.2.4.

ideas of what kurtosis is. Consider Figure 4.11 for three examples. These are illustrated in the same way as the notions of location, dispersion and skewness are illustrated in Figure 3.1. The density f of F is depicted in the upper panel and the x-axis is divided in equidistant intervals. The x-axis of the density g of G in the lower panel is also divided into intervals, which are chosen in such a way that the probability mass on each interval is the same as the probability mass on the corresponding interval of f. This interval bounds for g are obtained by plugging the interval bounds of f into the RIDF R_{FG}.

Recall the basic ideas behind the transformations for the first three convex characteristics: for location, all interval bounds are shifted to the right; for dispersion, all intervals are stretched out; and for skewness the lengths of the intervals increases from left to right.

On the left side of Figure 4.11, we compare the density f of the standard normal distribution with the density g of the t_2-distribution. The intervals, which are equidistant for f, are rescaled for g: they are longer for large values on the x-axis and shorter if they are close to the centre of the distributions. As discussed in Section 4.2.3, both $\mathcal{N}(0,1) \leq_3 t_2$ and $\mathcal{N}(0,1) \leq_{gs} t_2$ hold. Since all t-distributions are symmetric, $t_0 = 0$ is the canonical choice for the threshold of the concave-convex order. Note that $F \leq_2 G$, which is equivalent to $R''_{FG} \geq 0$, implies that the length of the intervals for g increases from left to right. $F \leq_3 G$, which is equivalent to $R'''_{FG} \geq 0$, implies that the rate, at which the length of the intervals increases, is increasing from left to right. If the length of the j-th interval in the lower left panel of Figure 4.11 is divided by the length of the $(j-1)$-th interval, that ratio increases from left to right. Furthermore, $F \leq_{gs} G$ means that the lengths of the intervals decreases up to zero and increases from there on out.

The densities in the central panels of Figure 4.11 are from the counterexample from the proof of Proposition 4.16 with g being rescaled to obtain the same support for both distributions. Here, $F \leq_{gs} G$ obviously holds again because the intervals for g decrease in length up to zero and increase from there on out. However, it is also obvious that the ratio between the lengths of the fourth and the third interval is smaller that the ratio between the lengths of the second and the first interval. These ratios can therefore not be increasing and $F \leq_3 G$ does not hold. This illustrates in which way $\leq_{gs}$ and $\leq_{gs}^{t_0}$ are weaker than $\leq_3$.

The panels on the right side of Figure 4.11 show the densities of two Weibull distributions for which $F \leq_3 G$ and $F \leq_{gs}^{t_0} G$ holds for all reasonable thresholds t_0 (see Section 4.2.3). Since the two Weibull distributions are skewness-comparable in the sense of $G \leq_2 F$, the lengths of the intervals for g are decreasing. Thus, $F \leq_{gs} G$ holds, but the statement is not meaningful because $0 \notin (-\infty, 0) = \operatorname{int}(R''_{FG}(D_F))$ is not a reasonable threshold. $F \leq_3 G$ also holds because the ratio between two neighbouring intervals, while being negative throughout, still increases. Contrary to $\leq_{gs}^{t_0}$, the fact that the order $\leq_3$ holds is always meaningful since it is not dependent on any parameter. Any threshold $t_0 \in \mathbb{R}$ for the order $\leq_{gs}^{t_0}$ can be converted into a threshold for the ratios between two neighbouring intervals. Thus, for all reasonable thresholds $t_0 \in \operatorname{int}(R''_{FG}(D_F))$, the ratio is smaller than the corresponding converted threshold up to some point and larger than the threshold from that point onward. This yields $F \leq_{gs}^{t_0} G$.

We can conclude that both $\leq_3$ and $\leq_{gs}$ make sense as orders of kurtosis for symmetric distributions. However, for asymmetric distributions, there is no canonical threshold for the order $\leq_{gs}^{t_0}$, making it difficult to use and interpret. Therefore, using $\leq_3$ for general distributions is advisable.

4.3. Kurtosis Measures on Asymmetric Distributions

Contrary to kurtosis orders, measures of kurtosis are often used on arbitrary distributions without ensuring that they are symmetric or nearly symmetric, particularly in practice (see, e.g., Wheeler, 1975 or Hanook et al., 2013). Measures are mostly chosen based on historical relevance and on how easy they are to use and interpret. Since this practice obviously lacks rigour, we start out by giving a general framework for the definition of kurtosis measures. Let $\mathcal{Q} \subseteq \mathcal{P}_I^3$ be the subset of all cdf's that are sufficiently regular for a given kurtosis measure candidate to be defined. Oja (1981) proposed for a mapping $\kappa : \mathcal{Q} \to \mathbb{R}$ to be said to be a measure, if satisfies the following two properties:

(K1) $\kappa(aX+b) = \kappa(X)$ for all $a \in \mathbb{R} \setminus \{0\}, b \in \mathbb{R}$ and $F \in \mathcal{Q}$, (K2) $\kappa(F) \leq \kappa(G)$ for all $F, G \in \mathcal{Q}$ such that $F \leq_K G$ for some kurtosis order $\leq_K$.

This definition is in line with that of measures of central location, dispersion and skewness from Definition 3.1. Usually, the crucial property is the one requiring the measure preserve a certain order with respect to the relevant characteristic, so in this case (K2). However, as discussed in the previous sections, no suitable proposal for a kurtosis order that is applicable to both symmetric and asymmetric distributions has been explicitly considered in the literature so far. Consequently, the usage of kurtosis measures on asymmetric distributions lacks any kind of foundation.

In Section 4.2, we concluded that the order of the third convex characteristic is, in general, superior to the concave-convex order for two major reasons. First, while $\leq_3$ is unambiguous, $\leq_{gs}^{t_0}$ depends on the used threshold t_0 for which there is no obvious preferred choice if the distribution is not symmetric. Second, $\leq_3$ is stronger than $\leq_{gs}^{t_0}$ for any threshold t_0, meaning that the former imposes a more basic requirement on corresponding kurtosis measures. $\leq_3$ is also the kurtosis analogue of the fundamental orders $\leq_{st}$, $\leq_{disp}$ and $\leq_c$ for the lower convex characteristics. However, by reusing Example 4.2, we observe that even this most basic requirement cannot be satisfied in a meaningful way.

Theorem 4.27. *a) There exists no mapping $\kappa : \mathcal{P}_I^3 \to \mathbb{R}$ such that $F \leq_3 G$ implies $\kappa(F) \leq \kappa(G)$ and $F <_3 G$ implies $\kappa(F) < \kappa(G)$ for all $F, G \in \mathcal{P}_I^\infty$.*

b) Let $t_0 \in \operatorname{int}(R''_{FG}(D_F))$. There exists no mapping $\kappa : \mathcal{P}_I^3 \to \mathbb{R}$ such that $F \leq_{gs}^{t_0} G$ implies $\kappa(F) \leq \kappa(G)$ and $F <_{gs}^{t_0} G$ implies $\kappa(F) < \kappa(G)$ for all $F, G \in \mathcal{P}_I^3$.

Proof. a) As mentioned before, we consider the three cdf's from Example 4.2:

$$\begin{aligned} F &: [0,1] \to [0,1], \quad t \mapsto t^3, \\ G &: [0,1] \to [0,1], \quad t \mapsto t, \\ H &: [0,1] \to [0,1], \quad t \mapsto 1 - \sqrt[3]{1-t}. \end{aligned}$$

Note that all three cdf's are three times differentiable and have interval support as well as strictly positive densities, i.e. $F, G, H \in \mathcal{P}_I^3$. As already shown in Example 4.2, $F \leq_3 G$ and $G \leq_3 H$ holds, but also $F \not\leq_3 H$. Note that this does not imply $H \leq_3 F$ as $\leq_3$ is not a total relation, which means that for two cdf's F and H it is possible that neither $F \leq_3 H$ nor $H \leq_3 F$.

Now we assume that there exists a mapping $\kappa : \mathcal{P}_I^3 \to \mathbb{R}$ that preserves the order $\leq_3$. It follows that $\kappa(F) \leq \kappa(G) \leq \kappa(H)$. We now contradict this by showing $H \leq_3 F$, which then implies $H <_3 F$ and $\kappa(H) < \kappa(F)$. To this end, it holds

$$R_{HF}(t) : [0,1] \to [0,1], \quad t \mapsto F^{-1}(H(t)) = \sqrt[3]{1 - \sqrt[3]{1-t}} = \left(1 - (1-t)^{1/3}\right)^{1/3}.$$

It follows for $t \in [0,1]$

$$\begin{aligned}
R'_{HF}(t) &= \tfrac{1}{9}\left(1-(1-t)^{1/3}\right)^{-2/3}(1-t)^{-2/3},\\
R''_{HF}(t) &= \tfrac{2}{27}\left(1-(1-t)^{1/3}\right)^{-2/3}(1-t)^{-5/3} - \tfrac{2}{81}\left(1-(1-t)^{1/3}\right)^{-5/3}(1-t)^{-4/3},\\
R'''_{HF}(t) &= \tfrac{10}{81}\left(1-(1-t)^{1/3}\right)^{-2/3}(1-t)^{-8/3} - \tfrac{4}{81}\left(1-(1-t)^{1/3}\right)^{-5/3}(1-t)^{-7/3}\\
&\qquad + \tfrac{10}{729}\left(1-(1-t)^{1/3}\right)^{-8/3}(1-t)^{-2},
\end{aligned}$$

and then

$$\begin{aligned}
H \le_3 F &\Leftrightarrow R'''_{HF}(t) \ge 0 \quad \forall t \in [0,1]\\
&\Leftrightarrow \tfrac{10}{81}\left(1-(1-t)^{1/3}\right)^2 - \tfrac{4}{81}\left(1-(1-t)^{1/3}\right)(1-t)^{1/3} + \tfrac{10}{729}(1-t)^{2/3} \ge 0\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall t \in [0,1]\\
&\Leftrightarrow (1-t)^{2/3} - \tfrac{27}{17}(1-t)^{1/3} + \tfrac{45}{68} \ge 0 \quad \forall t \in [0,1]\\
&\Leftrightarrow \left((1-t)^{1/3} - \tfrac{27}{34}\right)^2 \ge -\left(\tfrac{3}{17}\right)^2 \quad \forall t \in [0,1].
\end{aligned}$$

Since the last inequality is obviously true for all $t \in [0,1]$ because the left side is non-negative and the right side is negative, this concludes the proof.

b) Let F_i denote the cdf of a Weibull distributed random variable with shape parameter $i > 0$ and consider the triple (F_j, F_k, F_ℓ) with $2\ell < k$ and $\ell \in (\frac{j}{2}, j)$. It follows that the conditions $j \notin (k, 2k)$, $k \notin (\ell, 2\ell)$ and $\ell \in (\frac{j}{2}, j)$ are satisfied. According to the consideration of the Weibull distribution in Section 4.2.3, these three conditions are, in order, equivalent to $F_j \le_{gs}^{t_0} F_k$, $F_k \le_{gs}^{t_0} F_\ell$ and $F_\ell <_{gs}^{t_0} F_j$, thus concluding the proof by contradiction.

Note that the triple of cdf's used for the proof of part b) can also be used to prove part a) since the statements $F \le_3 G$ and $F \le_{gs}^{t_0} G$ are equivalent if F and G are Weibull distributed and the same equivalence is true for the corresponding strict orders.

Obviously, the statement of Theorem 4.27 is also valid if the set $\mathcal{P}_I^3$ is replaced by any other set of cdf's that includes the three cdf's used for the counterexample in the proof or any other triple of cdf's that poses an analogous contradiction.

Since there is an additional assumption made in Theorem 4.27, we cannot generally conclude that there is no kurtosis measure candidate that satisfies property (K2) on a sufficiently rich set of cdf's. This, however, is due to the fact that one can define a trivial kurtosis measure by just defining the constant mapping $\kappa_T \equiv k$ for some $k \in \mathbb{R}$. It is easy to verify that this defines a kurtosis measure in the sense that it satisfies both properties (K1) and (K2). The mapping $\kappa_T \equiv 0$ is also a measure of central location, dispersion and skewness, mainly due to the fact that the crucial properties of all these definitions are always stated using the non-strict version of the corresponding order. One way to exclude this special case would be to just assume that the kurtosis measure candidate is not constant on the given set of cdf's. This, however, would be very difficult to deal with since the existence of a kurtosis measure with these properties heavily depends on the considered set of cdf's and its structure. Showing that any kurtosis measure is constant on certain families like the Weibull family would be fairly easy. Whether different kurtosis values could be assigned to different distribution families would depend upon their interconnectedness in terms of the kurtosis order. Since this variant for defining kurtosis measures seems to be very convoluted, we will not analyze it here any further.

Another variant that excludes the trivial measure without exhibiting said transferability issue is the one used in Theorem 4.27: requiring the measure to preserve both the strict and the non-strict version of the underlying order. This variant comes with its own set of drawbacks, most notably that the strict order needs to be considered separately from the usually used non-strict one. For example, the fact that the non-strict order $\leq_3$ implies the non-strict order $\leq_{gs}^{t_0}$ does not mean that the same implication holds for the corresponding strict orders $<_3$ and $<_{gs}^{t_0}$. However, the latter implication can indeed be proved similarly to the former one (see Theorem 4.16 and Corollary 4.21). So, if property (K2) is extended to

(K2′) $\kappa(F) \leq \kappa(G)$ for all $F, G \in \mathcal{Q}$ such that $F \leq_K G$ and
$\kappa(F) < \kappa(G)$ for all $F, G \in \mathcal{Q}$ such that $F <_K G$ for some kurtosis order $\leq_K$ with strict version $<_K$

in the definition of kurtosis measures, Theorem 4.27 proves that there exist no kurtosis measures based on the kurtosis orders $\leq_3$ or $\leq_{gs}^{t_0}$. Furthermore, we can conclude that there exist no kurtosis measures based on any kurtosis order, which is weaker than $\leq_3$ in both the strict and non-strict version. Since $\leq_3$ is the strongest kurtosis order that can be found in the literature, this result deems the definition of kurtosis measures via (K1) and (K2′) to be too strong and therefore unusable.

Since the proof of Theorem 4.27 is based upon the fact that $\leq_3$ is not transitive, the question of how candidates for kurtosis measures behave on the transitivity sets of $\leq_3$ arises. However, that question needs to be addressed separately for each candidate and also for each kind of transitivity set. The three major kinds of transitivity sets found in Section 4.2 are $\mathcal{T}_{D,p}^{t}, t \in \mathbb{R}, p \in (0,1)$ (see Theorem 4.8), $\mathcal{T}_{Mode}^{\tilde{p}}, \tilde{p} \in (-1,1)$ (see Theorem 4.9), and the set of all sinh-arsinh distributions (see Corollary 4.26). Candidates for kurtosis measures are discussed in the following subsections. There, we focus on the fulfilment of (K2) for the order of the third convex characteristic and the concave-convex order. The modified requirement (K2′) was mainly defined to obtain Theorem 4.27. If the measure is constructed in a symmetric way, the property (K1) usually follows easily by using Lemma 3.5 or Proposition 2.22.

4.3.1. Moment-Based Approaches to Measuring Kurtosis

The Standardized Fourth Moment

The earliest attempt at measuring kurtosis is attributed to Pearson (1895) and is given by the standardized fourth moment

$$\kappa_M : \mathcal{L}^4 \to \mathbb{R}, \quad X \mapsto \mathbb{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^4\right]$$

(denoted by β_2 by Pearson). Ever since, the concept of kurtosis and what it describes has been much discussed. However, its oldest measure is still its most prominent one and often understood as synonymous with the notion of kurtosis itself (see, e.g., McAlevey and Stent, 2018 or Crack, 2019). The utilization of κ_M as a kurtosis measure is in line with that of the first three (standardized) moments as measures of central location, dispersion and skewness, respectively. The fourth moment is, however, the first one that cannot be standardized with respect to an earlier moment, namely the third one. The fact that κ_M is not invariant to skewness (in terms of the standardized third moment γ_M) is represented in the generally valid inequality

$$\kappa_M(X) \geq (\gamma_M(X))^2 + 1, \quad F \in \mathcal{L}^4 \tag{4.18}$$

(see Pearson, 1916, p. 432). The inequality essentially states that any distribution that is notably skewed in either direction is necessarily higher in kurtosis than less skewed distributions. For example, consider a normally distributed random variable Z. Since it satisfies $\kappa_M(Z) = 3$, it is less kurtotic with respect to κ_M than any random variable that is sufficiently skewed to satisfy $|\gamma_M(X)| > \sqrt{2}$ (like, e.g., $X \sim \mathrm{Exp}(\lambda), \lambda > 0$, with $\gamma_M(X) = 2$). Hence, similarly to the kurtosis orders, distributions are generally not comparable with respect to κ_M if they exhibit a notable difference in skewness. In the particular case of symmetric distributions, the measure κ_M was shown to preserve the order $\leq_s$ by van Zwet (1964, pp. 20-21). The fact that $\leq_{gs}$ is generally weaker than $\leq_3$ and equivalent to $\leq_s$ for symmetric distributions gives the following result.

Theorem 4.28. *If the mapping κ_M is restricted to the domain $\mathcal{S}$, it satisfies property (K2) for the kurtosis orders $\leq_3$ and $\leq_{gs}$.*

In fact, the result by van Zwet includes more than that: it states that every even standardized moment except for the second satisfies property (K2) if it is restricted to symmetric distributions. This seems to be related to the fact that the $2k$-th moment, $k \geq 2$, can only be standardized with respect to the first two moments and not with respect to the third up to the $(2k-1)$-th moment. Analogously, it is easy to show that the generalization of the standard deviation to $\sqrt[2k]{\mathbb{E}[(X-\mu_X)^{2k}]}$ is a measure of dispersion for all $k \geq 1$, since the dispersion is not standardized out of the measure. (This can be seen by replicating the proof that the standard deviation is a measure of dispersion from Oja (1981, p. 159). The crucial property here is that $x \mapsto x^{2k}$ is convex for all $k \geq 1$.) Because there is no known way of standardizing with respect to kurtosis, the difference in terms of kurtosis keeps on being represented in higher-order even standardized moments. If one wanted to meaningfully compare two random variables X and Y with respect to a higher convex characteristic like, e.g., the fifth, X and Y would have to be sufficiently similar with respect to all lower convex characteristics.

This vague concept can be made more concrete by switching from the order of the k-th convex characteristic $\leq_k$, $k \in \mathbb{N}_0$, to the k-convex order $\leq_{k-cx}$, $k \in \mathbb{N}$. Recall Proposition 2.20, which states that, for $k \in \mathbb{N}_0$, $F \leq_k G$ implies $F \leq_{(k+1)-cx} G$, if $\mathbb{E}[X^j] = \mathbb{E}[Y^j]$ holds for all $j \in \{1, \dots, k\}$. As noted before, $\leq_{k-cx}$ orders distributions with respect to their $(k-1)$-th convex characteristic. So, for $k = 2$ it is a dispersion order, for $k = 3$ a skewness order and for $k = 4$ a kurtosis order. However, for the order $\leq_{k-cx}$ to hold at all, all lower convex characteristics are required to be standardized (see Proposition 2.19). The fact that the same condition is required for $\leq_{k-1}$ to be stronger than $\leq_{k-cx}$ suggests that the standardization with respect to lower convex characteristics is not intrinsic to the orders $\leq_k$ of the k-th convex characteristic, contrary to the k-convex orders $\leq_{k-cx}$.

If the condition $\mathbb{E}[X^j] = \mathbb{E}[Y^j]$ for all $j \in \{1, \dots, k-1\}$ in the characterization of the k-convex order $\leq_{k-cx}$ (condition (i) in Proposition 2.19), in which the standardizing takes place, is omitted, we obtain the order $\leq_{k-icx}$. Its most popular representative is the increasing convex order $\leq_{icx}$, which is obtained for $k = 2$. This order is very popular in mathematical finance and risk analysis and quantifies location as well as dispersion of a distribution (see, e.g., Whitt, 1980, p. 1063). Thus, the additional requirement $\mathbb{E}[X] = \mathbb{E}[Y]$ is indeed necessary to obtain the pure dispersion order $\leq_{cx}$.

A Kurtosis Measure Using L-Moments

Along with the measures for location, dispersion and skewness discussed in Section 3.1.1, Hosking (1989, 1990) also introduced a kurtosis measure based on L-moments, a class of alternative moments that are also formally introduced in Section 3.1.1. Analogously to the skewness measure, the kurtosis measure is defined as the fourth L-moments divided by the second L-moment, which is a dispersion measure:

$$\tau_{LM} : \mathcal{L}^1 \to \mathbb{R}, \quad F \mapsto \frac{\lambda_4^F}{\lambda_2^F}.$$

Hosking (1989, pp. 6–7, Theorem 4) claims that τ_{LM} preserves the order $\leq_3$ for symmetric distributions, but does not explicitly prove it. To the knowledge of the author, this is the only place in the literature except for Oja (1981), where the order $\leq_3$ is used. In the following, it is proved explicitly that τ_{LM} preserves both $\leq_3$ and $\leq_{gs}$ for symmetric distributions.

Theorem 4.29. *If the mapping κ_{LM} is restricted to the domain $\mathcal{S}$, it satisfies property (K2) for the kurtosis orders $\leq_3$ and $\leq_{gs}$.*

Proof. Let $F, G \in \mathcal{S}$. Hosking (1989, pp. 3–4) showed that the k-th L-moment can also be written as

$$\lambda_k^F = \frac{1}{k-1} \int_{D_F} F(t)\,(1-F(t))\, Q_{k-2}(F(t))\,\mathrm{d}t, \quad k \geq 2,$$

where $Q_k(p), p \in (0,1)$, is the Jacobi polynomial $P_k^{(1,1)}(2p-1)$. Consequently,

$$\lambda_2^F = \int_0^1 \frac{p\,(1-p)}{f(F^{-1}(p))}\,\mathrm{d}p, \quad \lambda_4^F = \frac{1}{3}\int_0^1 \frac{p\,(1-p)}{f(F^{-1}(p))} \cdot Q_2(p)\,\mathrm{d}p,$$

where $Q_2(p) = 3(5p^2 - 5p + 1), p \in (0,1)$, and

$$\begin{aligned}
\lambda_2^G &= \int_0^1 p(1-p) \cdot \left(G^{-1}\right)'(p)\,\mathrm{d}p = \int_0^1 p(1-p) \cdot R'_{FG}(F^{-1}(p)) \cdot \left(F^{-1}\right)'(p)\,\mathrm{d}p \\
&= \int_0^1 \frac{p(1-p)}{f(F^{-1}(p))} \cdot R'_{FG}(F^{-1}(p))\,\mathrm{d}p, \\
\lambda_4^G &= \frac{1}{3}\int_0^1 \frac{p(1-p)}{f(F^{-1}(p))} \cdot Q_2(p) \cdot R'_{FG}(F^{-1}(p))\,\mathrm{d}p.
\end{aligned}$$

Since the integrand is symmetric around $\frac{1}{2}$ in the above representations of all four L-moments, each of them is equal to twice the same integral from 0 to $\frac{1}{2}$. Because of

$$\kappa_{LM}(G) - \kappa_{LM}(F) = \frac{\lambda_4^G \lambda_2^F - \lambda_4^F \lambda_2^G}{\lambda_2^F \lambda_2^G},$$

$\kappa_{LM}(F) \leq \kappa_{LM}(G)$ is equivalent to

$$\begin{aligned}
&\int_0^{\frac{1}{2}} \varphi(p) \cdot Q_2(p) \cdot R'_{FG}(F^{-1}(p))\,\mathrm{d}p \cdot \int_0^{\frac{1}{2}} \varphi(p)\,\mathrm{d}p \\
&\quad \geq \int_0^{\frac{1}{2}} \varphi(p) \cdot Q_2(p)\,\mathrm{d}p \cdot \int_0^{\frac{1}{2}} \varphi(p) \cdot R'_{FG}(F^{-1}(p))\,\mathrm{d}p,
\end{aligned} \tag{4.19}$$

where $\varphi(p) = \frac{p(1-p)}{f(F^{-1}(p))}, p \in (0,1)$, is a positive function. Now, the fact that both Q_2 and $R'_{FG} \circ F^{-1}$ are decreasing on $(0, \frac{1}{2})$ lets us invoke Chebyshev's generalized inequality for integrals (see Theorem 10 in Mitrinović, 1970, p. 40), which states that (4.19) holds. ▨

4.3.2. Quantile-Based Approaches to Measuring Kurtosis

In order to establish a connection between quantiles and convexity of any order, it is useful to consider difference-based characterizations of convexity instead of considering derivatives. More specifically, Proposition 2.14b) states that the k-convexity of a function can be equivalently characterized using divided differences. In this section, we consider multiple quantile-based approaches for measuring kurtosis, which are driven by that observation in the case $k = 3$. All of the approaches do not need the general requirement $F \in \mathcal{P}_I^3$, it is only kept for the sake of consistency throughout Chapter 4.

According to Proposition 2.14c), $F \leq_3 G$ is equivalent to

$$[F^{-1}(p_0), \ldots, F^{-1}(p_3)|\Delta_{FG}] \geq 0 \quad \forall\, 0 < p_0 < \ldots < p_3 < 1.$$

Using the recursive formula for divided differences from Definition 2.13 and the results of Example 2.15, where this is done for the skewness order $\leq_2$, we obtain that $F \leq_3 G$ holds, if and only if

$$\frac{\frac{G^{-1}(p_3)-G^{-1}(p_2)}{F^{-1}(p_3)-F^{-1}(p_2)} - \frac{G^{-1}(p_2)-G^{-1}(p_1)}{F^{-1}(p_2)-F^{-1}(p_1)}}{F^{-1}(p_3) - F^{-1}(p_1)} - \frac{\frac{G^{-1}(p_2)-G^{-1}(p_1)}{F^{-1}(p_2)-F^{-1}(p_1)} - \frac{G^{-1}(p_1)-G^{-1}(p_0)}{F^{-1}(p_1)-F^{-1}(p_0)}}{F^{-1}(p_2) - F^{-1}(p_0)} \geq 0 \tag{4.20}$$

for all $0 < p_0 < p_1 < p_2 < p_3 < 1$. Just like the equivalent condition (4.1), this inequality seems like it can not be rewritten in a way that is symmetric in F and G without any further assumptions. This conjecture is confirmed in Corollary 4.3.

Now, assume that F is a symmetric cdf, and choose $0 < \alpha < \eta < 1/2$. Further, put $p_0 = \alpha, p_1 = \eta, p_2 = 1-\eta, p_3 = 1-\alpha$, and define $c = F^{-1}(\eta)-F^{-1}(\alpha) = F^{-1}(1-\alpha)-F^{-1}(1-\eta)$ and $d = F^{-1}(1-\eta) - F^{-1}(\eta)$. Then, (4.20) takes the specific form

$$\frac{1}{c}\Big(G^{-1}(1-\alpha) - G^{-1}(\alpha)\Big) - \left(\frac{2}{d} + \frac{1}{c}\right)\Big(G^{-1}(1-\eta) - G^{-1}(\eta)\Big) \geq 0.$$

This is equivalent to

$$\frac{F^{-1}(1-\alpha) - F^{-1}(\alpha)}{F^{-1}(1-\eta) - F^{-1}(\eta)} \leq \frac{G^{-1}(1-\alpha) - G^{-1}(\alpha)}{G^{-1}(1-\eta) - G^{-1}(\eta)}. \tag{4.21}$$

As a consequence, for $0 < \alpha < \eta < 1/2$, the mapping

$$\kappa_Q^{\alpha,\eta} : \mathcal{P}_I^3 \to \mathbb{R}, \quad F \mapsto \frac{F^{-1}(1-\alpha) - F^{-1}(\alpha)}{F^{-1}(1-\eta) - F^{-1}(\eta)}$$

preserves the order $\leq_3$ on the subset $\mathcal{S} \subseteq \mathcal{P}_I^3$ of symmetric distributions. The same is true for the alternative mapping

$$\kappa_{QA}^{\alpha,\eta} : \mathcal{P}_I^3 \to \mathbb{R}, \quad F \mapsto \frac{F^{-1}(1-\alpha) - 3F^{-1}(1-\eta) + 3F^{-1}(\eta) - F^{-1}(\alpha)}{F^{-1}(1-\eta) - F^{-1}(\eta)},$$

as (4.21) is equivalent to

$$\kappa_{QA}^{\alpha,\eta}(F) \leq \kappa_{QA}^{\alpha,\eta}(G),$$

which can be seen by subtracting 3 on either side of (4.21). Even more can be shown, as stated in the following result.

Theorem 4.30. *Let $0 < \alpha < \eta < 1/2$. If the mappings $\kappa_Q^{\alpha,\eta}$ and $\kappa_{QA}^{\alpha,\eta}$ are restricted to the domain $\mathcal{S}$, they both satisfy property (K2) for the kurtosis orders $\leq_3$ and $\leq_{gs}$.*

Proof. Let $F, G \in \mathcal{S}$ satisfy $F \leq_{gs} G$. It is now sufficient to show $\kappa_Q^{\alpha,\eta}(F) \leq \kappa_Q^{\alpha,\eta}(G)$. Since R_{FG} is antisymmetric, it is concave on $\operatorname{supp}(F) \cap (-\infty, 0]$ and convex on $\operatorname{supp}(F) \cap [0, \infty)$. Additionally, $F^{-1}(\frac{1}{2}) = 0$ holds. Analogous to the equivalence of $F \leq_2 G$ and (2.11) in Example 2.10, it follows that

$$\frac{G^{-1}(p_2) - G^{-1}(p_1)}{F^{-1}(p_2) - F^{-1}(p_1)} - \frac{G^{-1}(p_1) - G^{-1}(p_0)}{F^{-1}(p_1) - F^{-1}(p_0)} \begin{cases} \leq 0 & \text{, if } 0 < p_0 < p_1 < p_2 \leq \frac{1}{2}, \\ \geq 0 & \text{, if } \frac{1}{2} \leq p_0 < p_1 < p_2 < 1. \end{cases} \tag{4.22}$$

Because of the symmetry of F and G, one of the two inequalities is redundant and we limit ourselves to the upper inequality. By applying $p_0 = \alpha, p_1 = \eta, p_2 = \frac{1}{2}$ to the upper part of (4.22), we obtain

$$\begin{aligned} \frac{F^{-1}(1-\alpha) - F^{-1}(\alpha)}{F^{-1}(1-\eta) - F^{-1}(\eta)} &= \frac{F^{-1}(\frac{1}{2}) - F^{-1}(\alpha)}{F^{-1}(\frac{1}{2}) - F^{-1}(\eta)} = \frac{F^{-1}(\eta) - F^{-1}(\alpha)}{F^{-1}(\frac{1}{2}) - F^{-1}(\eta)} + 1 \\ \leq \frac{G^{-1}(\eta) - G^{-1}(\alpha)}{G^{-1}(\frac{1}{2}) - G^{-1}(\eta)} + 1 &= \frac{G^{-1}(\frac{1}{2}) - G^{-1}(\alpha)}{G^{-1}(\frac{1}{2}) - G^{-1}(\eta)} = \frac{G^{-1}(1-\alpha) - G^{-1}(\alpha)}{G^{-1}(1-\eta) - G^{-1}(\eta)} \end{aligned}$$

The outer identities follow from the symmetry of F and G, which yields $F^{-1}(1-\alpha) - F^{-1}(\frac{1}{2}) = F^{-1}(\frac{1}{2}) - F^{-1}(\alpha)$ (and the same if F is replaced by G or α is replaced by η). ▨

In fact, Theorem 4.30 is still true, if only the less kurtotic distribution F is assumed to be symmetric. This is already clear for $\leq_3$ from the derivation of (4.21), but the following explanation for $\leq_{gs}$ offers more insight into how that relates to the relationship between kurtosis and skewness.

Let $0 < \alpha < \eta < \frac{1}{2}$ and $F \in \mathcal{S}, G \in \mathcal{P}_I^3$ with $F \leq_{gs} G$. Then, (4.22) still holds. By choosing $p_0 = \alpha, p_1 = \eta, p_2 = \frac{1}{2}$ in the upper case of (4.22) and choosing $p_0 = \frac{1}{2}, p_1 = 1 - \eta, p_2 = 1 - \alpha$ in the lower case, we obtain

$$\begin{aligned} \frac{G^{-1}(\frac{1}{2}) - G^{-1}(\alpha)}{G^{-1}(\frac{1}{2}) - G^{-1}(\eta)} &\geq \frac{F^{-1}(\frac{1}{2}) - F^{-1}(\alpha)}{F^{-1}(\frac{1}{2}) - F^{-1}(\eta)} \\ &= \frac{F^{-1}(1-\alpha) - F^{-1}(\frac{1}{2})}{F^{-1}(1-\eta) - F^{-1}(\frac{1}{2})} \leq \frac{G^{-1}(1-\alpha) - G^{-1}(\frac{1}{2})}{G^{-1}(1-\eta) - G^{-1}(\frac{1}{2})}. \end{aligned} \tag{4.23}$$

Since $\kappa_Q^{\alpha,\eta}(F)$ and $\kappa_Q^{\alpha,\eta}(G)$ are weighted averages of the one-sided quantities in (4.23), it follows that

$$\kappa_Q^{\alpha,\eta}(F) \leq \min\left\{ \frac{G^{-1}(\frac{1}{2}) - G^{-1}(\alpha)}{G^{-1}(\frac{1}{2}) - G^{-1}(\eta)}, \frac{G^{-1}(1-\alpha) - G^{-1}(\frac{1}{2})}{G^{-1}(1-\eta) - G^{-1}(\frac{1}{2})} \right\} \tag{4.24}$$

$$\begin{aligned}
&\leq \frac{G^{-1}(\frac{1}{2}) - G^{-1}(\eta)}{G^{-1}(1-\eta) - G^{-1}(\eta)} \frac{G^{-1}(\frac{1}{2}) - G^{-1}(\alpha)}{G^{-1}(\frac{1}{2}) - G^{-1}(\eta)} \qquad (4.25)\\
&\quad + \frac{G^{-1}(1-\eta) - G^{-1}(\frac{1}{2})}{G^{-1}(1-\eta) - G^{-1}(\eta)} \frac{G^{-1}(1-\alpha) - G^{-1}(\frac{1}{2})}{G^{-1}(1-\eta) - G^{-1}(\frac{1}{2})}\\
&= \kappa_Q^{\alpha,\eta}(G).
\end{aligned}$$

Inequality (4.24) is basically the same inequality that was used to prove Theorem 4.30 and its tightness is influenced by both F and G. However, the tightness of inequality (4.25) only depends on G and it becomes an equation if G is symmetric, suggesting that an increase of asymmetry for G tends to decrease the tightness of the inequality $\kappa_Q^{\alpha,\eta}(F) \leq \kappa_Q^{\alpha,\eta}(G)$. This can be made more concrete if (4.25) is analyzed more closely.

If we additionally assume $F <_2 G$,

$$\begin{aligned}
\frac{G^{-1}(\eta) - G^{-1}(\alpha)}{F^{-1}(\eta) - F^{-1}(\alpha)} &< \frac{G^{-1}(\frac{1}{2}) - G^{-1}(\eta)}{F^{-1}(\frac{1}{2}) - F^{-1}(\eta)}\\
&< \frac{G^{-1}(1-\eta) - G^{-1}(\frac{1}{2})}{F^{-1}(1-\eta) - F^{-1}(\frac{1}{2})} < \frac{G^{-1}(1-\alpha) - G^{-1}(1-\eta)}{F^{-1}(1-\alpha) - F^{-1}(1-\eta)}
\end{aligned}$$

follows. Since F is symmetric, we obtain that

$$\frac{G^{-1}(\frac{1}{2}) - G^{-1}(\alpha)}{G^{-1}(\frac{1}{2}) - G^{-1}(\eta)} < \frac{G^{-1}(1-\alpha) - G^{-1}(\frac{1}{2})}{G^{-1}(1-\eta) - G^{-1}(\frac{1}{2})}$$

and

$$\frac{G^{-1}(\frac{1}{2}) - G^{-1}(\eta)}{G^{-1}(1-\eta) - G^{-1}(\eta)} < \frac{G^{-1}(1-\eta) - G^{-1}(\frac{1}{2})}{G^{-1}(1-\eta) - G^{-1}(\eta)}.$$

This means that the two quantities in the minimum in (4.24) are not equal and that the weighted average in (4.25) is dominated by the larger of the two quantities. Hence, the inequality (4.24) is strict and its tightness is low if G is markedly skewed. If we assume $G <_2 F$ instead of $F <_2 G$, the situation is analogous with the roles of the two quantities in the minimum in (4.24) swapped.

These observations for the quantile-based kurtosis measures are reminiscent of inequality (4.18) for the moment-based kurtosis measure. Both state that the kurtosis measure value of markedly skewed distributions tends to be higher than the kurtosis measure value of symmetric distributions.

The measures $\kappa_Q^{\alpha,\eta}$ and $\kappa_{QA}^{\alpha,\eta}$ have appeared quite often in the literature, see Ruppert (1987), Balanda and MacGillivray (1988), and Jones et al. (2011), where further references can be found. Specific choices of the parameters in the literature are $\eta = \frac{1}{4}$ and $\alpha = 0.05$ or $\alpha = 0.01$. Alternative parameter choices can be obtained through equidistant evaluation of the quantile function. For example, the quintile-based measure $\kappa_{QA}^{1/5,2/5}$ was introduced in Jones et al. (2011, p. 90), motivated by the analogy to Bowley's skewness measure $\gamma_Q^{1/4}$, which takes second differences instead of third ones. Furthermore, Moors (1988, p. 26) defined the octile-based measure

$$M = \frac{F^{-1}(\frac{7}{8}) - F^{-1}(\frac{5}{8}) + F^{-1}(\frac{3}{8}) - F^{-1}(\frac{1}{8})}{F^{-1}(\frac{6}{8}) - F^{-1}(\frac{2}{8})}.$$

Noting that

$$M = \kappa_Q^{1/8,2/8} - \kappa_Q^{3/8,2/8} = \kappa_Q^{1/8,2/8} - \left(\kappa_Q^{2/8,3/8}\right)^{-1},$$

it follows that M preserves the orders $\leq_3$ and $\leq_{gs}$ for symmetric distributions as well.

A Kurtosis Functional for Asymmetric Distributions

All quantile-based kurtosis measures discussed so far are tailored to symmetric distributions. In order to obtain an easy to use kurtosis functional that preserves the order $\leq_3$ for general distributions, we start again from (4.20). Given $\alpha \in (0, \frac{1}{2})$, we set again $p_0 = \alpha, p_3 = 1 - \alpha$. Now, we choose p_1 and p_2 in such a way that d equals c; this leads to

$$p_1^{F,\alpha} = F\left(\frac{2}{3}F^{-1}(\alpha) + \frac{1}{3}F^{-1}(1-\alpha)\right), \quad p_2^{F,\alpha} = F\left(\frac{1}{3}F^{-1}(\alpha) + \frac{2}{3}F^{-1}(1-\alpha)\right).$$

With this choice, (4.20) boils down to

$$G^{-1}(1-\alpha) - 3G^{-1}(p_2^{F,\alpha}) + 3G^{-1}(p_1^{F,\alpha}) - G^{-1}(\alpha) \geq 0.$$

After dividing by the outer interquantile range $\tau_Q^\alpha(G)$ to obtain a scale invariant functional, we end up with the kurtosis functional

$$\kappa_{QF}^\alpha(F,G) = \frac{G^{-1}(1-\alpha) - 3G^{-1}(p_2^{F,\alpha}) + 3G^{-1}(p_1^{F,\alpha}) - G^{-1}(\alpha)}{G^{-1}(1-\alpha) - G^{-1}(\alpha)}, \quad 0 < \alpha < \frac{1}{2}. \tag{4.26}$$

Note that the functional is not symmetric in its two arguments F and G. Summing up, we have the following result.

Proposition 4.31. *If $F, G \in \mathcal{P}_I^3$ satisfy $F \leq_3 G$, then $\kappa_{QF}^\alpha(F,G) \geq 0$.*

This kurtosis functional has the advantage over $\kappa_Q^{\alpha,\eta}$ and $\kappa_{QA}^{\alpha,\eta}$ that it can be applied to all distributions without any requirement concerning symmetry. However, this comes along with the disadvantage that it can only quantify kurtosis of a pair of distributions and not of individual distributions. While this is also true for kurtosis orders, the functional κ_{QF}^α returns a number for any pair of distributions and therefore offers a more specific statement than the kurtosis orders.

4.3.3. Density-Based Approaches to Measuring Kurtosis

A different approach to quantify skewness and kurtosis than the ones presented in this Section was made by Critchley and Jones (2008). However, the utilized quantities at times bear a certain resemblance.

Density-based measures of location, dispersion and skewness are discussed in Section 3.1.3 These measures are obtained from the conditions $\Delta_{FG}(t) \geq 0$, $\Delta'_{FG}(t) \geq 0$ and $R''_{FG}(t) \geq 0$ for all $t \in D_F$. Since $F \leq_3 G$ is equivalent to $R'''_{FG}(t) \geq 0$ for all $t \in D_F$, we start out by calculating that third derivative based on (4.5), yielding

$$\begin{aligned} R'''_{FG}(t) = \left(\frac{1}{g(R_{FG}(t))}\right)^5 \cdot \Big[& f''(t)(g(R_{FG}(t)))^4 - 3f(t)f'(t)(g(R_{FG}(t)))^2 g'(R_{FG}(t)) \\ & + 3(f(t))^3(g'(R_{FG}(t)))^2 - (f(t))^3 g(R_{FG}(t))g''(R_{FG}(t))\Big]. \end{aligned}$$

Based on this, we find

$$\begin{aligned} F \leq_3 G \Leftrightarrow\ & f''(t)(g(R_{FG}(t)))^4 - (f(t))^3 g(R_{FG}(t))g''(R_{FG}(t)) \\ & \geq 3\left[f(t)f'(t)(g(R_{FG}(t)))^2 g'(R_{FG}(t)) - (f(t))^3(g'(R_{FG}(t))^2)\right] \quad \forall t \in D_F \\ \Leftrightarrow\ & \frac{f''(t)}{(f(t))^3} - \frac{g''(R_{FG}(t))}{(g(R_{FG}(t)))^3} \geq 3\frac{f'(t)}{(f(t))^2}\frac{g'(R_{FG}(t))}{(g(R_{FG}(t)))^2} - 3\frac{(g'(R_{FG}(t)))^2}{(g(R_{FG}(t)))^4} \\ & \hspace{20em} \forall t \in D_F \\ \Leftrightarrow\ & \frac{f''(F^{-1}(p))}{(f(F^{-1}(p)))^3} - \frac{g''(G^{-1}(p))}{(g(G^{-1}(p)))^3} \\ & \geq 3\frac{g'(G^{-1}(p))}{(g(G^{-1}(p)))^2}\left(\frac{f'(F^{-1}(p))}{(f(F^{-1}(p)))^2} - \frac{g'(G^{-1}(p))}{(g(G^{-1}(p)))^2}\right) \qquad \forall p \in (0,1). \end{aligned} \tag{4.27}$$

By plugging in the definition of $\gamma_D^p, p \in (0,1)$, $F \leq_3 G$ is equivalent to

$$\frac{f''(F^{-1}(p))}{(f(F^{-1}(p)))^3} - \frac{g''(G^{-1}(p))}{(g(G^{-1}(p)))^3} \geq 3\gamma_D^p(G)(\gamma_D^p(F) - \gamma_D^p(G)) \quad \forall p \in (0,1).$$

If we now assume that both F and G are symmetric, $\gamma_D^{1/2}(F) - \gamma_D^{1/2}(G) = 0$ holds for the specific choice of $p = \frac{1}{2}$. In that case, $F \leq_3 G$ implies

$$-\frac{f''(F^{-1}(p))}{(f(F^{-1}(p)))^3} \leq -\frac{g''(G^{-1}(p))}{(g(G^{-1}(p)))^3}.$$

This suggests that a reasonable choice for a density-based kurtosis measure is obtained as the special case $p = \frac{1}{2}$ from the following class of mappings

$$\kappa_D^p : \mathcal{P}_I^3 \to \mathbb{R}, \quad F \mapsto -\frac{f''(F^{-1}(p))}{(f(F^{-1}(p)))^3},$$

where $p \in (0,1)$. As for the other density-based measures, we introduce the short hand $\kappa_D = \kappa_D^{1/2}$. Finally, this means that $F \leq_3 G$ is equivalent to

$$\kappa_D^p(G) - \kappa_D^p(F) \geq 3\gamma_D^p(G)(\gamma_D^p(F) - \gamma_D^p(G)) \quad \forall p \in (0,1). \tag{4.28}$$

Additionally, swapping the roles of F and G yields that $G \leq_3 F$ is equivalent to

$$\kappa_D^p(G) - \kappa_D^p(F) \leq 3\gamma_D^p(F)(\gamma_D^p(F) - \gamma_D^p(G)) \quad \forall p \in (0,1). \tag{4.29}$$

Now, if κ_D^p is heuristically understood to measure kurtosis (without specifying exactly what that means) and γ_D^p is heuristically understood to measure skewness, this statement can be interpreted as follows: if two distributions are equally skewed, the right sides of both (4.28) and (4.29) vanish, meaning that the kurtosis comparison between F and G is centred around zero. G exhibits more kurtosis than F, if $\kappa_D^p(G) \geq \kappa_D^p(F)$, and vice versa, with both cdf's being equivalent in terms of kurtosis, if the two quantities are equal. This behaviour can be summarized as follows.

Theorem 4.32. *Let $t \in \mathbb{R}$ and $p \in (0,1)$. If $F, G \in \mathcal{T}_{D,p}^t$ satisfy $F \leq_3 G$, then $\kappa_D^p(F) \leq \kappa_D^p(G)$.*

Similarly to the proof of (S1) for γ_D, it can easily be shown that κ_D^p satisfies (K1), if and only if $p = \frac{1}{2}$. Thus, for any $t \in \mathbb{R}$, $\kappa_D = \kappa_D^{1/2}$ restricted to the set $\mathcal{T}_D^t$ is a measure of kurtosis in the sense of (K1) and (K2). Note that $\mathcal{T}_D^0$ is a superset of other transitivity sets like $\mathcal{T}_{Mode}^0$ and $\mathcal{S}$ (see Remark 4.10a)).

However, the situation gets more complex if F and G differ in terms of skewness. Define $\ell_{FG}^p = 3\gamma_D^p(G)(\gamma_D^p(F) - \gamma_D^p(G))$ as the lower limit for $F \leq_3 G$ and $u_{FG}^p = 3\gamma_D^p(F)(\gamma_D^p(F) - \gamma_D^p(G))$ as the upper limit for $G \leq_3 F$, both for all $p \in (0,1)$. Because of

$$u_{FG}^p - \ell_{FG}^p = 3(\gamma_D^p(F) - \gamma_D^p(G))^2 \geq 0,$$

the centre of their comparison in terms of skewness extends to an interval of length $3(\gamma_D^p(F) - \gamma_D^p(G))^2$. So, the bigger the difference in skewness is between the two distributions, the larger the interval that is associated with equivalence with respect to the kurtosis order $\leq_3$. The aforementioned centre of the kurtosis comparison not only extends to an interval but also shifts, depending on the concrete values of $\gamma_D^p(F)$ and $\gamma_D^p(G)$ and, more specifically, on their signs. It is easy to see from their definitions that ℓ_{FG}^p and u_{FG}^p have the same sign, if and only if $\gamma_D^p(F)$ and $\gamma_D^p(G)$ have the same sign. Note that ℓ_{FG}^p and u_{FG}^p having the same sign means that $\kappa_D^p(F) = \kappa_D^p(G)$ for all $p \in (0,1)$ implies $F <_3 G$ if said sign is negative and $G <_3 F$ if said sign is positive. Furthermore, 0 is contained in the interval $[\ell_{FG}^p, u_{FG}^p]$, around which the kurtosis comparison is centred, if and only if $\gamma_D^p(F)$ and $\gamma_D^p(G)$ have differing signs (including the case that either is zero).

The interplay of the difference of kurtosis measures and the upper and lower limits with respect to the order $\leq_3$ is considered for specific asymmetric distributions in the following example.

Example 4.33. a) Let $X \sim \text{Weib}(1)$ with cdf F and let $Y \sim \text{Weib}(k)$ with cdf G for $k > 0$, where the Weibull distribution is defined as in Section 4.2.3. The difference $\kappa_D^p(G) - \kappa_D^p(F)$ for $p = \frac{1}{2}$ is plotted in the left panel of Figure 4.12 as a function of the distributional parameter k. The same panel additionally shows in red the lower and upper limits $\ell_{FG}^{1/2}$ and $u_{FG}^{1/2}$ of the 'centre of kurtosis comparison' between F and G from (4.28) and (4.29). All three graphs are obviously zero at $k = 1$ since $F = G$ holds in that case. For $1 < k < 2$, we have $\kappa_D(G) - \kappa_D(F) > u_{FG}^{1/2}$. This observation is in line with $F <_3 G$ holding for exactly those values of k as mentioned in (4.10) as $F \leq_3 G$ is equivalent to $\kappa_D^p(G) - \kappa_D^p(F) \geq \ell_{FG}^p$ for all $p \in (0,1)$. Note the equivalence

$$F <_3 G \Leftrightarrow \kappa_D^p(G) - \kappa_D^p(F) \geq \ell_{FG}^p \quad \forall p \in (0,1) \qquad \text{and} \tag{4.30}$$
$$\exists p_0 \in (0,1): \; \kappa_D^{p_0}(G) - \kappa_D^{p_0}(F) > u_{FG}^{p_0}.$$

In this case, the latter inequality holds for $p_0 = \frac{1}{2}$. The observation $\kappa_D(G) - \kappa_D(F) < \ell_{FG}^{1/2}$ for $\frac{1}{2} < k < 1$ is in accordance with (4.11) in a similar way. This is not immediately obvious from Figure 4.12, but can be validated by rescaling the plot window. The fact that $\kappa_D(G) - \kappa_D(F) \in [\ell_{FG}^{1/2}, u_{FG}^{1/2}]$ holds for $k < \frac{1}{2}$ and $k > 2$ is an implication of (4.12). Since $F =_3 G$ holds in that case, $\kappa_D^p(G) - \kappa_D^p(F) \in [\ell_{FG}^p, u_{FG}^p]$ follows for all $p \in (0,1)$.

Note that the family of Weibull distributions decreases in skewness with respect to $\leq_2$ as k increases. Hence, the sign change of both $\ell_{FG}^{1/2}$ and $u_{FG}^{1/2}$ at $k = 1$ stems from the fact that $\gamma_D(G) - \gamma_D(F)$ changes its sign from positive to negative. The additional sign change of $\ell_{FG}^{1/2}$ at $k \approx 3.26$ stems from $\gamma_D(G)$ changing sign. The Weibull distribution changes from right-skewed to left-skewed around that value of k with the exact value being determined by the utilized skewness measure.

b) Let $X \sim \Gamma(1)$ with cdf F and $Y \sim \Gamma(k)$ with cdf G for $k > 0$, where the gamma distribution is defined as in Section 4.2.3. The same quantities as for the Weibull distribution are plotted in the right panel of Figure 4.12. It suggests that $\kappa_D(G) - \kappa_D(F) > u_{FG}^{1/2}$ for $k > 1$ and $\kappa_D(G) - \kappa_D(F) < \ell_{FG}^{1/2}$ for $k < 1$. This would be perfectly in line with the observations in Section 4.2.3, namely $F <_3 G$ for $k > 1$ and $G <_3 F$ for $k < 1$. Exploration of the graphs outside the plot window yields $\kappa_D(G) - \kappa_D(F) \in [\ell_{FG}^{1/2}, u_{FG}^{1/2}]$ for $k > 25.5$ and for $k < 0.24$. This, however, does not

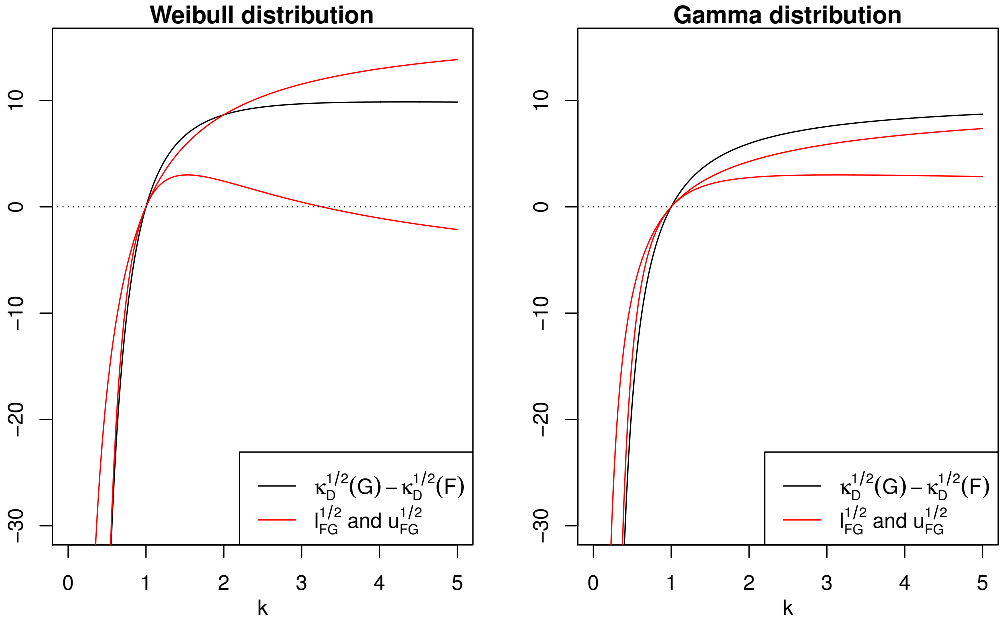

Figure 4.12.: Graphs of the differences of density-based kurtosis measures. Left panel: F and G cdf's of the Weib(1)- and the Weib(k)-distributions respectively. Right panel: F and G cdf's of the $\Gamma(1)$- and the $\Gamma(k)$-distributions respectively.

contradict the observations on the gamma distribution from Section 4.2.3 because of the equivalent characterisation of $F <_3 G$ in (4.30).

The interpretation of (the signs of) $\ell_{FG}^{1/2}$ and $u_{FG}^{1/2}$ is fairly straightforward: Both $\gamma_D(F)$ and $\gamma_D(G)$ take positive values throughout, but their difference changes sign at $k = 1$ as the Gamma distributions decrease in skewness with respect to $\leq_2$ as k increases.

A analogous result to Theorem 4.32 cannot be obtained for the family of orders $\leq_{gs}^{t_0}, t_0 \in \mathbb{R}$. However, the slightly altered version $<_{gss}$ of $\leq_{gs}$, which is defined in Remark 4.23 is indeed preserved by κ_D^p on a corresponding density-based transitivity set $\mathcal{T}_{D,p}^t, t \in \mathbb{R}$. Since $<_{gss}$ is a strict order, it also implies that the kurtosis measures are ordered in a strict sense.

Theorem 4.34. *Let $t \in \mathbb{R}$ and $p \in (0,1)$. If $F, G \in \mathcal{T}_{D,p}^t$ satisfy $F <_{gss} G$, then $\kappa_D^p(F) < \kappa_D^p(G)$.*

Proof. Recalling the equivalences in (4.6), $F, G \in \mathcal{T}_{D,p}^t$ implies $R''_{FG}(F^{-1}(p)) = 0$. Furthermore, Remark 4.23b) states that $F <_{gss} G$ implies the existence of a $p_{FG} \in (0,1)$ such that $R''_{FG}(t) < 0$ for $t < F^{-1}(p_{FG})$ and $R''_{FG}(t) > 0$ for $t > F^{-1}(p_{FG})$. Obviously, $p_{FG} = p$ holds. This implies

$$R'''_{FG}(F^{-1}(p)) = \lim_{\varepsilon \searrow 0} \frac{R''_{FG}(F^{-1}(p) + \varepsilon) - R''_{FG}(F^{-1}(p) - \varepsilon)}{2\varepsilon} > 0,$$

because of $R''_{FG}(F^{-1}(p) + \varepsilon) > 0$ and $R''_{FG}(F^{-1}(p) - \varepsilon) < 0$ for all $\varepsilon > 0$ small enough to satisfy $F^{-1}(p) \pm \varepsilon \in D_F$. By replicating the equivalences (4.27) for strict inequalities, we obtain that $R'''_{FG}(F^{-1}(p)) > 0$ is equivalent to

$$\kappa_D^p(G) - \kappa_D^p(F) > 3\gamma_D^p(G)(\gamma_D^p(F) - \gamma_D^p(G)).$$

Because of $F, G \in \mathcal{T}_{D,p}^t$, $\gamma_D^p(F) = t = \gamma_D^p(G)$ holds, which concludes the proof. ▨

Note that this result is not valid if $<_{gss}$ is replaced by $\leq_{gs}$. The latter order is not strong enough to ensure the identity $p_{FG} = p$, which is a crucial point in the proof of Theorem 4.34.

A major drawback present in all four density-based measures (for location, dispersion, skewness and kurtosis) is the fact that they are only based on one single evaluation of the quantile function, the density function and its derivatives. A possible solution using integration is given in Remark 3.16 for location, dispersion and skewness measures. The same idea is applied to kurtosis in the following.

Remark 4.35. Let μ be a symmetric, finite measure on the set $(0, 1)$. Because of the symmetry of μ, the mapping $\kappa_{ID}^\mu(F) = \int_0^1 \kappa_D^p(F)\,\mu(\mathrm{d}p), F \in \mathcal{P}_I^3$, can easily be shown to satisfy the kurtosis property (K1). The crucial property (K2) (with respect to $\leq_3$) cannot be inherited as easily as for the integrated density-based measures in Remark 3.16 because the set on which (K2) is fulfilled differs based on p. Thus, for any function $t : (0, 1) \to \mathbb{R}, p \mapsto t(p)$, κ_{ID}^μ fulfils (K2), if it is restricted to the set $\bigcap_{p\in(0,1)} \mathcal{T}_{D,p}^{t(p)}$. However, it is unclear whether there exists such a function t for which the given set is non-empty.

For the exemplary measure $\mu = \mathcal{U}(\alpha, 1-\alpha), \alpha \in [0, \frac{1}{2})$, used in Remark 3.16, we obtain

$$\kappa_{ID}^\mu(F) = \int_\alpha^{1-\alpha} -\frac{f''(F^{-1}(p))}{(f(F^{-1}(p)))^3}\,\mathrm{d}p = -\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \frac{f''(t)}{(f(t))^2}\,\mathrm{d}t.$$

As opposed to the corresponding measures of location, dispersion and skewness, this cannot be expressed explicitly and is therefore no viable candidate for a kurtosis measure.

Overall, (4.28) and (4.29) suggest that the comparison of distributions in terms of kurtosis has to account for possible differences in skewness. The same observation is made in Sections 4.3.1 and 4.3.2. In the first case, the evidence comes in the form of the well-known inequality (4.18). In the second case, it is shown that the tightness of the inequality $\kappa_Q^{\alpha,\eta}(F) \leq \kappa_Q^{\alpha,\eta}(G)$ is reduced if F or G are markedly skew or if they exhibit a large difference in skewness.

A similar statement can be made about the density-based measurement of kurtosis in this section. However, compared to the ideas from Section 4.3.2, the inequalities (4.28) and (4.29) represent the influence of skewness on the measurement of kurtosis in a more precise way. Since these inequalities also characterize the kurtosis order $\leq_3$, this helps to heuristically understand why it is difficult to meaningfully apply the order to distributions with a significant difference in skewness.

An alternative to kurtosis measures in the traditional form is given by the kurtosis functional κ_{QF}^{α} defined in (4.26). It requires two cdf's F and G as arguments and returns a scalar that represents the difference in kurtosis between F and G. In Proposition 4.31, it is shown to preserve the order $\leq_3$ in the sense that $F \leq_3 G$ implies $\kappa_{QF}^{\alpha}(F,G) \geq 0$. Analogously, a density-based kurtosis functional can be defined by

$$\kappa_{DF}^{p}(F,G) = \kappa_{D}^{p}(G) - \kappa_{D}^{p}(F) - 3\gamma_{D}^{p}(G)\left(\gamma_{D}^{p}(F) - \gamma_{D}^{p}(G)\right)$$

for all $p \in (0,1)$. An advantage κ_{DF}^{p} has over κ_{QF}^{α} is that it characterizes the order $\leq_3$ in the sense that $F \leq_3 G$ is equivalent to $\kappa_{DF}^{p}(F,G) \geq 0$ for all $p \in (0,1)$. However, neither κ_{DF}^{p} nor κ_{QF}^{α} is symmetric in their two arguments F and G.

Another connection between density-based measures and quantile-based measures of lower convex characteristics is discussed in Remark 3.15. More specifically, it is shown that density-based measures can be interpreted as limiting values of quantile-based measures. In the following remark, this is replicated for measures of kurtosis.

Remark 4.36. The quantile-based kurtosis measure that seems to be most closely related to τ_{Q}^{α} and γ_{Q}^{α} is $\kappa_{QA}^{\alpha,\eta}$, since it has a higher-order difference in its numerator. Because $\kappa_{QA}^{\alpha,\eta}$ is dependent upon two parameters $0 < \alpha < \eta < \frac{1}{2}$, further assumptions have to be made to calculate a meaningful limit. In order to obtain a natural continuation of the previous limits, α and η are chosen in such a way that the evaluation points of the quantile function are equidistant. Hence, we choose $\eta = \frac{1}{2} - \beta$ and $\alpha = \frac{1}{2} - 3\beta$ for an $\beta \in (0, \frac{1}{6})$, meaning that the distance between to neighbouring evaluation points is fixed to 2β. After slightly altering the rescaling factor used in Remark 3.15, the following limit is obtained for $\beta \searrow 0$

$$\begin{aligned}
\lim_{\beta \searrow 0} \frac{2}{\beta^2} \cdot \kappa_{QA}^{\frac{1}{2}-3\beta,\frac{1}{2}-\beta}(F) &= 8 \cdot \lim_{\beta \searrow 0} \frac{\frac{F^{-1}(\frac{1}{2}+3\beta)-3F^{-1}(\frac{1}{2}+\beta)+3F^{-1}(\frac{1}{2}-\beta)-F^{-1}(\frac{1}{2}-3\beta)}{(2\beta)^3}}{\frac{F^{-1}(\frac{1}{2}+\beta)-F^{-1}(\frac{1}{2}-\beta)}{2\beta}} \\
&= \frac{\left(F^{-1}\right)'''\left(\frac{1}{2}\right)}{\left(F^{-1}\right)'\left(\frac{1}{2}\right)} = -\frac{f''(F^{-1}(\frac{1}{2}))}{(f(F^{-1}(\frac{1}{2})))^3} + 3\left(\frac{f'(F^{-1}(\frac{1}{2}))}{(f(F^{-1}(\frac{1}{2})))^2}\right)^2 \\
&= \kappa_{D}^{1/2}(F) + 3(\gamma_{D}^{1/2}(F))^2. \qquad (4.31)
\end{aligned}$$

In contrast to the two parts of Remark 3.15, we do not obtain the already known density-based measure as limiting value. However, if the limit above is defined as an alternative density-based kurtosis measure, most of the results concerning κ_D can be replicated.

Particularly, if we define

$$\kappa_{DA}^{p} : \mathcal{P}_{I}^{3} \to \mathbb{R}, \quad F \mapsto \kappa_{D}^{p}(F) + 3(\gamma_{D}^{p}(F))^2,$$

we obtain that $F \leq_3 G$ is equivalent to

$$\kappa_{DA}^p(G) - \kappa_{DA}^p(F) \geq 3\gamma_D^p(F)(\gamma_D^p(G) - \gamma_D^p(F)) \quad \forall p \in (0,1). \tag{4.32}$$

This lower bound coincides with ℓ_{GF}^p, so the lower bound for the corresponding difference of measures κ_D^p with the roles of F and G reversed. Therefore, κ_{DA}^p also preserves the order $\leq_3$, if F and G are equally skewed in the sense of $\gamma_D^p(F) = \gamma_D^p(G)$, or if one of the cdf's (in this case F) is symmetric and $p = \frac{1}{2}$.

The fact that κ_{DA}^p is more closely related to the quantile-based kurtosis measures than κ_D^p can also be seen in another way. In Section 4.3.2, it is shown that the symmetry of the less kurtotic cdf F is sufficient for $\kappa_Q^{\alpha,\eta}$ to preserve the order $\leq_3$. The same statement for κ_{DA}^p follows directly from (4.32). However, the symmetry of F is not sufficient for κ_D^p to preserve $\leq_3$, as evidenced by (4.28). Instead, a sufficient condition is given by the symmetry of the more kurtotic cdf G.

4.3.4. Expectile-Based Approaches to Measuring Kurtosis

For any cdf $F \in \mathcal{L}^1$, we define an expectile-based kurtosis measure by

$$\kappa_E^{\alpha,\eta}(F) = \frac{e_F(1-\alpha) - 3e_F(1-\eta) + 3e_F(\eta) - e_F(\alpha)}{e_F(1-\eta) - e_F(\eta)}, \quad 0 < \alpha < \eta < \frac{1}{2}.$$

We use the analogue to $\kappa_{QA}^{\alpha,\eta}$ instead of the equivalent analogue to $\kappa_Q^{\alpha,\eta}$ because the former is better suited for determining its limiting value, as noted in Remark 4.36. In contrast to the expectile-based measures considered in Section 3.2, the expectile kurtosis $\kappa_E^{\alpha,\eta}$ has the major problem that it is not known whether preserves any kurtosis order on some subset of distributions like $\mathcal{S}$. Thus, its limiting value is needed to draw a connection to kurtosis orders. Recall that, for the expectile skewness γ_E^α, its limiting value for $\alpha \nearrow \frac{1}{2}$, given by $\gamma_{EL}(F) = 2F(\mu_F) - 1$, is much simpler and easier to handle. Specifically, the proof that γ_{EL} is a skewness measure in Theorem 3.29 is very simple. In order to proceed analogously for the expectile kurtosis, we consider the same limit as in (4.31) and replace the quantile function with the corresponding expectile function, yielding

$$\lim_{\beta \searrow 0} \frac{2}{\beta^2} \cdot \kappa_E^{\frac{1}{2}-3\beta,\frac{1}{2}-\beta}(F) = 8 \cdot \lim_{\beta \searrow 0} \frac{\frac{e_F(\frac{1}{2}+3\beta) - 3e_F(\frac{1}{2}+\beta) + 3e_F(\frac{1}{2}-\beta) - e_F(\frac{1}{2}-3\beta)}{(2\beta)^3}}{\frac{e_F(\frac{1}{2}+\beta) - e_F(\frac{1}{2}-\beta)}{2\beta}} = \frac{(e_F)'''\left(\frac{1}{2}\right)}{(e_F)'\left(\frac{1}{2}\right)}. \tag{4.33}$$

The first two derivatives of the expectile function e_F, evaluated at $\frac{1}{2}$, are given by

$$e_F'(\tfrac{1}{2}) = 2\tau_{EL}(F),$$
$$e_F''(\tfrac{1}{2}) = 8\tau_{EL}(F) \cdot (2F(\mu_F) - 1),$$

where $\tau_{EL}(F) = \mathbb{E}[|X - \mu_F|]$ denotes the mean absolute deviation from the mean (see (3.16) and (3.23)). To shorten the following calculations, we use the abbreviation $\delta_F = \tau_{EL}(F)$. The third derivative at $\frac{1}{2}$ can be calculated similarly to the second, which utilizes the general form of the first derivative from Proposition 2.22f), given by

$$e'_F(\alpha) = \frac{\mathbb{E}[|X - e_F(\alpha)|]}{(1-\alpha)F(e_F(\alpha)) + \alpha\,(1 - F(e_F(\alpha)))}, \quad \alpha \in (0,1).$$

If we denote the numerator and the denominator by $u(\alpha)$ and $v(\alpha)$, respectively, we obtain

$$e'''_F(\tfrac{1}{2}) = \lim_{\alpha \nearrow \frac{1}{2}} \frac{u''(\alpha)(v(\alpha))^2 - 2u'(\alpha)v'(\alpha)v(\alpha) + 2u(\alpha)(v'(\alpha))^2 - u(\alpha)v''(\alpha)v(\alpha)}{(v(\alpha))^3}.$$

This can be calculated using

$$\begin{aligned}
u(\alpha) &= \mathbb{E}[|X - e_F(\alpha)|] \overset{\alpha \to \frac{1}{2}}{\to} \delta_F, \\
u'(\alpha) &= e'_F(\alpha)(2F(e_F(\alpha)) - 1) \overset{\alpha \to \frac{1}{2}}{\to} e'_F(\tfrac{1}{2})(2F(\mu_F) - 1) = 2\delta_F \gamma_{EL}(F), \\
u''(\alpha) &= e''_F(\alpha)(2F(e_F(\alpha)) - 1) + 2(e'_F(\alpha))^2 f(e_F(\alpha)) \overset{\alpha \to \frac{1}{2}}{\to} 8\delta_F \left((\gamma_T(F))^2 + \delta_F f(\mu_F)\right)
\end{aligned}$$

and

$$\begin{aligned}
v(\alpha) &= (1-\alpha)F(e_F(\alpha)) + \alpha(1 - F(e_F(\alpha))) \overset{\alpha \to \frac{1}{2}}{\to} \tfrac{1}{2}, \\
v'(\alpha) &= (1 - 2F(e_F(\alpha))) + (1-2\alpha)f(e_F(\alpha))e'_F(\alpha) \overset{\alpha \to \frac{1}{2}}{\to} -\gamma_{EL}(F), \\
v''(\alpha) &= (1-2\alpha)f'(e_F(\alpha))(e'_F(\alpha))^2 + (1-2\alpha)f(e_F(\alpha))e''_F(\alpha) - 4f(e_F(\alpha))e'_F(\alpha) \\
&\overset{\alpha \to \frac{1}{2}}{\to} -8f(\mu_F)\delta_F
\end{aligned}$$

(see also (3.21) and (3.22)). Hence,

$$\begin{aligned}
e'''_F(\tfrac{1}{2}) &= 8\left[2\delta_F\left((\gamma_{EL}(F))^2 + \delta_F f(\mu_F)\right) + 2\delta_F(\gamma_{EL}(F))^2 + 2\delta_F(\gamma_{EL}(F))^2 + 4f(\mu_F)\delta_F^2\right] \\
&= 48\delta_F\left(\delta_F f(\mu_F) + (\gamma_{EL}(F))^2\right).
\end{aligned}$$

By plugging this into (4.33), we obtain the limiting expectile-based kurtosis measure

$$\begin{aligned}
\lim_{\beta \searrow 0} \frac{2}{\beta^2} \cdot \kappa_E^{\frac{1}{2}-3\beta, \frac{1}{2}-\beta}(F) &= \frac{(e_F)'''\left(\frac{1}{2}\right)}{(e_F)'\left(\frac{1}{2}\right)} = \frac{48\delta_F\left(\delta_F f(\mu_F) + (\gamma_{EL}(F))^2\right)}{2\delta_F} \\
&= 24\left(\delta_F f(\mu_F) + (\gamma_{EL}(F))^2\right).
\end{aligned}$$

We disregard the factor of 24 for the sake of simplicity and define the mapping

$$\kappa_{EL} : \mathcal{L}^1 \to \mathbb{R}, \quad F \mapsto \tau_{EL}(F) \cdot f(\mu_F) + (\gamma_{EL}(F))^2.$$

If we restrict that mapping to symmetric distributions, the second summand vanishes as then, $\gamma_T(F) = 0$. In that case, κ_{EL} satisfies the requirements for a kurtosis measure.

Theorem 4.37. *If the mapping κ_{EL} is restricted to the domain $\mathcal{S}$, it is a kurtosis measure in the sense that it satisfies (K1) and (K2) for the kurtosis orders $\leq_3$ and $\leq_{gs}$.*

Proof. Let $F \in \mathcal{S}$. We start out by showing that κ_{EL} satisfies (K1) in two steps, first $\kappa_{EL}(aX+b) = \kappa_{EL}(X)$ for all $a > 0$ and $b \in \mathbb{R}$. To this end, consider $\tau_{EL}(aX+b) = a\tau_{EL}(X)$ and

$$f_{aX+b}(\mathbb{E}[aX+b]) = \frac{1}{a} f\left(\frac{\mathbb{E}[aX+b] - b}{a}\right) = \tfrac{1}{a} f(\mu_F).$$

For (K1), it remains to be shown that $\kappa_{EL}(-X) = \kappa_{EL}(X)$, which is implied by $\tau_{EL}(-X) = \tau_{EL}$ and $f_{-X}(\mathbb{E}[-X]) = f(-\mathbb{E}[-X]) = f(\mu_F)$.

Now, assume $F, G \in \mathcal{S}$ such that $F \leq_{gs} G$. Furthermore, assume $\mu_F = \mu_G = 0$ and $f(\mu_F) = g(\mu_G)$, which poses no restriction because κ_{EL} is invariant under shifts and rescaling since it satisfies (K1). It follows from $F \leq_{gs} G$ as well as $F, G \in \mathcal{S}$ that R_{FG} is concave on $(-\infty, 0] \cap D_F$ and convex on $[0, \infty) \cap D_F$. Thus, $R'_{FG}(t) = \frac{f(t)}{g(R_{FG}(t))}, t \in D_F$, is decreasing for $t \leq 0$ and increasing for $t \geq 0$. So, the function R'_{FG} reaches its global minimum $\frac{f(\mu_F)}{g(\mu_G)} = \frac{f(0)}{g(0)} = 1$ at 0. (Since both F and G are symmetric, their medians and means coincide, yielding $R_{FG}(0) = 0$.) It follows that $R'_{FG}(t) \geq 1$ for all $t \in D_F$, which is equivalent to $F \leq_{disp} G$. Since τ_{EL} is a dispersion measure, $\tau_{EL}(F) \leq \tau_{EL}(G)$ holds, thus concluding the proof. ▨

In fact, the result above holds for any mapping $\kappa : \mathcal{S} \to \mathbb{R}, F \mapsto \tau(F) f(\nu(F))$, where ν is an arbitrary measure of central location and τ is an arbitrary measure of dispersion. The second generalization is valid because the fact that τ_{EL} is a dispersion measure is the only property of τ_{EL} that is utilized in the proof of Theorem 4.37. The mean can be replaced by any other location measure since all location measures are equal on a symmetric distribution according to Proposition 3.4a).

We make use of this flexibility by centring the kurtosis measure around the median instead of around the mean. The resulting mapping

$$\kappa_{EM} : \mathcal{L}^1 \to \mathbb{R}, \quad F \mapsto \tau_{IQ}(F) \cdot f(F^{-1}(\tfrac{1}{2})),$$

where $\tau_{IQ}(F) = \mathbb{E}[|X - F^{-1}(\frac{1}{2})|]$ denotes the mean absolute deviation around the median (see (3.4)), preserves the order $\leq_3$ under weaker assumptions.

Theorem 4.38. *Let $t \in \mathbb{R}$. If $F, G \in \mathcal{T}_D^t$ satisfy $F \leq_3 G$, then $\kappa_{EM}(F) \leq \kappa_{EM}(G)$.*

Proof. Since κ_{EM} is invariant to affine linear transformations (for analogous reasons as κ_{EL} on $\mathcal{S}$), we can assume without restriction that $F^{-1}(\frac{1}{2}) = G^{-1}(\frac{1}{2}) = 0$ and $f(F^{-1}(\frac{1}{2})) = g(G^{-1}(\frac{1}{2}))$. According to Proposition 4.7, $\frac{1}{2} \in \Pi_{FG}$ holds for all pairs $F, G \in \mathcal{T}_D^t$ with $F \leq_3 G$. It follows that R_{FG} is concave on $(-\infty, F^{-1}(\frac{1}{2})] \cap \operatorname{supp}(F)$ and convex on $[F^{-1}(\frac{1}{2}), \infty) \cap \operatorname{supp}(F)$. Noting $F^{-1}(\frac{1}{2}) = 0$ and that the mapping τ_{IQ} is also a dispersion measure, the remainder of the proof is analogous to that of Theorem 4.37. ▨

4.3.5. Ratios of Dispersion Measures as Measures of Kurtosis

One of the first attempts at describing kurtosis measures in a general way was made by Bickel and Lehmann (1975a, pp. 469–470). They defined a kurtosis measure as a 'suitable' ratio of two (possibly rescaled) dispersion measures. Of the measures we have discussed in Section 4.3 thus far, several fit this description.

The most obvious instance is the quantile-based measure κ_Q, which is defined as a wider interquantile range divided by a more narrow interquantile range. Any interquantile range is by definition a measure of dispersion (see Theorem 3.8b)). The situation is similar for the expectile kurtosis κ_E, which is equivalent to a ratio of two interexpectile ranges τ_E. While any interexpectile range is a dispersion measure according to Theorem 3.24, the ratio could not be shown to preserve the order $\leq_3$ on any notable subset of distributions.

Another example is the expectile limit measure around the median, which can be rewritten as $\kappa_{EM}(F) = \frac{\tau_{IQ}(F)}{\tau_D(F)}$. Note that $\tau_D(F) = \frac{1}{f(F^{-1}(\frac{1}{2}))}$ is also a dispersion measure according to Theorem 3.12. Finally, the moment-based measure κ_M is equivalent to the monotonically transformed

$$\sqrt[4]{\kappa_M(F)} = \frac{\sqrt[4]{\mathbb{E}[(X-\mu_F)^4]}}{\sigma_F}, \quad F \in \mathcal{L}^4.$$

(Two measures κ_1 and κ_2 are said to be equivalent, if $\kappa_1(F) \leq \kappa_1(G)$ is equivalent to $\kappa_2(F) \leq \kappa_2(G)$ for all cdf's F and G, so if one measure can be monotonically transformed into the other.) Since the generalized standard deviation $\sqrt[2k]{\mathbb{E}[(X-\mu_F)^{2k}]}$ is shown to be a dispersion measure for all $k \in \mathbb{N}$ in Section 4.3.1, $\sqrt[4]{\kappa_M}$ is also a ratio of two dispersion measures. All other kurtosis measure candidates in this work have no apparent representation as a ratio of two dispersion measures. Hence, while the rather vague definition from Bickel and Lehmann (1975a) obviously does have some merit, it does not seem to coincide with the order-based approach at defining kurtosis measures.

A pattern can be observed in the kurtosis measures that are the ratio of two dispersion measures. In all cases, the dispersion measure in the numerator puts more emphasis on the tails of the distribution relative to the measure in the denominator, which focuses more on the centre of the distribution. This begs the question as to why these kinds of constructions tend to preserve the order $\leq_3$, at least on the subset of symmetric distributions. While the notion of 'putting more emphasis on the tails' seems to be too vague to obtain any general rigorous result explaining this behaviour, the proof of Theorem 4.37 is fairly instructive for this kind of situation. If $F, G \in \mathcal{S}, F \neq G$ satisfy $F \leq_{gs} G$, then the function R'_{FG} is decreasing up to 0 and increasing from there on out. We now assume the more centre-focussed dispersion measures in the denominator τ_C to be equal, i.e. $\tau_C(F) = \tau_C(G)$. If τ_C not only satisfies (D2) but also its strict version, meaning that $F <_{disp} G$ implies $\tau_C(F) < \tau_C(G)$, then neither $R'_{FG} > 1$ nor $R'_{FG} < 1$ is true. It follows that there exists a $t_0 \in D_F \cap [0, \infty)$ such that $R'_{FG}(t) > 1$ holds for $|t| \geq t_0$. If the more tail-focussed dispersion measure in the numerator τ_T is sufficiently similar to τ_C, then $\tau_T(F) \leq \tau_T(G)$ follows from the fact that the dispersions of F and G are similar around the centre (since $\tau_C(F) = \tau_C(G)$) and that F is more dispersed than G on the tails (since $R'_{FG}(t) > 1$ for large absolute values of t). Overall, the mapping $F \mapsto \frac{\tau_T(F)}{\tau_C(F)}$ then preserves the concave-convex order $\leq_{gs}$ and therefore also the order $\leq_3$ of the third convex characteristic.

This similarity in the construction of a number of kurtosis measures strengthens the interpretation of the concept of kurtosis by Balanda and MacGillivray (1988) that is noted at the beginning of Section 4.1. The density of a typical symmetric and unimodal distribution that is very kurtotic has a sharp peak in the centre, declines steeply away from it, and has fat tails. Thus, the dispersion of the distribution mostly lies far away from its centre. If a distribution exhibits little kurtosis, the shoulders of its density are very prominent compared to its centre and tails. Here, the dispersion of the distribution is mostly close to the centre. Both graphical intuitions are illustrated in the lower right panel of Figure 1.1. Overall, a centre-focussed dispersion measure τ_C tends to take larger values for distributions with less kurtosis, and a tail-focussed dispersion measure τ_T tends to take larger values for more kurtotic distributions.

CHAPTER 5

Conclusion and Outlook

As evidenced in Chapter 3, the quantification of location, dispersion and skewness of sufficiently regular continuous probability distributions is well-founded through the literature. Specifically Sections 3.1.1 and 3.1.2, where the most popular measures are discussed, mostly consist of already known results without any major gaps. There is widespread agreement on what the fundamental underlying orders of these characteristics are and their usage is usually unproblematic.

The largest contribution of novel results concerning these three most basic characteristics within Chapter 3 comes in the form of expectile-based methods in Section 3.2. The usage of expectiles to describe distributions and their characteristics may seem unappealing at first because they are less intuitive than more well-established quantities like quantiles or moments. However, along with their rising popularity in actuarial and financial mathematics, Theorem 3.23 provides a convincing argument for expectiles to be considered an important tool for the description of probability distributions. That result states that the dilation order, which is one of the most utilized orders of dispersion, can be equivalently characterized using expectiles. If the expectiles in that characterization are replaced by quantiles, the so-called weak dispersive order $\leq_{w-disp}$ is obtained. An open question for future research is to analyze the exact relationship between that order $\leq_{w-disp}$ and the dilation order $\leq_{dil}$, which coincides with the weak expectile dispersive order $\leq_{we-disp}$. The answer to that question should also shed some light on whether centring a distribution around some central location measure like the mean or the median can be reversed and switched to another choice of centre. Since most measures of convex characteristics higher than location utilize some sort of centring of the distribution, this is a quite interesting question. Of course, it only affects asymmetric distributions since otherwise, all location measures coincide (see Proposition 3.4a)). A hint is given by Bellini et al. (2018b, p. 1852), who suggest that the interquantile range, as a dispersion measure centred around the median, does not preserve the dilation order, which is centred around the mean.

The majority of novel results for continuous distributions in this thesis is given in Chapter 4 concerning the characteristic of kurtosis. It is already known from the literature that the theory underlying the quantification of kurtosis fundamentally differs from all lower convex characteristics. More specifically, the lack of transitivity of underlying orders was mentioned by Oja (1981) and the intrinsic entanglement of kurtosis and skewness for asymmetric distributions was pointed out and tackled in different ways by several publications (see, e.g., MacGillivray and Balanda, 1988, Balanda and MacGillivray, 1990, Blest, 2003 or Jones et al., 2011). However, no prior publication has analyzed $\leq_3$, which seems like the canonical choice for a fundamental kurtosis order, with respect to these ideas.

At first glance, $\leq_3$ exhibits the same problems that have already been documented for the quantification of kurtosis in general. However, closer inspection reveals the transitivity sets that are developed over the course of Section 4.2.1. The restriction of attention to subsets of distributions that are comparable in terms of skewness acts as a substitute for standardization, which is not possible via an arithmetic operation for skewness.

A class of kurtosis orders, which has received much more attention in the literature, is given by the concave-convex orders. The corresponding order $\leq_s$ is generalized to asymmetric distributions in Section 4.2.2 in a more flexible and adaptive way than through the antiskewness-order $\leq_a$ by MacGillivray and Balanda (1988). While statements by Oja (1981) give the impression that this kind of order is superior to $\leq_3$, our considerations do not prove this to be true. The generalized concave-convex order is also not transitive in general and while it does also have transitivity sets, they are more difficult to find and less plentiful than for $\leq_3$. Overall, $\leq_3$ is concluded to be the superior order of kurtosis, also considering that it is preserved by more candidates for kurtosis measures, sometimes under weaker assumptions.

A notable exception concerning the entanglement of kurtosis and skewness is posed by the class of sinh-arsinh-distributions introduced by Jones and Pewsey (2009). These distributions depend on four parameters associated with location, dispersion, skewness and kurtosis. The already known result that certain quantile-based kurtosis measures are invariant to the skewness parameter of this class of distributions (see Jones et al., 2011) is generalized to the underlying orders in Theorem 4.24. It states the following: whether two distributions can be ordered with respect to either $\leq_3$ or a generalized concave-convex order $\leq_{gs}^{t_0}$ is solely dependent on their kurtosis parameters and, in particular, independent of their skewness parameters. Since any legitimate measure of kurtosis should preserve at least one of these orders, this property is inherited by any such measure. This result makes the sinh-arsinh-distributions much more appealing than other four-parameter distribution families that include the normal distributions. Thus, it is an interesting topic for future research.

For distributions outside of that specific class, skewness and kurtosis remain entangled and the subsequent problems persist. However, Section 4.3.3 on density-based kurtosis measures provides explicit formulas on the structure of that entanglement, which arise from an equivalent characterisation of the order $\leq_3$. They suggest that a small difference of a kurtosis measure for two distributions is meaningless, if one distribution is sufficiently left-skewed and the other is sufficiently right-skewed. The larger the difference in skewness is, the larger the difference in the kurtosis measure needs to be in order to have any merit. If the kurtosis measure takes two values that are too close to each other, then that result is meaningless. Heuristically, one might say that a difference in skewness blurs the view on their comparison with respect to kurtosis. Furthermore, if the two distributions are both differently skewed in the same direction, then the same value of a kurtosis measure for both distributions no longer means that they are equally kurtotic. When talking about the difference of the two kurtosis measure values, the comparison is no longer centred around zero. If the distributions are both right-skewed, the centre is a positive value of that difference; and if they are both left-skewed, the centre is a negative value of that difference.

It is important to note that all of these rather vague and heuristic statements rest upon results for density-based measures. Similarly explicit results for other classes of kurtosis measures could not be obtained. Almost all considered candidates for kurtosis measures satisfy the crucial property (K2) for symmetric distributions, which can be seen as a kind of minimal requirement. Stronger results hold for $\kappa_{EM}(F) = \mathbb{E}[|X - F^{-1}(\frac{1}{2})|] \cdot f(F^{-1}(\frac{1}{2}))$, which satisfies (K2) for cdf's F in the more general transitivity sets $\mathcal{T}_D^t, t \in \mathbb{R}$; and for the quantile-based measures $\kappa_Q^{\alpha,\eta}$ and $\kappa_{QA}^{\alpha,\eta}$, which already satisfy (K2) if only the less kurtotic cdf F is symmetric with no further requirement on the other cdf G. The proof of (K2) for the quantile-based measures also contains some heuristic information on how asymmetries make a comparison with respect to kurtosis difficult. It is similar, but less explicit than the information obtained from density-based measures.

However, most other kurtosis measures are much better suited to be used in applications than the density-based measures. Because the densities and especially their derivatives are virtually impossible to estimate with sensible accuracy, the density-based measures are only useful in a theoretical context.

Convex characteristics of order four and higher have seldom been considered in the literature. In his concluding remarks, Oja (1981, p. 168) stated that the interpretation of the orders $\leq_4, \leq_5, \ldots$ are an open question. Oja argues that an increase with respect to $\leq_5$ could be associated with 'more tendency to bimodality'. Other publications have presented arguments for interpreting kurtosis as a lack of tendency to bimodality (see Darlington, 1970, Chissom, 1970, Hildebrand, 1971 or Balanda and MacGillivray, 1988, pp. 113–114). This is somewhat plausible when recalling the interpretation of Balanda and MacGillivray (1988, p. 116) that an increase in kurtosis comes with less probability mass on the 'shoulders' of the distribution and more on the centre and the tails. Thus, a distribution with minimal kurtosis has its entire probability mass on its shoulders and practically none in the centre or the tails, which intuitively gives the image of a bimodal distribution with small tails.

There are a number of patterns than can be observed in the zeroth to third convex characteristic. Continuing these in a straightforward way may shed some light on the nature of higher-order characteristics. By using the derivative-based characterization of $\leq_k$, one can generalize the idea that suggests $\leq_s$ to be a suitable kurtosis order. For that, let F and G be two sufficiently regular cdf's that are not overly asymmetric. Then, $F \leq_k G$ is equivalent to $\Delta_{FG}^{(k)} \geq 0$, which often implies the existence of a $t_0 \in D_F$ such that $\Delta_{FG}^{(k-1)}(t) \leq 0$ for $t \leq t_0$ and $\Delta_{FG}^{(k-1)}(t) \geq 0$ for $t \geq t_0$. Thus, an increase in the k-th convex characteristic can prototypically be characterized by a decrease in the $(k-1)$-the convex characteristic on the left side and an increase in the $(k-1)$-th convex characteristic on the right side. For $k = 1$, a decrease in location on the left side and an increase in location on the right side yields an increase in dispersion; analogous statements can be made for $k = 2$ and $k = 3$. This means that an increase in the fourth convex characteristic is associated with asymmetry that is located in the shoulder on the left side and asymmetry that is located in the tail on the right side. However, this description becomes fuzzy for larger values of k.

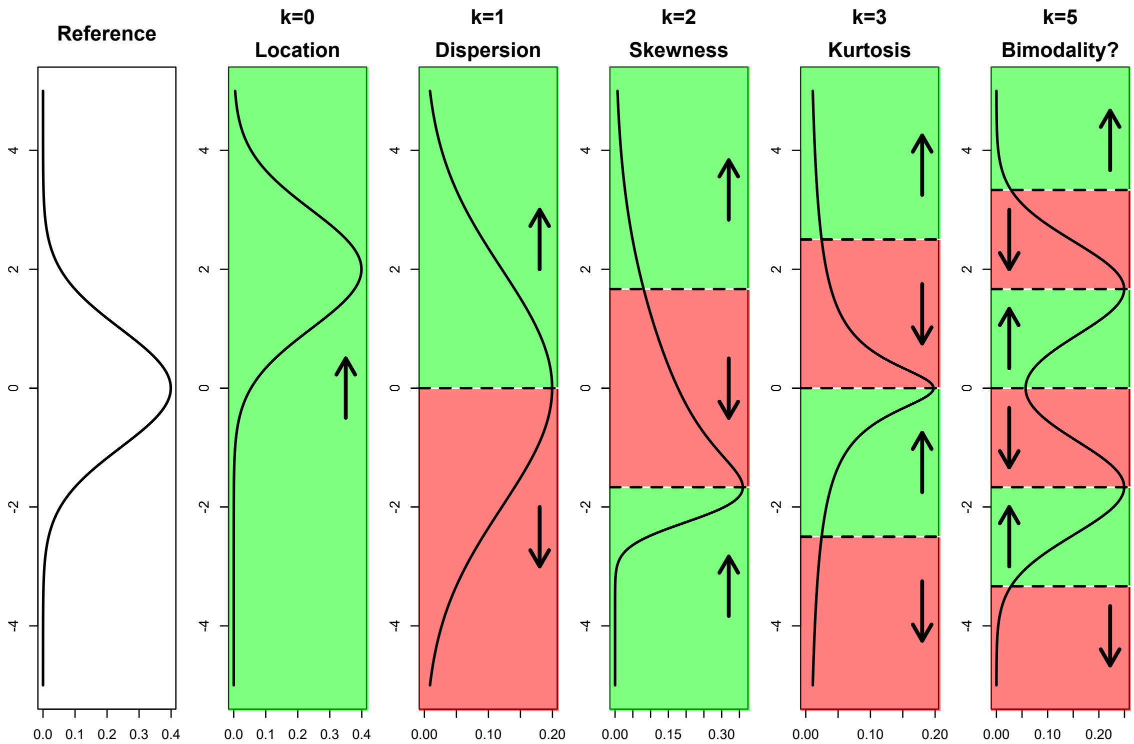

Figure 5.1.: Illustration of how an increase in the k-th convex characteristic can be interpreted for different values of k.

Another possible approach is based on the quantile-based measures. For $k \in \{0, 1, 2, 3\}$, the numerator of the measure of the k-th convex characteristic is given by a k-th order difference of quantiles. Heuristically, one can imagine the distribution to be divided into $k + 1$ parts and each part is represented by one evaluation of the quantile function. The sign of the rightmost evaluation is positive and it alternates for the subsequent evaluations. For $k = 0$, an increase in location means that the single evaluation of the quantile function becomes larger. This is represented in the two leftmost panels of Figure 5.1. The first panel depicts the density of the reference distribution F and the second panel depicts the density of the transformed distribution G. For $k = 1$, an increase in dispersion means that the smaller evaluation decreases and the larger evaluation increases, overall stretching out the distribution. In the third panel of Figure 5.1, the decrease is represented by the colour red and a downward arrow and the increase is represented by the colour green and an upward arrow. This heuristic can be continued for skewness in the fourth panel and kurtosis in the fifth panel; the reference distribution is always the one in the leftmost panel. Applying this heuristic to the case $k = 5$ suggests that Oja (1981, p. 168) might indeed be correct in suggesting that this convex characteristic is associated with more tendency to bimodality (see rightmost panel of Figure 5.1). Of course, this principle can also be applied to any other $k \in \mathbb{N}_0$. A pattern that can be observed is that the k-th convex characteristic is asymmetric in nature for even k and symmetric in nature for odd k.

The heuristic from Figure 5.1 is also supported by the intersection criterion for the order of the k-th convex characteristic given in Proposition 2.17b). If $F \leq_k G$ holds for a $k \in \mathbb{N}_0$ and F and G are equal in all lower-order convex characteristics in some sense, then F and G intersect exactly k times and the last non-zero value of $F - G$ is positive. Thus, both distributions are divided into $k + 1$ parts and, in the last part, $G^{-1} - F^{-1}$ is positive. So, in the last part, a G-distributed random variable takes larger values than a F-distributed random variable. Throughout the rest of the parts, the direction of the transformation from F to G alternates. This, again, leaves us with the heuristic illustrated by Figure 5.1.

However, the consideration of higher convex characteristics has a major problem. Since distributions can only be standardized with respect to location and dispersion, all characteristics of order $k \geq 2$ remain entangled, just like skewness and kurtosis are. An indicator for this is given by results by van Zwet (1964) concerning the $(k + 1)$-th standardized moment ($k \geq 4$), which is an intuitive candidate for a measure of the k-th convex characteristic. If $k + 1$ is odd, that moment is always a measure of skewness because it preserves the order $\leq_c$; and if $k + 1$ is even, that moment is always a measure of kurtosis for symmetric distributions because it preserves the order $\leq_s$. This strengthens the conjecture that all higher-order convex characteristics remain entangled and are difficult to separate. A solution we found for the treatment of kurtosis is to consider it on sets of comparable skewness as a substitute for standardization. This means that convex characteristics of an order $k \geq 4$ can only meaningfully discussed on sets of distributions on which all characteristics of order $2, 3, \dots, k - 1$ are equal in some sense. It seems highly doubtful that such sets have any merit. Another possible solution comes in the form of special distribution families like the sinh-arsinh-distributions in which the characteristics can be separated.

Overall, it is unclear whether the consideration of higher convex characteristics than kurtosis is sensible. On one hand, they may be helpful in characterizing and categorizing distributional shape, e.g. by adding further dimensions the skewness-kurtosis-plane. It would also be very useful to be able to associate intuitive properties like bimodality with such a characteristic. On the other hand, its meaningful description and discussion seems more unlikely, the higher the order of the characteristic.

PART II

Discrete Setting

Chapter 6

Basic Orders on Discrete Distributions

In Part I of this thesis, numerous mappings are shown to be measures of central location, dispersion, skewness and kurtosis. The importance of rigorously assessing their suitability to measure the corresponding characteristic cannot be understated because these kinds of measures are often used in applications without being questioned. While most candidates for measures considered in Part I indeed fulfil the corresponding defining properties, there are a few exceptions, particularly for the characteristic of skewness (see (3.6) and (3.8)). Another mapping that does not satisfy the crucial skewness property (S2) was nonetheless introduced as a measure of skewness or asymmetry by Patil et al. (2012) on the basis of heuristic observations. However, Eberl and Klar (2021) pointed out multiple examples in which the proposed measure behaves counterintuitively. It was shown that the quantity can be better described as a measure of similarity to the exponential distribution. The order-based definition of measures of characteristics is the best known method to exclude such misconstructions, while only imposing a basic requirement and leaving the exact interpretation of the characteristic up to the measures.

All cdf's considered throughout Part I are assumed to be absolutely continuous and to have interval support. These assumptions were often made in the literature on this topic, e.g. by Oja (1981), who formalized and unified the order-based approach to the quantification of the first four convex characteristics. However, both of these assumptions are not met by discrete distributions, which is a far too rich class of distributions to not be included in this fundamental theory. Measures of these characteristics are usually applied to both continuous and discrete distributions. Besides application to popular lattice distributions like the binomial, Poisson or geometric distribution, this also includes empirical measures, which are usually obtained by applying theoretical measures to empirical, and therefore discrete, distributions. It is the purpose of Part II of this thesis to meaningfully extend the order-based approach for the quantification of convex characteristics to discrete distributions.

Orders of location and dispersion have been considered without any assumptions on the involved distributions in the literature, e.g. by Müller and Stoyan (2002) and Belzunce et al. (2015). To the knowledge of the author, the concepts of skewness and kurtosis have been exclusively discussed in a continuous setting, with rare or brief exceptions like Eberl and Klar (2019) or a counterexample given by van Zwet (1964, pp. 16–17). Another exception is, of course, posed by the usage of measures of these characteristics on discrete distributions. Throughout Part II, we mostly focus on location and dispersion because groundwork has been laid in the literature for these characteristics and they are also easier to handle, as evidenced in Part I. Ideas concerning a discrete skewness order are briefly discussed in Section 8.2.

We restrict our attention to the orders of the convex characteristics in question, so to $\leq_{st}$, $\leq_{disp}$ and $\leq_c$. For location and skewness, this choice is fairly straightforward because the corresponding orders are very basic, do not explicitly favour any kind of measure, and they are by far the most popular orders of these characteristics. For dispersion, the choice is not quite as easy because the dilation order $\leq_{dil}$ is also very popular in the literature, maybe even more so than $\leq_{disp}$ in the financial and actuarial literature. However, it is established in Section 3.2.2 that $\leq_{dil}$ centres the comparison with respect to dispersion around the mean and therefore favours dispersion measures that are centred around the mean. It is also known that $\leq_{dil}$ is a weakening of $\leq_{disp}$ (see Proposition 2.20b) and Example 2.21b) or Müller and Stoyan, 2002, p. 42), and a more basic order is more suitable for our purposes because it allows for an interpretation of dispersion that is as general as possible.

6.1. The Usual Stochastic Order

The usual stochastic order $\leq_{st}$ is generally very compatible with discrete distributions. This is evidenced by the fact that most of its properties and characterizations are valid without any assumptions of smoothness (see Müller and Stoyan, 2002, pp. 2–7 or Belzunce et al., 2015, pp. 28–36). Exceptions are, of course, posed by results relating $\leq_{st}$ to other stochastic orders that are not as compatible with discrete distributions. Furthermore, the stochastic order is known to be applicable to many important classes of discrete distributions. Müller and Stoyan (2002, pp. 61, 63) noted that the stochastic order holds within a number of popular lattice distribution for suitable parameter choices, including the binomial, Poisson and geometric distributions. Further results of this type were derived by Klar et al. (2010) and Klenke and Mattner (2010).

The only notable irregularities that the stochastic order exhibits for discrete distributions concern its characterization via RIDF's and similar concepts. RIDF's are among the most important objects for the comparison of distributions throughout this thesis and they are connected to the orders of convex characteristics via Proposition 2.9 and Corollary 2.12. In the case of the stochastic order, they state that, for $F, G \in \mathcal{P}_I$, $F \leq_{st} G$ is equivalent to $\Delta_{FG}(t) \geq 0$ for all $t \in D_F$. If this requirement is applied to distributions that are not absolutely continuous, it makes sense to include the right endpoint of the support of F (if it exists), because it may have positive probability mass on it. This extends the requirement to $\Delta_{FG}(t) \geq 0$ for all $t \in D'_F$. However, as stated by the following result, an additional assumption is necessary to obtain equivalence to the stochastic order for arbitrary distributions.

Proposition 6.1. *Let $F, G \in \mathcal{P}$.*

a) $\Delta_{FG}(t) \geq 0$ for all $t \in D'_F$ implies $F \leq_{st} G$.

b) The reverse implication from a) holds, if and only if there does not exist a non-degenerate interval $I_0 \subseteq \mathbb{R}$ and a number $p_0 \in (0,1)$ such that $F(I_0) = G(I_0) = \{p_0\}$.

Proof. a) Assume $\Delta_{FG}(t) \geq 0$ for all $t \in D'_F$ and assume that there exists a $t_0 \in \mathbb{R}$ such that $G(t_0) > F(t_0)$ for a proof by contradiction. This eliminates the case $F(t_0) = 1$; and $F(t_0) = 0$ yields $t_0 \leq G^{-1}(F(t_0)) = \inf D_G$, which contradicts $G(t_0) > 0$. Thus, only the case $0 < F(t_0) < 1$ is left, which is equivalent to $t_0 \in D_F$. Furthermore, G has a jump discontinuity that skips $F(t_0)$, since otherwise it would follow $t_0 > \inf\{s \in \mathbb{R} : G(s) \geq F(t_0)\} = G^{-1}(F(t_0))$. Hence, $t_0 = G^{-1}(F(t_0))$ holds. Because F is right-continuous, there now exists a $\varepsilon > 0$ such that $F(t_0 + \varepsilon) < G(t_0) \leq G(t_0 + \varepsilon)$. This yields $G^{-1}(F(t_0 + \varepsilon)) = \inf\{s \in \mathbb{R} : G(s) \geq F(t_0 + \varepsilon)\} = t_0 < t_0 + \varepsilon$, which poses a contradiction to $\Delta_{FG}(t) \geq 0$ holding for all $t \in D'_F$ and thus completes the proof of part a).

b) First, we assume that $G(t) \leq F(t)$ holds for all $t \in \mathbb{R}$ while there exists a $t_1 \in D'_F$ such that $t_1 > G^{-1}(F(t_1)) = \inf\{s \in \mathbb{R} : G(s) \geq F(t_1)\}$. This yields $G(t) = F(t) = F(t_1)$ for all $t \in [G^{-1}(F(t_1)), t_1]$. In particular, there exists a non-degenerate interval $I_0 = [G^{-1}(F(t_1)), t_1]$ and a $p_0 = F(t_1)$ such that $F(I_0) = G(I_0) = \{p_0\}$ holds true.

In order to prove the other asserted implication, we now assume that there exists a non-degenerate interval $I_0 \subseteq \mathbb{R}$ and a number $p_0 \in (0,1)$ satisfying $F(I_0) = G(I_0) = \{p_0\}$. We choose I_0 to be of maximal length. This yields $\Delta_{FG}(t) = G^{-1}(F(t)) - t = G^{-1}(p_0) - t = \inf I_0 - t < 0$ for all $t \in I_0 \setminus \inf I_0$. Since I_0 is non-degenerate by assumption, the set $\{t \in D'_F : \Delta_{FG}(t) < 0\}$ is not empty. The proof is completed by the fact that the cdf's F and G of $X \sim \text{Bin}(1, \frac{1}{2})$ and $Y = X + \frac{1}{2}$ satisfy the assumption with $I_0 = [\frac{1}{2}, 1)$ and $p_0 = \frac{1}{2}$, and also satisfy $F \leq_{st} G$ (see left and central panel of Figure 6.1).

In particular, F and G having interval support is a sufficient condition for the equivalence of $\Delta_{FG}(t) \geq 0$ for all $t \in D'_F$ and $F \leq_{st} G$. Müller and Stoyan (2002, p. 4) noted that this equivalence does not generally hold, but they gave a wrong sufficient condition by requiring F and G to be continuous. They also gave an alternative characterization of the stochastic order using RIDF's with the advantage that it is generally valid.

Proposition 6.2. *For $F, G \in \mathcal{P}$, $\Delta_{GF}(t) \leq 0$ for all $t \in D'_G$ is equivalent to $F \leq_{st} G$.*

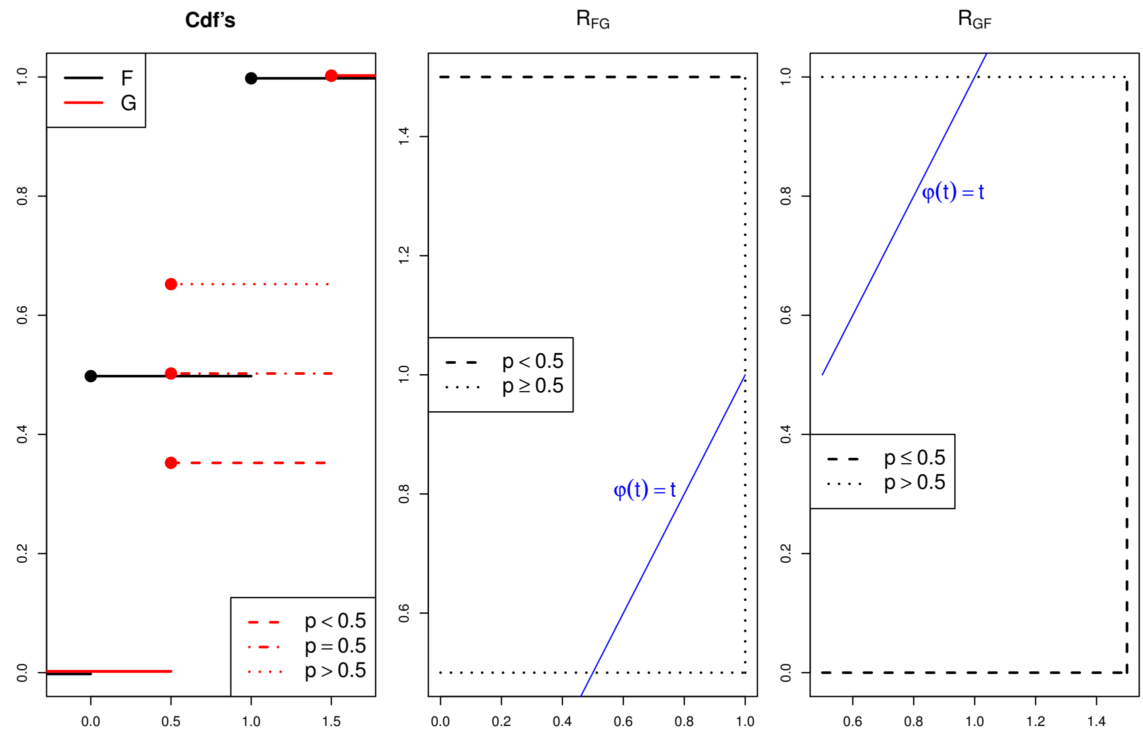

Figure 6.1.: Illustration of Example 6.3.

The differences between the three characterizations discussed so far are illustrated in the following example.

Example 6.3. Let $X \sim \text{Bin}(1, \frac{1}{2})$ and $Y = \tilde{Y} + \frac{1}{2}$ with $\tilde{Y} \sim \text{Bin}(1, 1-p)$ and $p \in (0,1)$. Obviously, $X \leq_{st} Y$ holds, if and only if $p \leq \frac{1}{2}$ (see left panel of Figure 6.1). Furthermore, we have $D'_F = [0,1]$, $D'_G = [\frac{1}{2}, \frac{3}{2}]$ and

$$R_{FG}(t) = \begin{cases} \frac{3}{2}, & \text{, if } p < \frac{1}{2}, \\ \frac{1}{2} + \mathbb{1}_{\{1\}}(t), & \text{, if } p \geq \frac{1}{2} \end{cases}, \qquad R_{GF}(t) = \begin{cases} \mathbb{1}_{\{\frac{3}{2}\}}(t), & \text{, if } p \leq \frac{1}{2}, \\ 1, & \text{, if } p > \frac{1}{2} \end{cases}$$

(see the two right panels of Figure 6.1). It follows that

$$\begin{aligned} \Delta_{GF}(t) \leq 0 \ \forall t \in D'_G \ &\Leftrightarrow \ p \leq \tfrac{1}{2} \ \Leftrightarrow \ F \leq_{st} G, \\ \Delta_{FG}(t) \geq 0 \ \forall t \in D'_F \ &\Leftrightarrow \ p < \tfrac{1}{2}. \end{aligned}$$

The difference between the two characterizations in the case $p = \frac{1}{2}$ is due to the utilized definition of the quantile function. If, instead, the definition $F^{-1}(q) = \sup\{t \in \mathbb{R} : F(t) \leq q\}, q \in (0,1)$, were used, we would obtain $R_{FG} \equiv \frac{3}{2}$ and $R_{GF} \equiv 1$ in the case $p = \frac{1}{2}$. Hence, the roles of the two characterizations would then be reversed.

A re-examination of the proofs of Propositions 6.1 and 6.2 yields that, for the alternative definition $F^{-1}(q) = \sup\{t \in \mathbb{R} : F(t) \le q\}, q \in (0,1)$, of the quantile function, the roles of the two characterizations of the stochastic order are generally reversed. This means that $\Delta_{FG}(t) \ge 0$ for all $t \in D'_F$ is then equivalent to $F \le_{st} G$. Furthermore, $\Delta_{GF}(t) \le 0$ for all $t \in D'_G$ then only implies $F \le_{st} G$, and equivalence holds, if and only if the condition in Proposition 6.1b) is fulfilled.

Relative inverse distribution functions on discrete distributions are discussed in detail in Appendix A. It is concluded that their use is generally not advisable because of Proposition A.3. It states that the crucial property of the RIDF from F to G, which is given by $R_{FG}(X) \stackrel{\mathcal{D}}{=} Y$, holds, if and only if $G(D_G) \subseteq \overline{F(D_F)}$. Since that condition is usually not fulfilled for two discrete cdf's F and G, the function Δ_{FG} then does not compare an F-distributed random variable and a G-distributed random variable. Instead, the second random variable has the closest cdf to G that can be obtained by (deterministically) transforming a F-distributed random variable. As a possible solution that is only partially successful, random transformations are introduced in Appendix A as a substitute. Note that, in these considerations, the domain of R_{FG} is always restricted to $\text{supp}(F)$. This is done to disregard all points that are not associated with any probability mass, which are not relevant for the crucial property of RIDF's.

The fact that RIDF's of discrete distributions lose some relevant information can be made obvious by comparing their graph to the corresponding Q-Q-plot (quantile-quantile-plot). The Q-Q-plot of two distributions or cdf's $F, G \in \mathcal{P}$ is the graphical representation of all points in the set $\{(F^{-1}(p), G^{-1}(p)) : p \in (0,1)\}$. Hence, it is very similar to the graph of the corresponding RIDF R_{FG}, but without the limitation that each point on the x-axis can only be assigned one point on the y-axis. This limitation is not relevant for distributions in $\mathcal{P}_I$, which is why the two plots coincide in that case. For discrete distributions, there are some discrepancies, which are analyzed exemplarily in the following. Here, the domain of the RIDF R_{FG} and the modified RIDF Δ_{FG} is restricted to $\text{supp}(F)$ for previously explained reasons.

Example 6.4. We define X and Y as in Example 6.3 and limit ourselves to the consideration of the RIDF R_{FG} and the corresponding Q-Q-plots. The graphs of the RIDF's only contain two points each because their domain is restricted to $\text{supp}(F) = \{0, 1\}$. The results are depicted in Figure 6.2.

Specifically the upper right panel shows that considering the RIDF on the domain $\text{supp}(F)$ yields worse results than on the domain D'_F. It suggests that $F \le_{st} G$ holds for $p > \frac{1}{2}$ because all points of the RIDF lie above the main diagonal. However, the corresponding Q-Q-plot in the lower right panel shows that there exists another point that is associated with positive probability mass and that lies below the main diagonal.

Generally, the fact that all points in the Q-Q-plot lie above or on the main diagonal is equivalent to $F^{-1}(p) \le G^{-1}(p)$ for all $p \in (0,1)$, and therefore to $F \le_{st} G$. Thus, the Q-Q-plot is the optimal graphical tool for the comparison of two distributions with respect to the usual

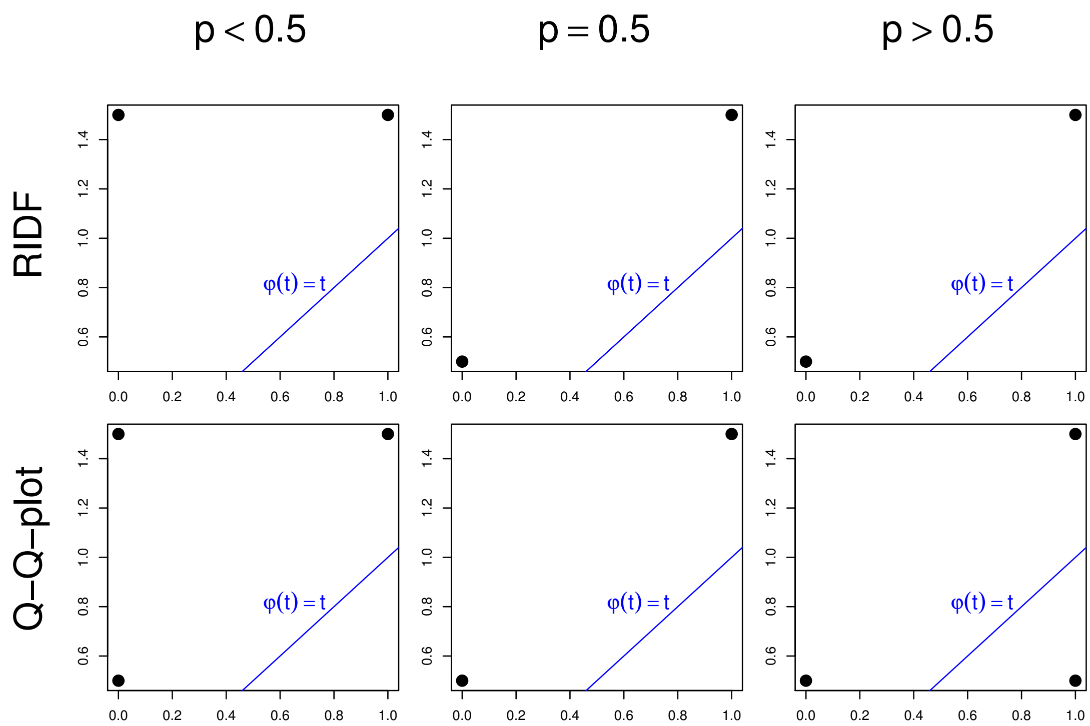

Figure 6.2.: Illustration of Example 6.4.

stochastic order. This is also mentioned by Müller and Stoyan (2002, p. 3–4) and Belzunce et al. (2015, p. 36).

Both of these references also consider the same methodology based on the so-called P-P-plot (probability-probability-plot), which is the graphical representation of all points in the set $\{(F(t), G(t)) : t \in \mathbb{R}\}$. It is contained within the set $[0, 1] \times [0, 1]$ and, if $F, G \in \mathcal{P}_I$, it coincides with the graph of the function $G \circ F^{-1}$. This function can be seen as a kind of alternative RIDF and was extensively considered by Handcock and Morris (1998, 1999), including applications to social sciences. However, while its graph is easier to overview because of its boundedness, it does not have the crucial property of the original RIDF based on the probability-integral-transform. Because this property plays a key role in the use of the RIDF for comparing distributions, we restrict ourselves to Q-Q-plots and the original RIDF's. The results in Propositions 6.1 and 6.2 as well as the subsequent discussion concerning Q-Q-plots can nonetheless mostly be replicated for P-P-plots and the corresponding alternative RIDF's. For this, note that $F \leq_{st} G$ is equivalent to $G(t) \leq F(t)$ for all $t \in \mathbb{R}$ and see Müller and Stoyan (2002, p. 4) and Belzunce et al. (2015, pp. 35–36).

6.2. The Dispersive Order

The utilization of RIDF's to compare discrete distributions with respect to the dispersive order $\leq_{disp}$ comes with similar problems as for the stochastic order $\leq_{st}$. According to Corollary 2.12, $F \leq_{disp} G$ is equivalent to $\Delta'_{FG}(t) \geq 0$ for all $t \in D_F$, if $F, G \in \mathcal{P}_I^1$. This characterization can be made more suitable for discrete distributions by extending the domain of Δ_{FG} to D'_F and by reformulating the requirement that the function is increasing: the characterization is then given by

$$\Delta_{FG}(t) - \Delta_{FG}(s) \geq 0 \quad \forall s, t \in D'_F \text{ with } s < t. \tag{6.1}$$

The following example shows that both $\mathrm{supp}(F)$ and D'_F as choices for the domain of Δ_{FG} lead to counterintuitive behaviour of an dispersion order based on (6.1) when it is applied to discrete distributions.

Example 6.5. a) We first consider D'_F as the domain of Δ_{FG}. Let $X \sim \mathcal{U}(\{1,2\})$ and $Y \sim \mathcal{U}(\{1,2,3,4\})$, so $\mathbb{P}(X = 1) = \frac{1}{2} = \mathbb{P}(X = 2)$ and $\mathbb{P}(Y = k) = \frac{1}{4}$ for all $k \in \{1,2,3,4\}$. Then, it is easy to verify that $X \leq_{disp} Y$ holds. For that, let $p_0, p_1 \in (0,1)$ with $p_0 < p_1$. If $p_0, p_1 \in (0, \frac{1}{2}]$ or $p_0, p_1 \in (\frac{1}{2}, 1)$, then $F^{-1}(p_1) - F^{-1}(p_0) = 0$, which directly yields $F^{-1}(p_1) - F^{-1}(p_0) \leq G^{-1}(p_1) - G^{-1}(p_0)$. In the only remaining case $p_0 \leq \frac{1}{2} < p_1$, it follows that $F^{-1}(p_1) - F^{-1}(p_0) = 1 \leq G^{-1}(p_1) - G^{-1}(p_0)$, which overall proves $X \leq_{disp} Y$. However, for $s, t \in [1,2) \subset [1,2] = D'_F$ with $s < t$, we obtain

$$\Delta_{FG}(t) - \Delta_{FG}(s) = (2 - t) - (2 - s) = s - t < 0.$$

Thus, although Y is clearly more dispersed than X, which is also recognized by the dispersive order, the characterization in (6.1) does not come to this conclusion.

b) Now we consider the characterization in (6.1), where D'_F is replaced by $\mathrm{supp}(F)$. Let $X \sim \mathrm{Bin}(1, \frac{1}{2})$ and $Y \sim \mathrm{Bin}(1, 1-\pi)$, where $\pi \in (\frac{1}{2}, 1)$. It is easy to see that $X \not\leq_{disp} Y$ and $Y \not\leq_{disp} X$ holds, because of $F^{-1}(p_1) - F^{-1}(p_0) = 0 < G^{-1}(p_1) - G^{-1}(p_0)$ for $\frac{1}{2} < p_0 \leq \pi < p_1 < 1$ and $G^{-1}(p_1) - G^{-1}(p_0) = 0 < F^{-1}(p_1) - F^{-1}(p_0)$ for $0 < p_0 \leq \frac{1}{2} < p_1 < \pi$. On the other hand, note that $\mathrm{supp}(F) = \{0,1\}$ and

$$\Delta_{FG}(1) - \Delta_{FG}(0) = (G^{-1}(1) - 1) - G^{-1}(\tfrac{1}{2}) = (1 - 1) - 0 = 0.$$

Thus, (6.1) is fulfilled for $\mathrm{supp}(F)$ instead of D'_F, meaning that G is deemed more dispersed than F. However, this can neither be confirmed intuitively (see left panel of Figure 6.1), nor by way of the dispersive order $\leq_{disp}$ or popular dispersion measures like the standard deviation, which gives $\sigma_F = \frac{1}{2} > \sqrt{\pi(1-\pi)} = \sigma_G$. The distribution of G also converges to a one-point-distribution in zero for $\pi \nearrow 1$, which exhibits no dispersion at all.

Other than RIDF's, Q-Q-plots can again be used to equivalently characterize the dispersive order. $F \leq_{disp} G$ holds, if and only if any straight line connecting two points in the corresponding Q-Q-plot has a slope of at least one. In spite of this graphical characterization, the dispersive order has a major flaw when it is applied to discrete distribution. The corresponding result was derived by Müller and Stoyan (2002, p. 41).

Proposition 6.6. *Let $F, G \in \mathcal{P}$. Then, $F \leq_{disp} G$ implies $F(D_F) \subseteq G(D_G)$.*

Conversely, this means that, if neither range of two cdf's is a subset of the range of the other cdf, the two distributions are not ordered with respect to $\leq_{disp}$. This does not present an obstacle for (absolutely) continuous distributions since the range of their cdf's is always equal to the entire unit interval. Discrete cdf's, however, take on at most countable many values. This means that, if two discrete ranges were picked at random (via independent uniformly distributed random variables), the probability for the ranges to even coincide in one point would be zero. This problem persists when we consider specific families of distributions like the binomial, Poisson or geometric distributions. An exception is given by specific classes of empirical distributions. In particular, every pair of non-tied empirical distributions with the same sample size satisfies the condition in Proposition 6.6.

However, it is very easy to find examples where one distribution is unambiguously more dispersed than the other, but neither $F(D_F) \subseteq G(D_G)$ nor $G(D_G) \subseteq F(D_F)$ holds, which means that the two distributions are not comparable with respect to $\leq_{disp}$. A particularly simple example is given in the following.

Example 6.7. Let $X \sim \mathcal{U}(\{1,2\})$ and $Y \sim \mathcal{U}(\{1,\ldots,5\})$. We have $F(D_F) = \{\frac{1}{2}\}$ and $G(D_G) = \{\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}\}$, thus, $F \not\leq_{disp} G$ (and $G \not\leq_{disp} F$). The situation here is very similar to that in Example 6.5a), except that Y is stretched out more because its probability mass is spread onto an additional point. However, $X \leq_{disp} \tilde{Y}$ holds for $\tilde{Y} \sim \mathcal{U}(\{1,2,3,4\})$, but $X \leq_{disp} Y$ does not hold. This situation is examined graphically in Figure 6.3.

The pmf's of X and Y are given in the upper left panel and the pmf's of X and $\tilde{Y}$ are given in the upper right panel. It is obvious that both $X \leq_D Y$ and $X \leq_D \tilde{Y}$ should hold for any sensible dispersion order $\leq_D$. The difference between X and Y is even larger than that between X and $\tilde{Y}$, yet $X \not\leq_{disp} Y$ and $X \leq_{disp} \tilde{Y}$. The reason for this becomes more obvious by observing the corresponding Q-Q-plots in the two middle panels of Figure 6.3. The critical point in both Q-Q-plots is the jump of F^{-1} from the value 1 to the value 2. On the left side, the slope between the two corresponding points is zero, which contradicts $X \leq_{disp} Y$; on the right side, the slope between these two points is one, and therefore $X \leq_{disp} \tilde{Y}$ is true.

The difference in the two Q-Q-plots can be explained by the comparison of the cdf's, which are depicted in the lower panels of Figure 6.3. It is colour-coded into the cdf's how the probability mass is shared between the different points in the supports of both distributions. Since this is essentially what is represented by the Q-Q-plots, the colours of the points in the Q-Q-plots correspond to those used for the cdf's. The reason for the slope of zero in the Q-Q-plot on the left side is that a smaller jump of G lies in between two larger jumps of

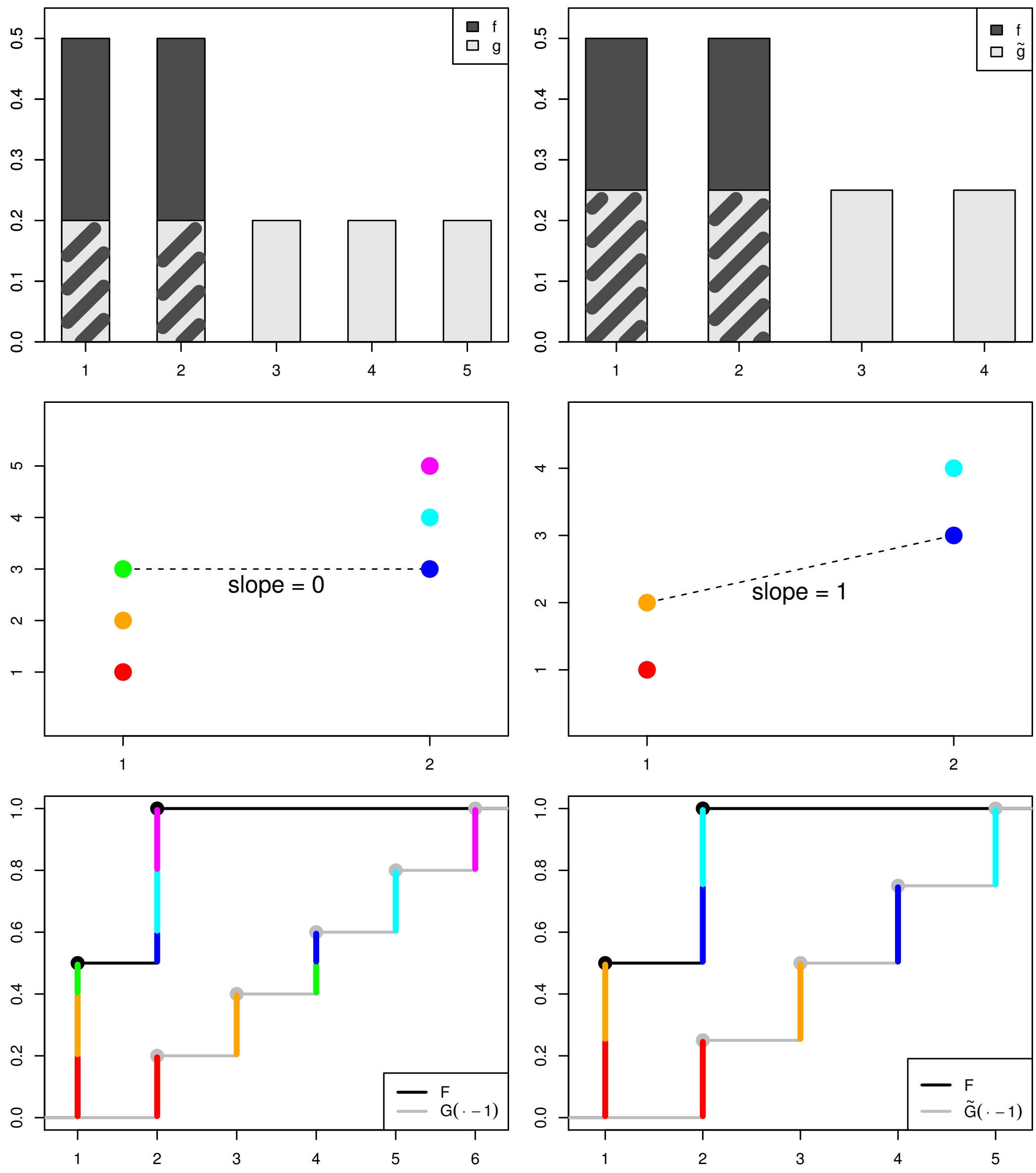

Figure 6.3.: Different illustrations of the two pairs of distributions considered in Example 6.7: $X \sim \mathcal{U}(\{1,2\})$ and $Y \sim \mathcal{U}(\{1,\ldots,5\})$ in the left panels; X and $\tilde{Y} \sim \mathcal{U}(\{1,2,3,4\})$ in the right panels. Upper: barplots of pmf's; Middle: Q-Q-plots; Lower: colour-coded cdf's.

F. This is not the case on the right side precisely because the condition $F(D_F) \subseteq G(D_G)$ is fulfilled.

The fact that constellations as in Example 6.7 exist, implies that $\leq_{disp}$ does not order discrete distributions sufficiently well with respect to dispersion. Since the dispersive order is made out to be the fundamental order of dispersion in Section 3.2.2, this leaves the concept of dispersion on discrete distributions without a foundation. In particular, the idea that popular dispersion measures like the ones discussed in Sections 3.1 and 3.2.2 actually measure dispersion does not have any basis in a discrete setting. Usually, this idea is based on the preservation of $\leq_{disp}$, which is not a meaningful dispersion order for discrete distributions.

A possible solution for this would be to replace $\leq_{disp}$ in the role as foundational order by another order of dispersion, for instance the dilation order $\leq_{dil}$. However, as discussed in Section 3.2.2, $\leq_{dil}$ is not well suited for this role. First, it is weaker than $\leq_{disp}$ and therefore possibly too specific in its interpretation of dispersion. While $F \leq_{disp} G$ means that G is more dispersed than F in a pointwise fashion, $F \leq_{dil} G$ represents a comparison of averages that can, e.g., be expressed via the stop-loss transform. Second, $\leq_{dil}$ implicitly centres the dispersion of any distribution around its mean, which inherently favours this kind of dispersion measure over others, which are, e.g., centred around the median.

All of the above observations suggest that a rigorous foundation of dispersion for discrete distributions requires an order that has the properties of the dispersive order but is more suitable for the discrete setting. Since a concept this widely used should not lack any rigour, Chapter 7 is dedicated to establishing this kind of order. Several proposals are developed and discussed and it is subsequently analyzed whether they are suitable discrete versions of the dispersive order $\leq_{disp}$. Finally, their behaviour on well known families of discrete distributions as well as their compatibility with popular measures of dispersion is examined.

CHAPTER 7

Discrete Dispersive Orders

Before starting to develop proposals for a discrete version of the dispersive order, we first establish our general setting. First, since $\mathcal{D}$ contains a number of difficult to handle distributions with virtually no practical use, we limit our considerations to the class of purposive discrete distributions given in the following definition along with a number of subclasses.

Definition 7.1. Let $F \in \mathcal{D}$ be a cdf with pmf f and let $X \sim F$.

a) The class of *purposive discrete distributions* $\mathcal{D}_0 \subseteq \mathcal{D}$ is defined by

$$\begin{aligned} F \in \mathcal{D}_0 &\Leftrightarrow \text{supp}(F) \text{ is order-isomorphic to a subset of } \mathbb{Z} \text{ with at least two elements} \\ &\Leftrightarrow \exists A \subseteq \mathbb{Z}, |A| \geq 2, \text{bijection } \varphi : \text{supp}(F) \to A \text{ such that} \\ &\quad\ x \leq y \Leftrightarrow \varphi(x) \leq \varphi(y)\ \forall x, y \in \text{supp}(F). \end{aligned}$$

b) For each $n \in \mathbb{N}_{\geq 2}$, the class of *empirical distributions with sample size* n, $\mathcal{E}(n)$, is defined by

$$\begin{aligned} F \in \mathcal{E}(n) \Leftrightarrow\ & \exists x \in \mathbb{R}^n \text{ such that} \\ & \mathbb{P}(X = x_j) = \frac{|\{k \in \{1, \ldots, n\} : x_k = x_j\}|}{n} \quad \forall j \in \{1, \ldots, n\} \\ \Leftrightarrow\ & F(\mathbb{R}) \subseteq \left\{ \tfrac{i}{n} : i \in \{0, \ldots, n\} \right\} \\ \Leftrightarrow\ & f(\mathbb{R}) \subseteq \left\{ \tfrac{i}{n} : i \in \{0, \ldots, n\} \right\}, \end{aligned}$$

if F is additionally non-degenerate. The class of all *empirical distributions* is defined by $\mathcal{E} = \bigcup_{n \in \mathbb{N}_{\geq 2}} \mathcal{E}(n)$. The vector x in the first characterization is said to be the *defining vector* of F or the corresponding distribution.

c) For each $n \in \mathbb{N}_{\geq 2}$, the class of *non-tied empirical distributions with sample size* n, $\mathcal{E}_{nt}(n)$, is defined by

$$\begin{aligned} F \in \mathcal{E}_{nt}(n) &\Leftrightarrow \exists x \in \mathbb{R}^n \text{ with } x_i \neq x_j \text{ for } i \neq j \text{ such that} \\ &\qquad \mathbb{P}(X = x_j) = \frac{1}{n} \quad \forall j \in \{1, \dots, n\} \\ &\Leftrightarrow F(\mathbb{R}) = \left\{\tfrac{i}{n} : i \in \{0, \dots, n\}\right\} \\ &\Leftrightarrow f(\mathbb{R}) = \{0, \tfrac{1}{n}\}. \end{aligned}$$

The class of all *non-tied empirical distributions* is defined by $\mathcal{E}_{nt} = \bigcup_{n \in \mathbb{N}_{\geq 2}} \mathcal{E}_{nt}(n)$.

d) The class of $\mathbb{Z}$-*lattice distributions* $\mathcal{LD}(\mathbb{Z})$ is defined by

$$F \in \mathcal{LD}(\mathbb{Z}) \Leftrightarrow \exists a > 0, b \in \mathbb{R} \text{ such that } \operatorname{supp}(F) = a\mathbb{Z} + b.$$

Let $n \in \mathbb{N}_{\geq 2}$. The classes of $\mathbb{N}$-, $(-\mathbb{N})$- *and* n-*lattice distributions* $\mathcal{LD}(\mathbb{N}), \mathcal{LD}(-\mathbb{N}), \mathcal{LD}(n)$ are defined analogously by replacing $\mathbb{Z}$ by $\mathbb{N}$, $-\mathbb{N}$ or $\{1, \dots, n\}$ in the definition. The positive number a is said to be the *defining distance* of F or the corresponding distribution.

e) The class of *lattice distributions* $\mathcal{LD}$ is defined by

$$F \in \mathcal{LD} \Leftrightarrow F \in \mathcal{LD}(N), \text{ where } N \in \mathbb{N}_{\geq 2} \text{ or } N \in \{\mathbb{N}, -\mathbb{N}, \mathbb{Z}\}.$$

This directly implies the inclusions $\mathcal{E}_{nt}(n) \subseteq \mathcal{E}_{nt} \subseteq \mathcal{D}_0$, $\mathcal{E}_{nt}(n) \subseteq \mathcal{E}(n) \subseteq \mathcal{E} \subseteq \mathcal{D}_0$ for all $n \in \mathbb{N}_{\geq 2}$, and $\mathcal{LD}(N) \subseteq \mathcal{LD} \subseteq \mathcal{D}_0$ for all $N \in \mathbb{N}_{\geq 2}$ or $N \in \{\mathbb{N}, -\mathbb{N}, \mathbb{Z}\}$. Furthermore, note that there indeed exist non-degenerate discrete distributions in $\mathcal{D} \setminus \mathcal{D}_0$, i.e. discrete distributions that are not order-isomorphic to a subset of the whole numbers. Consider the following example.

Example 7.2. Let the distribution of the random variable X be defined by

$$\mathbb{P}\left(X = \tfrac{1}{k}\right) = \left(\tfrac{1}{2}\right)^{|k|+1}$$

for all $k \in \mathbb{Z} \setminus \{0\}$. This constitutes a probability distribution because of

$$\sum_{k \in \mathbb{Z} \setminus \{0\}} \mathbb{P}\left(X = \tfrac{1}{k}\right) = 2 \cdot \sum_{k=1}^{\infty} \left(\tfrac{1}{2}\right)^{k+1} = \sum_{k=1}^{\infty} \left(\tfrac{1}{2}\right)^{k} = 1.$$

Now assume that there exists a subset $A \subseteq \mathbb{Z}$ with $|A| \geq 2$ and an order-isomorphism $\varphi : \operatorname{supp}(X) \to A$. Let $j = \varphi(-1) \in A$. Then, because φ is an order-isomorphism, $\varphi(-\frac{1}{1+k}) \geq j + k$ for all $k \in \mathbb{N}$. Thus, $\varphi(1) \geq j + k$ for all $k \in \mathbb{N}$, which poses a contradiction to $\varphi(1) \in A \subseteq \mathbb{Z}$.

However, examples of this kind are rather pathological in nature and usually do not appear in practice. The class $\mathcal{D}_0$ particularly contains all lattice distributions and empirical distributions, which are the most frequently used kinds of discrete distributions.

The following result, which is crucial for the remainder of this thesis, is only valid for distributions in $\mathcal{D}_0$.

Proposition 7.3. *Define*

$$\mathbb{I} = \{\mathbb{Z}, \mathbb{N}, -\mathbb{N}\} \cup \{\{1, \ldots, n\} : n \in \mathbb{N}_{\geq 2}\}$$

and

$$\mathbb{S}_A = \left\{ (x_j, p_j)_{j \in A} \subseteq \mathbb{R} \times (0, 1] : x_i < x_j \text{ for } i < j,\ p_j > 0 \text{ for } j \in A,\ \sum_{j \in A} p_j = 1 \right\}$$

for $A \in \mathbb{I} \setminus \mathbb{Z}$ as well as

$$\mathbb{S}_\mathbb{Z} = \left\{ (x_j, p_j)_{j \in \mathbb{Z}} \subseteq \mathbb{R} \times (0, 1] : x_i < x_j \text{ for } i < j,\ p_j > 0 \text{ for } j \in \mathbb{Z},\ \sum_{j \in \mathbb{Z}} p_j = 1, \right.$$
$$\left. \inf\{j \in \mathbb{Z} : \sum_{i \leq j} p_i \geq \tfrac{1}{2}\} = 0 \right\}.$$

For any $F \in \mathcal{D}_0$, there exists a unique index set $A \in \mathbb{I}$ that is order-isomorphic to supp(F)*, and there exists a unique sequence $(x_j, p_j)_{j \in A} \in \mathbb{S}_A$ such that $\mathbb{P}(X = x_j) = p_j$ for all $j \in A$. This unique association is denoted by $F \mathrel{\widehat{=}} (A, (x_j, p_j)_{j \in A})$. A is said to be the* indexing set *of F and $(x_j, p_j)_{j \in A}$ is said to be the* identifying sequence *of F.*

Proof. We define a function $\varphi : \mathcal{D}_0 \to \bigcup_{A \in \mathbb{I}} \mathbb{S}_A$ and show that it is a bijection, which is an even stronger statement than the assertion. Note that the codomain of φ is a disjoint union of the sets $\mathbb{S}_A, A \in \mathbb{I}$. The value assignment of φ is now defined by cases. Therefore, let $F \in \mathcal{D}_0$, then supp(F) is order-isomorphic to a subset of $\mathbb{Z}$. Note that since minima and maxima are defined via the order $\leq$, this order-isomorphism preserves minima and maxima.

Case 1: $\min(\operatorname{supp}(F))$ and $\max(\operatorname{supp}(F))$ both exist.

A subset of $\mathbb{Z}$ has a minimum and a maximum, if and only if it is finite. Therefore, supp(F) is also finite. Let $n = |\operatorname{supp}(F)| \in \mathbb{N}_{\geq 2}$. Define $x_1 = \min(\operatorname{supp}(F))$ and $x_j = \min(\operatorname{supp}(F) \setminus \{x_1, \ldots, x_{j-1}\}), j = 2, \ldots, n$. Then, we define $\varphi(F) = (x_j, p_j)_{j \in \{1,\ldots,n\}} \in \mathbb{S}_{\{1,\ldots,n\}}$, where $p_j = \mathbb{P}(X = x_j) > 0, j = 1, \ldots, n$. Note that all of the steps of the value assignment in this case are unique, thus ensuring injectivity in this case.

Case 2: $\min(\operatorname{supp}(F))$ exists, but $\max(\operatorname{supp}(F))$ does not.

The only set within $\mathbb{I}$ with existing minimum but non-existing maximum is $\mathbb{N}$. Similarly to

Case 1, define $x_1 = \min(\mathrm{supp}(F))$ and $x_j = \min(\mathrm{supp}(F) \setminus \{x_1, \ldots, x_{j-1}\}), j = 2, 3, \ldots$. Then, the definition $\varphi(F) = (x_j, p_j)_{j \in \mathbb{N}} \in \mathbb{S}_{\mathbb{N}}$ with $p_j = \mathbb{P}(X = x_j) > 0, j = 1, 2, \ldots$, is once again unique.

Case 3: $\max(\mathrm{supp}(F))$ exists, but $\min(\mathrm{supp}(F))$ does not.
This is analogous to Case 2 by simply swapping to roles of minima and maxima and replacing $\mathbb{N}$ by $(-\mathbb{N})$.

Case 4: $\min(\mathrm{supp}(F))$ and $\max(\mathrm{supp}(F))$ both do not exist.
The only set within $\mathbb{I}$ with non-existing minimum and non-existing maximum is $\mathbb{Z}$. We now define

$$\begin{aligned}
x_0 &= F^{-1}(\tfrac{1}{2}) = \inf\{t \in \mathbb{R} : F(t) \geq \tfrac{1}{2}\} \in \mathrm{supp}(F),\\
x_j &= \min\{t \in \mathrm{supp}(F) \setminus \{x_0, \ldots, x_{j-1}\} : F(t) \geq \tfrac{1}{2}\}, \quad j = 1, 2, \ldots,\\
x_j &= \max\{t \in \mathrm{supp}(F) \setminus \{x_{j+1}, \ldots, x_0\} : F(t) < \tfrac{1}{2}\}, \quad j = -1, -2, \ldots.
\end{aligned}$$

Defining $\varphi(F) = (x_j, p_j)_{j \in \mathbb{Z}} \in \mathbb{S}_{\mathbb{Z}}$ with $p_j = \mathbb{P}(X = x_j) > 0, j \in \mathbb{Z}$, now once again leads to a unique value assignment in this case.

It is ensured in every case separately that φ is well-defined and injective. Now let $A \in \mathbb{I}$ and $(x_j, p_j)_{j \in A} \in \mathbb{S}_A$. We define a cdf F by $\mathbb{P}(X = x_j) = p_j > 0, j \in A$. It follows directly that $\mathrm{supp}(F) = \{x_j : j \in A\}$ is order-isomorphic to $A \subseteq \mathbb{Z}$ with $|A| \geq 2$. This implies $F \in \mathcal{D}_0$ and following the above value assignment for φ yields $\varphi(F) = (x_j, p_j)_{j \in A}$. Thus, φ is surjective and therefore a bijection. ▨

Note that if $\mathbb{S}_{\mathbb{Z}}$ was of the same structure as $\mathbb{S}_A$ for $A \neq \mathbb{Z}$, F could only be uniquely identified up to an arbitrary index shift of the identifying sequence. The additional condition in $\mathbb{S}_{\mathbb{Z}}$ assures that the index 0 relates to the median and thereby fixes the sequence.

Throughout the remainder of this thesis, let $F \mathrel{\hat{=}} (A, (x_j, p_j)_{j \in A})$ and $G \mathrel{\hat{=}} (B, (y_j, q_j)_{j \in B})$. Furthermore, we establish the conventions $x_a = -\infty$ and $F(x_a) = 0$ for $a < \min A$ as well as $x_a = \infty$ and $F(x_a) = 1$ for $a > \max A$, provided that the minimum and the maximum exist, respectively.

7.1. Derivation of Discrete Dispersive Orders

The reason for the need for a discrete dispersive order is given by the result of Proposition 6.6 in spite of the existence of situations like Example 6.7. The goal is to construct a dispersion order that is meaningfully applicable to discrete distributions and therein assumes the same role as the dispersive order $\leq_{disp}$ has for continuous distribution. This presents us with two starting points for our derivation. The first is to find a representation of the dispersive order for continuous distributions that has an easily applicable discrete analogue. The second is to find out what exactly is required by the original dispersive order $\leq_{disp}$ at the edge of its applicability to discrete distributions.

We start out by considering the second starting point, postponing the first one to afterwards. To this end, the following result refines Proposition 6.6 for discrete distributions as is gives an equivalent characterization of $\leq_{disp}$ in that case. The one-dimensional Lebesgue measure is denoted by λ^1. Consequently, for any $F \in \mathcal{D}_0$ and $p \in F(D_F)$, $\lambda^1(F^{-1}(\{p\}))$ describes how long F assumes the value p.

Proposition 7.4. *Let $F, G \in \mathcal{D}$. Then $F \leq_{disp} G$ is equivalent to*

$$F(D_F) \subseteq G(D_G) \quad \textit{and} \quad \lambda^1(F^{-1}(\{p\})) \leq \lambda^1(G^{-1}(\{p\}))\ \forall p \in F(D_F).$$

Proof. We start by proving the implication from left to right. Due to Proposition 6.6, only the inequality of the Lebesgue measures needs to be shown. This follows immediately as

$$\lambda^1(F^{-1}(\{p\})) = \lim_{r \searrow p}(F^{-1}(r) - F^{-1}(p)) \leq \lim_{r \searrow p}(G^{-1}(r) - G^{-1}(p)) = \lambda^1(G^{-1}(\{p\})) \tag{7.1}$$

holds for all $p \in F(D_F)$ by assumption since $F(D_F) \subseteq (0,1)$.

For the other implication, let $p, q \in (0,1), p < q$. Since F is discrete, the difference of its quantile function at p and q is equal to the summed lengths of all intervals, on which F is constant at a value between p and q. Thus,

$$F^{-1}(q) - F^{-1}(p) = \sum_{r \in F(D_F) \cap [p,q)} \lambda^1(F^{-1}(\{r\}))$$

and analogously for G. By assumption, we obtain

$$\begin{aligned} &\Big(G^{-1}(q) - G^{-1}(p)\Big) - \Big(F^{-1}(q) - F^{-1}(p)\Big) \\ = &\sum_{r \in F(D_F) \cap [p,q)} \Big(\lambda^1(G^{-1}(\{r\})) - \lambda^1(F^{-1}(\{r\}))\Big) + \sum_{r \in (G(D_G) \setminus F(D_F)) \cap [p,q)} \lambda^1(G^{-1}(\{r\})). \end{aligned}$$

Since both of these summands are non-negative, the assertion follows. ∎

Using this characterization, the following examples illustrate how the dispersive order manifests itself specifically on some of the distribution classes given in Definition 7.1.

Example 7.5. a) Let $n \in \mathbb{N}_{\geq 2}$ and $F, G \in \mathcal{E}_{nt}(n)$. Then there exist $x, y \in \mathbb{R}^n$ with $x_i < x_{i+1}$ and $y_i < y_{i+1}$ for all $i \in \{1, \dots, n-1\}$ such that $\mathbb{P}(X = x_i) = \frac{1}{n} = \mathbb{P}(Y = y_i)$ for all $i \in \{1, \dots, n\}$. Because of $F(D_F) = G(D_G) = \{\frac{i}{n} : i \in \{1, \dots, n-1\}\}$, the following equivalence holds

$$\begin{aligned} F \leq_{disp} G &\Leftrightarrow \lambda^1(F^{-1}(\{\tfrac{i}{n}\})) \leq \lambda^1(G^{-1}(\{\tfrac{i}{n}\})) \quad \forall i \in \{1, \dots, n-1\} \\ &\Leftrightarrow x_{i+1} - x_i \leq y_{i+1} - y_i \quad \forall i \in \{1, \dots, n-1\}. \end{aligned}$$

This means that G is at least as dispersed as F, if and only if the distance between every pair of neighbouring points in the support of F is smaller than between the corresponding pair in the support of F. Particularly, $y_{j+1} - y_j < x_{i+1} - x_i$ for $i \neq j$ does not contradict $F \leq_{disp} G$.

b) Let $G \in \mathcal{E}_{nt}(n)$ for some $n \in \mathbb{N}_{\geq 2}$ like in a) and let $F \in \mathcal{E}(n) \setminus \mathcal{E}_{nt}(n)$ have exactly one tie. Hence, there exists a unique $i_0 \in \{1, \dots, n-1\}$ such that $x_i < x_{i+1}$ for $i \neq i_0$ and $x_{i_0} = x_{i_0+1}$ holds for the defining vector $x \in \mathbb{R}^n$ of F. It follows that $F(D_F) = G(D_G) \setminus \{\frac{i_0}{n}\} \subset G(D_G)$ and therefore

$$\begin{aligned} F \leq_{disp} G &\Leftrightarrow \lambda^1(F^{-1}(\{\tfrac{i}{n}\})) \leq \lambda^1(G^{-1}(\{\tfrac{i}{n}\})) \quad \forall i \in \{1, \dots, n-1\} \setminus \{i_0\} \\ &\Leftrightarrow x_{i+1} - x_i \leq y_{i+1} - y_i \quad \forall i \in \{1, \dots, n-1\} \setminus \{i_0\}. \end{aligned}$$

Once again, the distances between neighbouring pairs of points in the supports of F and G are compared. Which pairs of points are compared depends on the value that the corresponding cdf takes on the interval between the points. For example, if $i_0 \neq 1$, the difference $x_2 - x_1$ is compared to $y_2 - y_1$ since $F((x_1, x_2)) = \{\frac{1}{n}\} = G((y_1, y_2))$. Note that the interval length $y_{i_0+1} - y_{i_0}$ is not compared to any interval length of F.

c) Let $F, G \in \mathcal{LD}$ both have the same defining distance $c > 0$. Then there exist $a, b \in \mathbb{R}$ and $A, B \in \mathbb{I}$ such that $\operatorname{supp}(F) = cA + a$ and $\operatorname{supp}(G) = cB + b$. It immediately follows for all $p \in F(D_F) = F(\operatorname{supp}(F)) \setminus \{1\}$ that

$$\lambda^1(F^{-1}(\{p\})) = a = \lambda^1(G^{-1}(\{p\})).$$

Thus, the statement of Proposition 7.4 simplifies to

$$F \leq_{disp} G \Leftrightarrow F(D_F) \subseteq G(D_G).$$

Proposition 7.4 specifies how any pair of discrete distributions can be compared with respect to dispersion. The dispersive order can only order pairs of cdf's within the set $\{(F, G) \in (\mathcal{D}_0)^2 : F(D_F) \subseteq G(D_G) \text{ or } G(D_G) \subseteq F(D_F)\}$ with respect to dispersion. For the pairs in this set, the first components of the identifying sequences are arbitrary, while the second components of the sequence are mostly fixed. We now turn our attention to a class of distributions, for which the first components of the identifying sequence are mostly fixed while the second components are arbitrary. Afterwards, the goal is to unite the two methodologies in some way to enable us to compare two purposive discrete distributions with respect to their dispersions.

Two discrete cdf's F and G cannot be ordered with respect to $\leq_{disp}$, if neither $F(D_F) \subseteq G(D_G)$ nor $G(D_G) \subseteq F(D_F)$ holds. However, it is not difficult to find an example where neither inclusion holds but one distribution is unambiguously more dispersed than the other, as evidenced by Example 6.7. The basic idea for how to compare F and G with respect to dispersion in this kind of situation is obtained through one of the starting points discussed at the beginning of Section 7.1. Specifically, the idea is to modify a characterization of $\leq_{disp}$ for sufficiently regular continuous distributions in such a way that it is applicable to discrete distributions.

Corollary 2.12 states that, for $F, G \in \mathcal{P}_I^1$, $F \leq_{disp} G$ is equivalent to $\Delta'_{FG}(t) \geq 0$ for all $t \in D_F$. According to the proof of Theorem 3.12b), this can be rewritten using the Lebesgue densities f and g of F and G. It is equivalent to

$$f(F^{-1}(p)) \leq g(G^{-1}(p)) \quad \forall p \in (0,1). \tag{7.2}$$

The discrete analogue of Lebesgue densities for absolutely continuous distributions are pmf's, which are also densities, only with respect to a suitable counting measure. Our proposed discrete generalization of the dispersion order is therefore obtained by taking characterization (7.2) of $\leq_{disp}$ and replacing the Lebesgue densities with the respective pmf's. Since the values of a pmf are the jump heights of the corresponding cdf, this gives a requirement concerning the second components of the identifying sequences in question. Here, we use the convention $F \mathrel{\hat{=}} (A, (x_j, p_j)_{j\in A})$ and $G \mathrel{\hat{=}} (B, (y_j, q_j)_{j\in B})$. We mostly fix the first components of the identifying sequences by assuming

$$x_a - x_{a-1} \leq y_b - y_{b-1} \quad \forall a \in A \setminus \{\min A\}, b \in B \setminus \{\min B\}, \tag{7.3}$$

For any set M with non-existent minimum, we define $\{\min M\} := \emptyset$; an analogous rule holds for maximums. This means that every interval, on which G is constant, is at least as long as any interval, on which F is constant. Since this puts more distance between the points in the support of G than between those in the support of F, this intuitively makes G more dispersed. Condition (7.3) is obviously equivalent to

$$\sup_{a\in A\setminus\{\min A\}} (x_a - x_{a-1}) \leq \inf_{b\in B\setminus\{\min B\}} (y_b - y_{b-1}). \tag{7.4}$$

With the rather strict condition (7.3)/(7.4), we come to our first definition of a discrete version of the dispersive order. Weakened versions of this order, with respect to the requirements on the supports, are presented at a later point.

Definition 7.6. Let $F, G \in \mathcal{D}_0$ have the pmf's f, g. Then, G is said to be *at least as discretely dispersed* as F, denoted by $F \leq_{disp}^{disc} G$, if (7.3) is satisfied and

$$g(G^{-1}(p)) \leq f(F^{-1}(p)) \quad \forall p \in (0,1). \tag{7.5}$$

We first apply this definition to the situation in Example 6.7 for which the original dispersive order is not sufficient. This is done in order to find out whether $\leq_{disp}^{disc}$ is a suitable idea for its intended purpose.

Example 7.7. a) Continuation of Example 6.7: Condition (7.3) is satisfied since $x_a - x_{a-1} = 1 = y_b - y_{b-1}$ holds for all $a \in A \setminus \{\min A\}, b \in B \setminus \{\min B\}$. Additionally, $\text{range}(F^{-1}) = \text{supp}(F)$ and $\text{range}(G^{-1}) = \text{supp}(G)$. Since f is constantly equal to $\frac{1}{2}$ on its support and g is constantly equal to $\frac{1}{5}$ on its support, we obtain

$$g(G^{-1}(p)) = \frac{1}{5} \leq \frac{1}{2} = f(F^{-1}(p)) \quad \forall p \in (0,1).$$

Thus, $F \leq_{disp}^{disc} G$ holds as anticipated.

b) The first observation from part a) can be generalized to the entire set of lattice distributions. For that, let $F, G \in \mathcal{LD}$ have the defining distances $c_F, c_G > 0$, respectively. Now, there exist $d_F, d_G \in \mathbb{R}$ and sets $S_F, S_G \subset \mathbb{Z}$ such that $A = d_F + c_F \cdot S_F$ and $B = d_G + c_G \cdot S_G$. Hence,

$$\sup_{a \in A \setminus \{\min A\}} (x_a - x_{a-1}) = x_a - x_{a-1} = c_F \quad \forall a \in A \setminus \{\min A\},$$
$$\inf_{b \in B \setminus \{\min B\}} (y_b - y_{b-1}) = y_b - y_{b-1} = c_G \quad \forall b \in B \setminus \{\min B\},$$

yielding that condition (7.3) is equivalent to $c_F \leq c_G$. In particular, (7.3) is satisfied if the lattice distributions F and G have the same defining distance, so the same distance between neighbouring points in their respective supports. For counting distributions like the binomial, Poisson or geometric distribution, this distance is equal to 1. For these distributions, $F \leq_{disp}^{disc} G$ holds if (7.5) is fulfilled.

Example 7.7a) is particularly simple because the two pmf's are constant on their supports. This somewhat hides the fact that the order $\leq_{disp}^{disc}$ does not require that all values of f are compared with all values of g; only the comparison between specific pairs of values are relevant. Formulated with respect to the cdf's F and G, the values to be compared are the heights of their jumps. However, the pairs of jumps to be compared are decided upon by the values of the cdf's, so the sum of all jumps up to that point. In the following, we introduce a relation that specifies which pairs of jumps have to be compared for any pair of purposive discrete distributions.

Definition 7.8. Let $F, G \in \mathcal{D}_0$. Then, the relation $\overset{F,G}{\leftrightharpoons}$ on the set $A \times B$ is defined by

$$a \overset{F,G}{\leftrightharpoons} b \Leftrightarrow \exists r \in (0,1) : F^{-1}(r) = x_a, G^{-1}(r) = y_b$$

for $a \in A, b \in B$. The set $R(\overset{F,G}{\leftrightharpoons})$ of all $(a,b) \in A \times B$ with $a \overset{F,G}{\leftrightharpoons} b$ is said to be the set of (F,G)-*dispersion-relevant* pairs of indices.
If F and G are fixed, we write $\leftrightharpoons$ instead of $\overset{F,G}{\leftrightharpoons}$.

The definition of $\leftrightharpoons$ is directly informed by Definition 7.6. To illustrate, consider the following result.

Proposition 7.9. *Let $F, G \in \mathcal{D}_0$ satisfy (7.3). Then, $F \leq_{disp}^{disc} G$ is equivalent to $q_b \leq p_a$ for all $(a,b) \in R(\leftrightharpoons)$.*

Proof. The following chain of equivalences proves the assertion:

$$\begin{aligned} F \leq_{disp}^{disc} G &\Leftrightarrow g(G^{-1}(r)) \leq f(F^{-1}(r)) && \forall r \in (0,1) \\ &\Leftrightarrow g(y_b) \leq f(x_a) && \forall (a,b) \in A \times B \text{ such that} \\ & && \quad F^{-1}(r) = x_a, G^{-1}(r) = y_b \text{ for some } r \in (0,1) \\ &\Leftrightarrow g(y_b) \leq f(x_a) && \forall (a,b) \in R(\leftrightharpoons) \\ &\Leftrightarrow \mathbb{P}(Y = y_b) \leq \mathbb{P}(X = x_a) && \forall (a,b) \in R(\leftrightharpoons) \\ &\Leftrightarrow q_b \leq p_a && \forall (a,b) \in R(\leftrightharpoons) \end{aligned}$$

The second equivalence holds due to the fact that $\operatorname{range}(F^{-1}) = \operatorname{supp}(F)$ and $\operatorname{range}(G^{-1}) = \operatorname{supp}(G)$. ▨

Before putting this new relation to use on our simple example from before, we give an equivalent characterization that is easier to handle.

Proposition 7.10. *Let $F, G \in \mathcal{D}_0$. For $a \in A, b \in B$, we have*

$$a \leftrightharpoons b \Leftrightarrow (F(x_{a-1}), F(x_a)) \cap (G(y_{b-1}), G(y_b)) \neq \emptyset.$$

Proof. Let $(a,b) \in A \times B$. Then,

$$\begin{aligned} F^{-1}(r) = x_a &\Leftrightarrow r \in (F(x_{a-1}), F(x_a)], \\ G^{-1}(r) = y_b &\Leftrightarrow r \in (G(y_{b-1}), G(y_b)]. \end{aligned}$$

So the existence of an $r \in (0,1)$ such that $F^{-1}(r) = x_a$ and $G^{-1}(r) = y_b$ is equivalent to the existence of an $r \in (F(x_{a-1}), F(x_a)] \cap (G(y_{b-1}), G(y_b)]$. This, in turn, is equivalent to that set being non-empty.

It remains to be shown that $(c_1, c_2] \cap (d_1, d_2] \neq \emptyset$ implies $(c_1, c_2) \cap (d_1, d_2) \neq \emptyset$ for $c_1, c_2, d_1, d_2 \in \mathbb{R}$ with $c_1 < c_2, d_1 < d_2$. If that implication is not true, the four constants can be chosen in such a way that $\emptyset \neq (c_1, c_2] \cap (d_1, d_2] \subseteq \{c_2, d_2\}$. This, however, is contradicted by the implication

$$\begin{aligned} c_2 \in (c_1, c_2] \cap (d_1, d_2] &\Rightarrow c_2 > d_1 \wedge c_2 \leq d_2 \\ &\Rightarrow \exists \varepsilon > 0 : c_2 - \varepsilon > d_1 \wedge c_2 - \varepsilon \leq d_2 \wedge c_2 - \varepsilon > c_1 \\ &\Rightarrow \exists \varepsilon > 0 : c_2 - \varepsilon \in (c_1, c_2] \cap (d_1, d_2] \end{aligned}$$

and the analogous implication

$$d_2 \in (c_1, c_2] \cap (d_1, d_2] \Rightarrow \exists \varepsilon > 0 : d_2 - \varepsilon \in (c_1, c_2] \cap (d_1, d_2].$$

$\blacksquare$

Proposition 7.10 states that every pair $(a, b) \in R(\leftrightharpoons)$ is associated with a non-empty interval subset of the unit interval. We define

$$r_{(a,b)} = \lambda^1\big((F(x_{a-1}), F(x_a)) \cap (G(y_{b-1}), G(y_b))\big) \tag{7.6}$$

for all $(a, b) \in A \times B$ as the length of that interval. Note that, due to Proposition 7.10, $r_{(a,b)} > 0$ is equivalent to $a \leftrightharpoons b$.

Moreover, there exists a union N of at most countably many atoms in $(0, 1)$ such that

$$N \cup \bigcup_{(a,b) \in R(\leftrightharpoons)} (F(x_{a-1}), F(x_a)) \cap (G(y_{b-1}), G(y_b)) = (0, 1),$$

with all of the unions on the left hand side being disjoint. Specifically, $N = F(D_F) \cup G(D_G)$. For any two pairs $(a, b), (\alpha, \beta) \in R(\leftrightharpoons)$, we say that (α, β) is *higher (lower) than* (a, b) if $\gamma > c$ $(\gamma < c)$ holds for all $\gamma \in (F(x_{\alpha-1}), F(x_\alpha)) \cap (G(y_{\beta-1}), G(y_\beta))$ and all $c \in (F(x_{a-1}), F(x_a)) \cap (G(y_{b-1}), G(y_b))$. One of these two situations is guaranteed to hold since the interval associated with (a, b) and the interval associated with (α, β) are disjoint.

Example 7.11. a) We revisit Example 7.7a) in order to explore the relation $\leftrightharpoons$. The indexing sets A and B of F and G are equal to their supports. The identifying sequences are given by $(j, \frac{1}{2})_{j \in \{1,2\}}$ and $(j, \frac{1}{5})_{j \in \{1,\dots,5\}}$, respectively. We now go through the elements of A one by one, starting with $a = 1$, which yields

$$\begin{aligned} a \leftrightharpoons b &\Leftrightarrow (F(x_{a-1}), F(x_a)) \cap (G(y_{b-1}), G(y_b)) \neq \emptyset \\ &\Leftrightarrow (0, \tfrac{1}{2}) \cap (\tfrac{b-1}{5}, \tfrac{b}{5}) \neq \emptyset \\ &\Leftrightarrow b \in \{1, 2, 3\} \end{aligned}$$

for $b \in B$. Similarly, for $a = 2$, we obtain

$$\begin{aligned} a \leftrightharpoons b &\Leftrightarrow (\tfrac{1}{2}, 1) \cap (\tfrac{b-1}{5}, \tfrac{b}{5}) \neq \emptyset \\ &\Leftrightarrow b \in \{3, 4, 5\} \end{aligned}$$

for $b \in B$. Since $A = \{1, 2\}$, it follows from Proposition 7.10 that

$$R(\leftrightharpoons) = \{(1, 1), (1, 2), (1, 3), (2, 3), (2, 4), (2, 5)\}.$$

This is depicted using the cdf's in the left panel of Figure 7.1, which is very similar to the lower left panel of Figure 6.3. In fact, the same jumps have the same colour on them, the colouring is simply extended to the entire heights of the jumps in Figure 7.1. In Figure 6.3, the colours in the cdf's are associated with the Q-Q-plot in the panel above. This is representative of another characterization of the order $\leftrightharpoons$: $a \leftrightharpoons b$ holds, if and only if the point (x_a, y_b) is part of the corresponding Q-Q-plot. Both formulations mean that the the same piece of probability mass lies on the point x_a in F and on the point y_b in G.

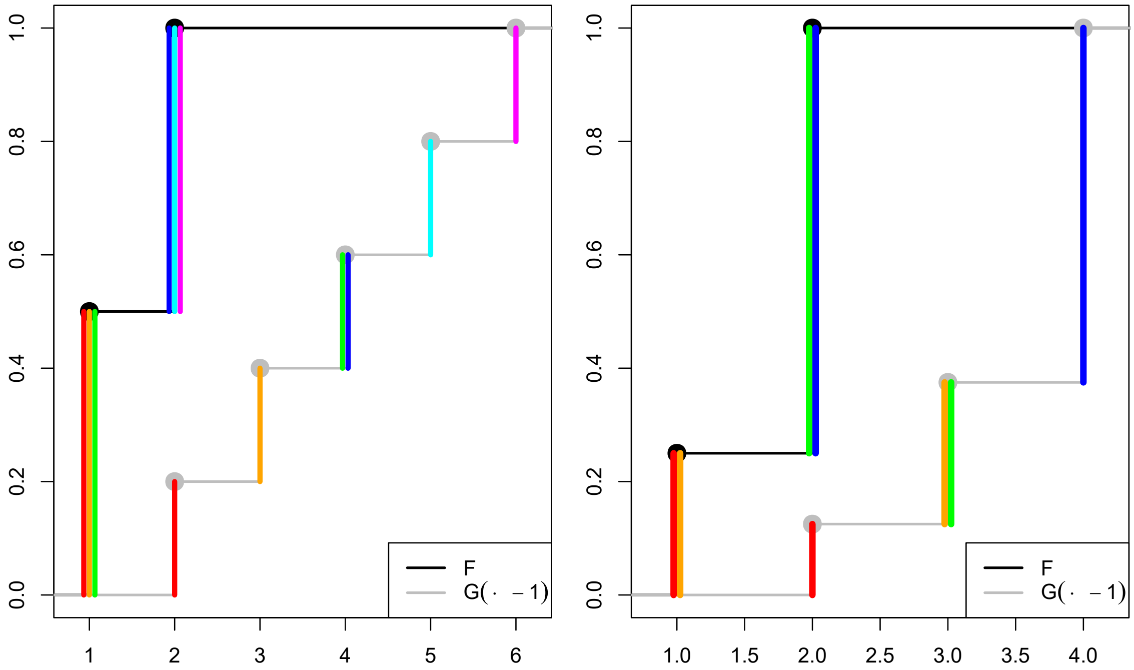

Figure 7.1.: Visualization of Example 7.11a) in the left panel and of Example 7.11b) in the right panel. The pairs of jumps, of which the heights are to be compared (and which are therefore connected by the relation $\leftrightharpoons$), are marked with the same colour.

b) Because of the simple structure of the cdf's in part a), they are not instructive for exploring the connection of the relation $\leftrightharpoons$ to the discrete dispersion order $\leq_{disp}^{disc}$ given in Proposition 7.9. Therefore, we also consider the following pair of cdf's. Let $X \sim F$ and $Y \sim G$ be defined by

$$\mathbb{P}(X = 1) = \tfrac{1}{4}, \quad \mathbb{P}(X = 2) = \tfrac{3}{4},$$
$$\mathbb{P}(Y = 1) = \tfrac{1}{8}, \quad \mathbb{P}(Y = 2) = \tfrac{1}{4}, \quad \mathbb{P}(Y = 3) = \tfrac{5}{8}.$$

Note that $F, G \in \mathcal{LD} \subseteq \mathcal{D}_0$ with the same defining distance 1; therefore, condition (7.3) is satisfied. Once again, the indexing sets A and B of F and G are given by their supports. As mentioned ahead of Example 7.11, the sets $(0,1) \setminus F(D_F)$ and $(0,1) \setminus G(D_G)$ are disjoint unions of intervals of the form $(F(x_{a-1}), F(x_a)), a \in A$, and $(G(y_{b-1}), G(y_b)), b \in B$, respectively. Specifically,

$$
\begin{aligned}
(0,1) \setminus F(D_F) &= (0, \tfrac{1}{4}) \cup (\tfrac{1}{4}, 1), \\
(0,1) \setminus G(D_G) &= (0, \tfrac{1}{8}) \cup (\tfrac{1}{8}, \tfrac{3}{8}) \cup (\tfrac{3}{8}, 1).
\end{aligned}
$$

By considering whether the pairwise intersections of these intervals are empty or not, we obtain

$$R(\leftrightharpoons) = \{(1,1),(1,2),(2,2),(2,3)\}$$

(see right panel of Figure 7.1). Going back to the definition of F and G, it is obvious that the first jump of F is at least as high as the first two jumps of G and the second jump of F is higher than the last two jumps of G. By Proposition 7.9, this yields $F \leq_{disp}^{disc} G$. Note that the third jump of G (height $\frac{5}{8}$) is higher than the first jump of F (height $\frac{1}{4}$), represented by the pair $(1,3) \in A \times B$ of indices. However, because the jumps do not overlap, $(1,3) \notin R(\leftrightharpoons)$, i.e. the comparison is not relevant to the discrete dispersion order.

Example 7.11b) shows that requirement (7.5) of the order $\leq_{disp}^{disc}$ compares the jump heights of the two involved distributions pointwise as opposed to uniformly. The original dispersion order $\leq_{disp}$ also compares in an pointwise manner as it just compares the gradient of the quantile functions at every point in the unit interval. The comparison of the gradient of F^{-1} at one point to the gradient of G^{-1} at another point is irrelevant. The meaning of pointwise comparisons is less obvious in a discrete setting, but as explained above, two jumps are to be compared if they overlap. As mentioned in Example 7.11a), this occurs, if and only if a point connecting the two jumps is part of the Q-Q-plot, which means that the two points share a common piece of probability mass among them. This is different from a uniform comparison since there are pairs of jumps, whose comparison is irrelevant, just like in the continuous setting.

With the very strict requirement (7.3), the discrete order $\leq_{disp}^{disc}$ seems to work fine. However, without this requirement, the order is not sufficient to capture all relevant aspects of dispersion. Specifically, the requirement (7.5) only looks at the distribution of the probability mass on the respective support, but not at the structure of that support, which explains the nature of the additional requirement (7.3). A simple example, in which (7.5) alone is not sufficient, is Example 7.5a), where $F, G \in \mathcal{E}_{nt}(n)$ for some $n \in \mathbb{N}_{\geq 2}$. There, the pmf's of both distributions are constantly equal on their respective supports and therefore, the difference in dispersion depends solely on the structure of the support.

Now, we want to obtain a weaker dispersion order than $\leq_{disp}^{disc}$ for arbitrary cdf's $F, G \in \mathcal{D}_0$ by maintaining the condition (7.5) for how the probability mass is distributed on the support and weakening the condition (7.3) for how the support is structured. The first condition, i.e.

$$g(G^{-1}(r)) \leq f(F^{-1}(r)) \quad \forall r \in (0,1),$$

can be applied to pairs of cdf's not satisfying condition (7.3). In the general setting, the condition is still well-defined and meaningful. While this is also true for condition (7.3), it is not pointwise, but uniform in nature. A pointwise requirement on the supports of $F, G \in \mathcal{D}_0$ is given in the equivalent characterization of $F \leq_{disp} G$ in Proposition 7.4 by

$$\lambda^1(F^{-1}(\{r\})) \leq \lambda^1(G^{-1}(\{r\})) \quad \forall r \in F(D_F). \tag{7.7}$$

This, however, is not a sensible condition, if the additional requirement $F(D_F) \subseteq G(D_G)$ is not satisfied. To this end, assume that there exists an $r_0 \in F(D_F) \setminus G(D_G)$ and that (7.7) holds. This yields

$$0 < \lambda^1(F^{-1}(\{r_0\})) \leq \lambda^1(G^{-1}(\{r_0\})) = 0,$$

a contradiction. Thus, a modification of condition (7.7) for arbitrary cdf's $F, G \in \mathcal{D}_0$ is needed. If the ranges of F and G do not satisfy $F(D_F) \subseteq G(D_G)$, it is reasonable to once again utilize the relation $\leftrightharpoons$ as an indicator for which comparisons are relevant. The subject of the comparison is (in (7.7) as well as in (7.3)) the distance, over which the cdf's equal one specific value. So the condition is given by

$$x_a - x_{a-1} \leq y_b - y_{b-1}$$

for whichever pairs $(a,b) \in A \times B$ of indices are to be compared, but generally not for all possible pairs. However, since we compare the constant intervals between two jumps and $\leftrightharpoons$ relates two jumps to each other, requiring the comparison for all $(a,b) \in R(\leftrightharpoons)$ results in some unreasonable asymmetries. This is exemplified and visualized in the following.

Example 7.12. a) Let $F, G \in \mathcal{D}_0$. Furthermore, let $A = \{1,2,3\}$, $B = \{1,2,3,4\}$ and let $p_j = \frac{1}{3}$ for all $j \in A$ and $q_j = \frac{1}{4}$ for all $j \in B$. The set $R(\leftrightharpoons)$ is given by

$$R(\leftrightharpoons) = \{(1,1),(1,2),(2,2),(2,3),(3,3),(3,4)\} = \bigcup_{a \in A} \{(a,a),(a,a+1)\}.$$

For the given cdf's, we have $\min A = \min B = 1$, implying $x_0 = y_0 = -\infty$. Since the quantities to be compared are $x_a - x_{a-1}$ and $y_b - y_{b-1}$, this leads to a problem in the case that either a or b are equal to 1. Then, both of the above quantities are infinite and graphically, we would not compare the length of a constant interval between to jumps but rather the length of the constant interval before all jumps. Since this is not a sensible procedure, we instead require

$$x_a - x_{a-1} \leq y_b - y_{b-1} \quad \forall (a,b) \in R(\leftrightharpoons) \cap ((A \setminus \{\min A\}) \times (B \setminus \{\min B\})). \tag{7.8}$$

The remaining pairs of indices to be compared are $(2, 2)$, $(2, 3)$, $(3, 3)$ and $(3, 4)$. Similarly to Example 7.11, the constant intervals to be compared are illustrated in the left panel of Figure 7.2 along with the cdf's. In this example, this seems to be a sensible choice.

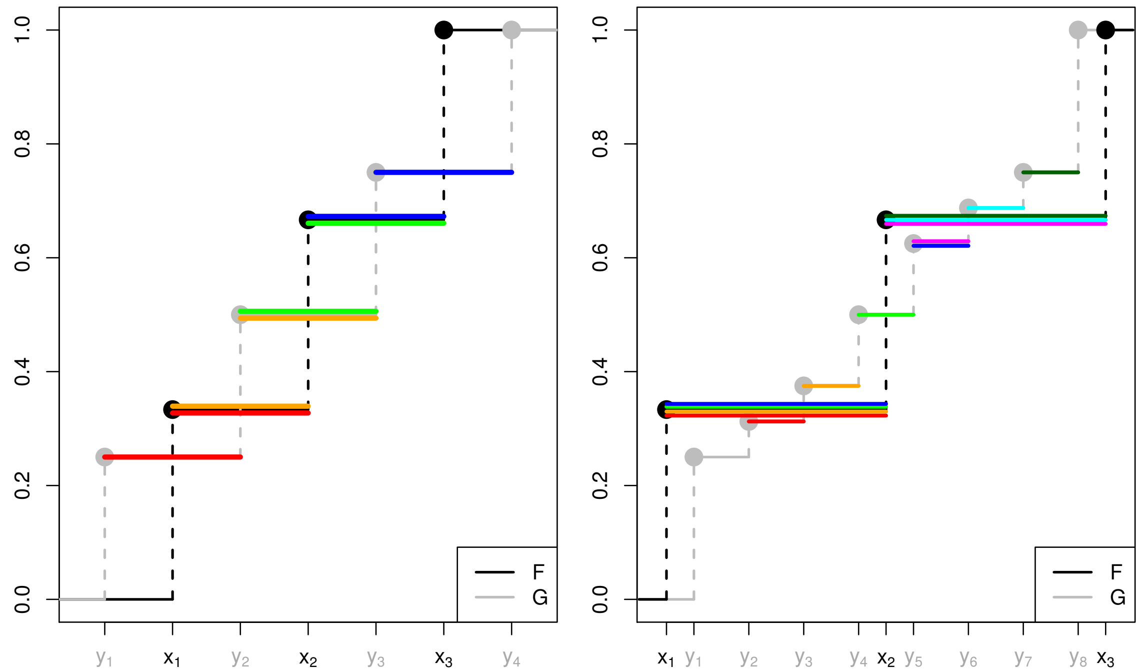

Figure 7.2.: Visualization of Example 7.12a) in the left panel and of Example 7.12b) in the right panel. The pairs of constant intervals, of which the lengths are deemed to be compared by $\leftrightharpoons$, are marked with the same colour.

b) Let F be defined as in part a) and let $G \in \mathcal{D}_0$ be defined by $B = \{1, \dots, 8\}$ and

$$(q_1, \dots, q_8)^\top = \frac{1}{16}(4, 1, 1, 2, 2, 1, 1, 4)^\top ,$$

yielding

$$F(D_F) = \left\{\frac{1}{3}, \frac{2}{3}\right\}, \quad G(D_G) = \left\{\frac{4}{16}, \frac{5}{16}, \frac{6}{16}, \frac{8}{16}, \frac{10}{16}, \frac{11}{16}, \frac{12}{16}\right\}.$$

Considering $\frac{5}{16} < \frac{1}{3} < \frac{6}{16}$ and $\frac{10}{16} < \frac{2}{3} < \frac{11}{16}$, we obtain

$$R(\leftrightharpoons) = \{(1, 1), (1, 2), (1, 3), (2, 3), (2, 4), (2, 5), (2, 6), (3, 6), (3, 7), (3, 8)\}.$$

Once again, we disregard the first three pairs for the comparisons of the constant intervals. The comparisons of the remaining pairs are illustrated in the right panel of Figure 7.2. Here, it becomes apparent that the relation $\leftrightharpoons$ is not fit to decide which pairs of constant intervals are to be compared with respect to their lengths. In spite of the obvious symmetry of both F and G, their comparison is highly asymmetric. This is evidenced by the pairs of indices $(2,2),(3,8) \in (A \setminus \{\min A\}) \times (B \setminus \{\min B\})$, which are symmetric, but satisfy $(3,8) \in R(\leftrightharpoons) \not\ni (2,2)$. An analogous statement is true for the pairs $(3,3),(2,5)$ and the pairs $(2,4),(3,4)$.

In Example 7.12b), the constant interval of F we identified with the index $a \in A \setminus \{\min A\}$ had the length $x_a - x_{a-1}$ and is the interval between the $(a-1)$-th and the a-th jump of the cdf. Therefore, it is just as reasonable to identify this interval with the index $a-1 \in A \setminus \{\max A\}$ instead of a. So, an alternative method of comparing the lengths of the pairs of constant intervals of F and G would be

$$x_{a+1} - x_a \leq y_{b+1} - y_b \quad \forall (a,b) \in R(\leftrightharpoons) \cap ((A \setminus \{\max A\}) \times (B \setminus \{\max B\})). \tag{7.9}$$

The resulting illustration for this type of comparison in Example 7.12a) is the same as in the left panel of Figure 7.2. However, the asymmetries in the resulting illustration for Example 7.12b) are reversed. This relationship is formalized and generalized by the following proposition.

Proposition 7.13. *Let $F, G \in \mathcal{D}_0$. Define the 'mirrored' cdf's $F^*(t) = 1 - F(-t)$, $G^*(t) = 1 - G(-t)$ for $t \in \mathbb{R}$ with $F^* \mathrel{\hat=} (A^*, (x_j^*, p_j^*)_{j \in A^*})$ and $G^* \mathrel{\hat=} (B^*, (y_j^*, q_j^*)_{j \in B^*})$. Furthermore, introduce the short hand $\leftrightharpoons^*$ for the relation $\overset{F^*,G^*}{\leftrightharpoons}$. Then, the following two statements are equivalent:*

(i) $x_a - x_{a-1} \leq y_b - y_{b-1} \quad \forall (a,b) \in R(\leftrightharpoons) \cap ((A \setminus \{\min A\}) \times (B \setminus \{\min B\}))$,

(ii) $x^*_{a^*+1} - x^*_{a^*} \leq y^*_{b^*+1} - y^*_{b^*} \ \forall (a^*, b^*) \in R(\leftrightharpoons^*) \cap ((A^* \setminus \{\max A^*\}) \times (B^* \setminus \{\max B^*\}))$.

Proof. The existence of $\min(\text{supp}(F))$ and therefore of $\min(A)$ is equivalent to the existence of $\max(A^*)$ with an analogous statement being true for $\max(A)$ and $\min(A^*)$. Examining the proof of Proposition 7.3 yields the following relationships between A and A^*:

$$A = A^* \text{ if } A \notin \{\mathbb{N}, -\mathbb{N}\}, \quad A = -A^* \text{ if } A \in \{\mathbb{N}, -\mathbb{N}\}.$$

Additionally, $\text{supp}(F)$ and $\text{supp}(F^*)$ contain the same elements, only with reversed sign. In the case $A = \{1, \ldots, n\}$ for some $n \in \mathbb{N}_{\geq 2}$, we define $a^* = n+1-a$, which yields $x_a = -x^*_{a^*}$ for any $a \in A = A^*$. Since $a > 1 = \min A$, it follows that $a^* = n+1-a < n = \max A = \max A^*$, and we obtain that, for any $a \in A \setminus \{\min A\}$, there exists an $a^* \in A^* \setminus \{\max A^*\}$ such that

$$x_a - x_{a-1} = x^*_{a^*+1} - x^*_{a^*}. \tag{7.10}$$

The cases $A = \mathbb{N}$ and $A = -\mathbb{N}$ also yield (7.10), if we define $a^* = -a \in A^* \setminus \{\max A^*\}$ for all $a \in A \setminus \{\min A\}$; the case $A = \mathbb{Z}$ yields (7.10), if we define $a^* = -a - 1 \in A^*$ for all $a \in A$. Analogous identities are true for G.

It remains to be shown that for all $(a, b) \in (A \setminus \{\min A\}) \times (B \setminus \{\min B\})$ and $(a^*, b^*) \in (A^* \setminus \{\max A^*\}) \times (B^* \setminus \{\max B^*\})$, the statement $a \leftrightharpoons b$ is equivalent to $a^* \leftrightharpoons^* b^*$. To this end, consider

$$\begin{aligned} a \leftrightharpoons b &\Leftrightarrow (F(x_{a-1}), F(x_a)) \cap (G(y_{b-1}), G(y_b)) \neq \emptyset \\ &\Leftrightarrow 1 - (F^*(-x_{a-1}), F^*(-x_a)) \cap 1 - (G^*(-y_{b-1}), G^*(-y_b)) \neq \emptyset \\ &\Leftrightarrow (F^*(-x_a), F^*(-x_{a-1})) \cap (G^*(-y_b), G^*(-y_{b-1})) \neq \emptyset \\ &\Leftrightarrow (F^*(x^*_{a^*}), F^*(x^*_{a^*+1})) \cap (G^*(y^*_{b^*}), G^*(y^*_{b^*+1})) \neq \emptyset \\ &\Leftrightarrow a^* \leftrightharpoons^* b^*, \end{aligned}$$

thus concluding the proof. ▨

Note that the proof of this proposition even specifies which pairs of indices are connected through the equivalence

$$x_a - x_{a-1} \leq y_b - y_{b-1} \Leftrightarrow x^*_{a^*+1} - x^*_{a^*} \leq y^*_{b^*+1} - y^*_{b^*}.$$

Hence, we know that the comparisons via (7.8) and via (7.9) both yield asymmetric results in a perfectly mirrored way. A possible strategy for obtaining a symmetric comparison is to combine the two methods in such a way that the asymmetries cancel out. There are two intuitive possibilities for this combination. They can either be combined via a logical and ($\wedge$) or via a logical or ($\vee$). In the following, both possibilities are explored. We start by defining two new relations that indicate the pairs of constant intervals to be compared for both proposed methods. For any indexing set $A \in \mathbb{I}$, we introduce the short hands $\underline{A} = A \setminus \{\min A\}$, $\overline{A} = A \setminus \{\max A\}$ and $\overline{\underline{A}} = A \setminus \{\min A, \max A\}$.

Definition 7.14. Let $F, G \in \mathcal{D}_0$.

a) The relation $\overset{F,G}{\leftrightharpoons}_\wedge$ on the set $\underline{A} \times \underline{B}$ is defined by

$$a \overset{F,G}{\leftrightharpoons}_\wedge b \Leftrightarrow (a \overset{F,G}{\leftrightharpoons} b) \wedge (a - 1 \overset{F,G}{\leftrightharpoons} b - 1)$$

for $a \in \underline{A}, b \in \underline{B}$. The set $R(\overset{F,G}{\leftrightharpoons}_\wedge)$ of all $(a, b) \in \underline{A} \times \underline{B}$ with $a \overset{F,G}{\leftrightharpoons}_\wedge b$ is said to be the set of (F, G)-$\wedge$-*dispersion-relevant* pairs of indices.

b) The relation $\overset{F,G}{\leftrightharpoons}_\vee$ on the set $\underline{A} \times \underline{B}$ is defined by

$$a \overset{F,G}{\leftrightharpoons}_\vee b \Leftrightarrow (a \overset{F,G}{\leftrightharpoons} b) \vee (a - 1 \overset{F,G}{\leftrightharpoons} b - 1)$$

for $a \in \underline{A}, b \in \underline{B}$. The set $R(\overset{F,G}{\leftrightharpoons}_\vee)$ of all $(a, b) \in \underline{A} \times \underline{B}$ with $a \overset{F,G}{\leftrightharpoons}_\vee b$ is said to be the set of (F, G)-$\vee$-*dispersion-relevant* pairs of indices.

Definition 7.15. Let $F, G \in \mathcal{D}_0$.

a) G is said to be *at least as discretely dispersed* as F *with respect to the probability mass*, denoted by $F \leq_{D-pm}^{disc} G$, if

$$q_b \leq p_a \quad \forall (a,b) \in R(\leftrightharpoons).$$

b) G is said to be *at least as* $\wedge$*-discretely dispersed* as F *with respect to the support*, denoted by $F \leq_{D-supp}^{\wedge-disc} G$, if

$$x_a - x_{a-1} \leq y_b - y_{b-1} \quad \forall (a,b) \in R(\leftrightharpoons_\wedge).$$

If $F \leq_{D-pm}^{disc} G$ and $F \leq_{D-supp}^{\wedge-disc} G$ hold, G is said to be *at least as* $\wedge$*-discretely dispersed* as F, denoted by $F \leq_{disp}^{\wedge-disc} G$.

c) G is said to be *at least as* $\vee$*-discretely dispersed* as F *with respect to the support*, denoted by $F \leq_{D-supp}^{\vee-disc} G$, if

$$x_a - x_{a-1} \leq y_b - y_{b-1} \quad \forall (a,b) \in R(\leftrightharpoons_\vee).$$

If $F \leq_{D-pm}^{disc} G$ and $F \leq_{D-supp}^{\vee-disc} G$ hold, G is said to be *at least as* $\vee$*-discretely dispersed* as F, denoted by $F \leq_{disp}^{\vee-disc} G$.

Obviously, the orders $\leq_{D-pm}^{disc}$ and $\leq_{D-supp}^{\wedge-disc}$ are a kind of split of the $\wedge$-discrete dispersive order, where the first ordering acts with respect to the y-axis or the probability mass and the second ordering with respect to the x-axis or the support of the distribution. The same split holds for the $\vee$-discrete dispersive order. Since $a \leftrightharpoons_\vee b$ implies $a \leftrightharpoons_\wedge b$ for all $(a,b) \in \underline{A} \times \underline{B}$, we have $R(\leftrightharpoons_\wedge) \subseteq R(\leftrightharpoons_\vee)$. By Definition 7.15, this yields

$$F \leq_{disp}^{\vee-disc} G \Longrightarrow F \leq_{disp}^{\wedge-disc} G$$

for all $F, G \in \mathcal{D}_0$, i.e. $\leq_{disp}^{\wedge-disc}$ is a weakening of $\leq_{disp}^{\vee-disc}$. Next, the performance of both orders is examined using the cdf's from Example 7.12.

Example 7.16 (Continuation of Example 7.12)**.** a) Consider the cdf's F and G as defined in Example 7.12a). Based on the set $R(\leftrightharpoons)$, it follows that

$$R(\leftrightharpoons_\wedge) = R(\leftrightharpoons_\vee) = \{(2,2),(2,3),(3,3),(3,4)\} = R(\leftrightharpoons) \cap (\underline{A} \times \underline{B}).$$

Hence, all four discussed discrete dispersion orders with respect to the support ($\leq_{D-supp}^{\wedge-disc}$, $\leq_{D-supp}^{\vee-disc}$, (7.8), and (7.9)) yield the same result for this simple example.

Since the pmf's of F and G are both constant on their supports, we obviously have $F \leq_{D-pm}^{disc} G$. Hence, the validity of $F \leq_{disp}^{\wedge-disc} G$ and $F \leq_{disp}^{\vee-disc} G$ depends solely on four conditions concerning the vectors (x_1, x_2, x_3) and (y_1, y_2, y_3, y_4), since $|R(\leftrightharpoons_\wedge)| =$

$|R(\leftrightharpoons_\vee)| = 4$. To be exact,

$$F \leq_{disp}^{\wedge-disc} G \Leftrightarrow F \leq_{disp}^{\vee-disc} G \Leftrightarrow x_2 - x_1 \leq y_2 - y_1,\ x_2 - x_1 \leq y_3 - y_2,$$
$$x_3 - x_2 \leq y_3 - y_2,\ x_3 - x_2 \leq y_4 - y_3.$$

b) Consider the cdf's F and G as defined in Example 7.12b). From the set $R(\leftrightharpoons)$ given in that example, we infer

$$R(\leftrightharpoons_\wedge) = \{(2,3),(2,4),(3,6),(3,7)\},$$
$$R(\leftrightharpoons_\vee) = \{(2,2),(2,3),(2,4),(3,4),(2,5),(3,5),(2,6),(3,6),(3,7),(3,8)\}.$$

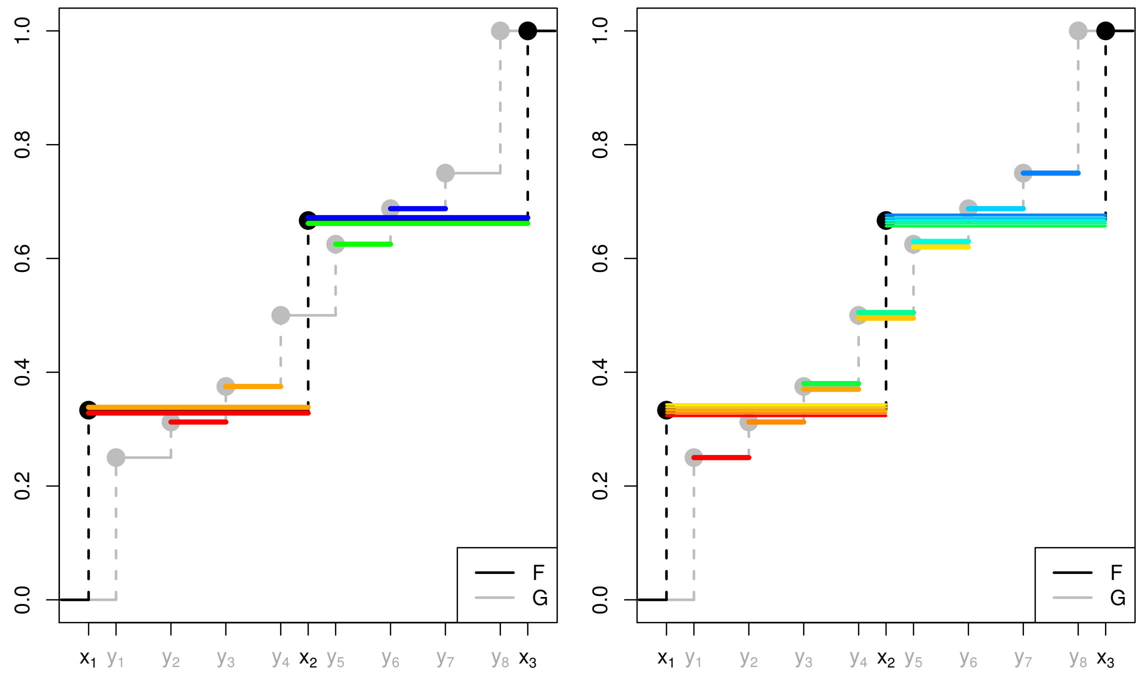

Figure 7.3.: Visualization of Example 7.16b). The pairs of constant intervals, of which the lengths are to be compared with respect to $\leftrightharpoons_\wedge$ (in the left panel) and $\leftrightharpoons_\vee$ (in the right panel), are marked with the same colour.

The difference in the number of comparisons dictated by $\leftrightharpoons_\wedge$ ($|R(\leftrightharpoons_\wedge)| = 4$) and $\leftrightharpoons_\vee$ ($|R(\leftrightharpoons_\vee)| = 10$) is quite large. Examining the structure of both sets suggests that there is a connection between $|R(\leftrightharpoons_\wedge)|$, i.e. the number of $\wedge$-comparisons, and $|\underline{A}|$, i.e. the cardinal number of the indexing set of F, the candidate for the less dispersed cdf. For each element in the latter set, there are two elements in the former set. As depicted in the left panel of Figure 7.3, the length of every constant interval of F is compared with the lengths of the two closest constant intervals of G, one from above and one from below. Since, in this example, $\underline{B}$ is much larger than $\underline{A}$, the constant intervals of G connected to the indices $2, 5, 8 \in \underline{B}$ are not used for any comparison. Hence, the statement $F \leq_{disp}^{\wedge-disc} G$ is completely independent from their lengths. One might say that the pointwise nature of the comparison via $\leq_{D-supp}^{\wedge-disc}$ is dictated by F.

A similar connection can be observed between $|R(\leftrightharpoons_{\vee})|$ and $|\underline{B}|$, i.e. the cardinal number of the indexing set of G, the candidate for the more dispersed cdf. Graphically, this means that every constant interval of G is compared with the lengths of the closest constant intervals of F from above and below, provided that the respective intervals exist. Hence, the pointwise comparison via $\leq_{D-supp}^{\vee-disc}$ seems to be dictated by G.

One might also use this example in combination with part a) to demonstrate that the order $\leq_{disp}^{\wedge-disc}$ already tends to be quite strong, similarly to the fact that $\leq_{disp}$ is a quite strong order in the continuous setting. To illustrate, let

$$\begin{aligned} c_x &= x_2 - x_1 = x_3 - x_2, \\ c_y &= y_3 - y_2 = y_4 - y_3 = y_6 - y_5 = y_7 - y_6 \\ \delta &= y_2 - y_1 = y_5 - y_4 = y_8 - y_7. \end{aligned}$$

Then, $F \leq_{disp}^{\wedge-disc} G$ is equivalent to $c_x \geq c_y$ with the value of δ being irrelevant. However, for $\delta \geq c_x$ and $c_y \to 0$, the example in part b) turns into the example in part a) with $F, G \in \mathcal{LD}$, c_x the defining distance of F and δ the defining distance of G. So in the limiting case $c_y \to 0$, we have $F \leq_{disp}^{\wedge-disc} G$, yet not if we approach that case with some $c_y \in (0, c_x)$. An approach to solve this inconsistency is presented in Section 8.1.

As noted in Example 7.16, one gets the impression that the pointwise comparisons of $\leq_{disp}^{\wedge-disc}$ are dictated by the candidate for the less dispersed cdf F and that the pointwise comparisons of $\leq_{disp}^{\vee-disc}$ are dictated by the candidate for the more dispersed cdf G. The following proposition formalizes and generalizes this idea. Preliminarily, we define the set of (upper and lower) nearest neighbours and prove a helpful lemma.

Definition 7.17. Let $F, G \in \mathcal{D}_0$ and let $a \in \underline{A}$. Then, the *set of (upper and lower) nearest neighbours of F in G with respect to a* (denoted by $\mathrm{NN}_F^G(a)$) is defined as follows.

(i) If $G(D_G) \cap (0, F(x_{a-1})] \neq \emptyset$ and $G(D_G) \cap [F(x_{a-1}), 1) \neq \emptyset$, define

$$\mathrm{NN}_F^G(a) = \{\sup\left(G(D_G) \cap (0, F(x_{a-1})]\right), \inf\left(G(D_G) \cap [F(x_{a-1}), 1)\right)\}.$$

(ii) If $G(D_G) \cap (0, F(x_{a-1})] = \emptyset$ and $G(D_G) \cap [F(x_{a-1}), 1) \neq \emptyset$, define

$$\mathrm{NN}_F^G(a) = \{\inf\left(G(D_G) \cap [F(x_{a-1}), 1)\right)\}.$$

(iii) If $G(D_G) \cap (0, F(x_{a-1})] \neq \emptyset$ and $G(D_G) \cap [F(x_{a-1}), 1) = \emptyset$, define

$$\mathrm{NN}_F^G(a) = \{\sup\left(G(D_G) \cap (0, F(x_{a-1})]\right)\}.$$

Here, it is impossible that both sets are empty since this would imply $\emptyset = G(D_G) \cap (0,1) = G(D_G)$ and thus $|\mathrm{supp}(G)| = 1$, which contradicts $G \in \mathcal{D}_0$. Furthermore, note that $F(x_{a-1})$ is the value that F takes on the interval $[x_{a-1}, x_a)$, which is generally associated with the index a.

Lemma 7.18. *Let $F, G \in \mathcal{D}_0$ satisfy $F \leq^{disc}_{D-pm} G$.*

a) Let $a \in A$. Then, exactly one of the following two statements holds:

(i) $\exists b \in \overline{B} : F(x_{a-1}) < G(y_b) < F(x_a)$,

(ii) $\exists b \in B : F(x_{a-1}) = G(y_{b-1})$ and $F(x_a) = G(y_b)$.

b) For all $a \in \overline{A}$, there exist $b_1, b_2 \in \overline{B}$ such that $G(y_{b_1}) \leq F(x_a) \leq G(y_{b_2})$ holds.

Proof. a) We prove the equivalence $\neg(i) \Leftrightarrow (ii)$. Note that $\neg(i)$ is equivalent to $G(D_G) \cap (F(x_{a-1}), F(x_a)) = \emptyset$.

'$\Leftarrow$': It follows that $(F(x_{a-1}), F(x_a)) \cap G(D_G) = (G(y_{b-1}), G(y_b)) \cap G(D_G) = \emptyset$.

'$\Rightarrow$': We start by proving that there exists a $b_u \in B$ such that $G(y_{b_u}) \geq F(x_a)$. If $F(x_a) < 1$, this follows directly from the fact that $\sup G(\mathrm{supp}(G)) = 1$. If $F(x_a) = 1$, $\sup G(\mathrm{supp}(G)) = 1$ also implies $\max G(\mathrm{supp}(G)) = 1$ since, otherwise, $\lim_{\beta\to\infty} G(y_\beta) - G(y_{\beta-1}) = 0 < 1 - F(x_{a-1}) = F(x_a) - F(x_{a-1})$ along with $\lim_{\beta\to\infty} G(y_\beta) = 1$ would contradict $F \leq^{disc}_{D-pm} G$. Obviously, there also exists a $b_\ell \in B \cup \{\min B - 1\}$ such that $G(y_{b_\ell}) \leq F(x_{a-1})$.

Now, define $b \in B$ by $G(y_b) = \min(G(\mathrm{supp}(G)) \cap [F(x_a), 1]) \leq G(y_{b_u})$, yielding $a \risingdotseq b$. By assumption $G(y_{b-1}) \leq F(x_{a-1}) < F(x_a) \leq G(y_b)$ holds and it follows that $q_b \geq p_a$. Equality holds, if and only if $G(y_{b-1}) = F(x_{a-1})$ and $G(y_b) = F(x_a)$, which corresponds to (ii). If equality does not hold, $F \leq^{disc}_{D-pm} G$ is contradicted.

b) Let $a \in \overline{A}$. The assertion follows by applying part a) to both a and $a+1 \in A$, and by considering all four arising cases separately.

▨

The following chain of inequalities follows directly from Lemma 7.18:

$$\inf G(D_G) \leq \inf F(D_F) \leq \sup F(D_F) \leq \sup G(D_G).$$

Proposition 7.19. *Let $F, G \in \mathcal{D}_0$ satisfy $F \leq^{disc}_{D-pm} G$. Then,*

a) $R(\risingdotseq_\wedge) = \bigcup_{a\in\underline{A}} \left(\{a\} \times \{\beta \in \underline{B} : G(y_{\beta-1}) \in \mathrm{NN}_F^G(a)\}\right)$, *b)* $R(\leftrightharpoons_\vee) = \bigcup_{b\in\underline{B}}\Big(\{\alpha \in \underline{A} : F(x_{\alpha-1}) \in \mathrm{NN}_G^F(b)\} \times \{b\}\Big)$.

Proof. a) First, note that, for all $a \in \underline{A}$, it follows from Lemma 7.18b) that $G(D_G) \cap (0, F(x_{a-1})] \neq \emptyset$ and $G(D_G) \cap [F(x_{a-1}), 1) \neq \emptyset$. Thus,

$$\mathrm{NN}_F^G(a) = \{\sup\left(G(D_G) \cap (0, F(x_{a-1})]\right), \inf\left(G(D_G) \cap [F(x_{a-1}), 1)\right)\}.$$

Because of $\sup G(\mathrm{supp}(G)) = 1$ and $F(x_{a-1}) < 1$, there exists a $b \in B$ such that $F(x_{a-1}) \in [G(y_{b-1}), G(y_b))$. It follows that

$$\sup(G(D_G) \cap (0, F(x_{a-1})]) = \max(G(D_G) \cap (0, F(x_{a-1})])$$

and, analogously,

$$\inf(G(D_G) \cap [F(x_{a-1}), 1)) = \min(G(D_G) \cap [F(x_{a-1}), 1)).$$

'$\supseteq$': Let $a \in \underline{A}$. Now $b-1 \in \overline{B}$ (or $b \in \underline{B}$) is defined uniquely by $G(y_{b-1}) = \max(G(D_G) \cap (0, F(x_{a-1})])$. If follows that $G(y_{b-1}) \leq F(x_{a-1}) < G(y_b)$ and thus, $a \leftrightharpoons b$. The case $G(y_{b-1}) = F(x_{a-1})$ is equivalent to

$$\min(G(D_G) \cap [F(x_{a-1}), 1)) = \max(G(D_G) \cap (0, F(x_{a-1})]) = F(x_{a-1}).$$

Then, there exists an $\varepsilon > 0$ such that

$$(F(x_{a-2}), F(x_{a-1})) \ni F(x_{a-1}) - \varepsilon = G(y_{b-1}) - \varepsilon \in (G(y_{b-2}), G(y_{b-1})), \quad (7.11)$$

$$(F(x_{a-1}), F(x_a)) \ni F(x_{a-1}) + \varepsilon = G(y_{b-1}) + \varepsilon \in (G(y_{b-1}), G(y_b)), \quad (7.12)$$

thus yielding $a \leftrightharpoons_\wedge b$. The case $G(y_{b-1}) < F(x_{a-1})$ remains to be considered. It immediately follows that $a-1 \leftrightharpoons b$. This, combined with $a \leftrightharpoons b$, yields

$$\begin{aligned} &G(y_b) - G(y_{b-1}) \leq F(x_{a-1}) - F(x_{a-2}),\ G(y_b) - G(y_{b-1}) \leq F(x_a) - F(x_{a-1}) \\ \Longrightarrow\ & F(x_{a-2}) < G(y_{b-1}) < F(x_{a-1}) < G(y_b) < F(x_a). \end{aligned}$$

It follows that $a-1 \leftrightharpoons b-1$ and $a \leftrightharpoons b+1$, thus, $a \leftrightharpoons_\wedge b$ and $a \leftrightharpoons_\wedge b+1$. Since $G(y_{(b+1)-1}) = G(y_b) = \min(G(D_G) \cap [F(x_{a-1}), 1))$, this concludes the proof of the implication from right to left.

'$\subseteq$': Let $(a, b) \in R(\leftrightharpoons_\wedge) \subseteq \underline{A} \times \underline{B}$.

Case 1: $G(y_{b-1}) > F(x_{a-1})$

Under this assumption, Lemma 7.18 states that $b-1 > \min B$. From $a-1 \leftrightharpoons b-1$, we then obtain $G(y_{b-2}) < F(x_{a-1})$. Consequently, $G(y_{b-1}) = \min(G(D_G) \cap [F(x_{a-1}), 1)) \in \mathrm{NN}_F^G(a)$.

Case 2: $G(y_{b-1}) < F(x_{a-1})$

Similarly to Case 1, we have $a \leftrightharpoons b$, which yields $G(y_b) > F(x_{a-1})$ and $G(y_{b-1}) = \max(G(D_G) \cap (0, F(x_{a-1})]) \in \mathrm{NN}_F^G(a)$.

Case 3: $G(y_{b-1}) = F(x_{a-1})$

It immediately follows that

$$\begin{aligned} G(y_{b-1}) &= \min(G(D_G) \cap [F(x_{a-1}), 1)) \\ &= \max(G(D_G) \cap (0, F(x_{a-1})]) \in \mathrm{NN}_F^G(a). \end{aligned}$$

b) '$\supseteq$': Let $b \in \underline{B}$. Assume first $F(D_F) \cap [G(y_{b-1}), 1) \neq \emptyset$. Then, with analogous reasoning to part a), there exists an $a \in \underline{A}$ such that

$$F(x_{a-1}) = \inf(F(D_F) \cap [G(y_{b-1}), 1)) = \min(F(D_F) \cap [G(y_{b-1}), 1)).$$

It follows that $F(x_{a-1}) \geq G(y_{b-1}) > F(x_{a-2})$, where it is possible that $a - 1 = \min A$ and therefore, $F(x_{a-2}) = 0$. Nonetheless, we obtain $a - 1 \leftrightharpoons b - 1$ and thus, $a \leftrightharpoons_\vee b$.

Now we assume $F(D_F) \cap (0, G(y_{b-1})] \neq \emptyset$ (which can occur simultaneously to $F(D_F) \cap [G(y_{b-1}), 1) \neq \emptyset$). Again analogously to part a), there exists an $a \in \underline{A}$ such that

$$F(x_{a-1}) = \sup(F(D_F) \cap (0, G(y_{b-1})]) = \max(F(D_F) \cap (0, G(y_{b-1})]).$$

We now infer $F(x_{a-1}) \leq G(y_{b-1}) < F(x_a)$, yielding $a \leftrightharpoons b$ and thereby $a \leftrightharpoons_\vee b$.

'$\subseteq$': Let $(a, b) \in R(\leftrightharpoons_\vee) \subseteq \underline{A} \times \underline{B}$.

Case 1: $F(x_{a-1}) < G(y_{b-1})$

If $a - 1 \leftrightharpoons b - 1$, it follows $G(y_{b-2}) < F(x_{a-1})$, which then yields $a \leftrightharpoons b - 1$. We obtain

$$F(x_a) = F(x_{a-1}) + p_a \geq F(x_{a-1}) + q_{b-1} > G(y_{b-2}) + q_{b-1} = G(y_{b-1}).$$

If $a \leftrightharpoons b$, it follows directly that $F(x_a) > G(y_{b-1})$. Since $a \leftrightharpoons_\vee b$ implies $a \leftrightharpoons b$ or $a - 1 \leftrightharpoons b - 1$, the inequality $F(x_a) > G(y_{b-1})$ holds generally. It yields $F(x_{a-1}) = \max(F(D_F) \cap (0, G(y_{b-1})]) \in \mathrm{NN}_G^F(b)$.

Case 2: $F(x_{a-1}) > G(y_{b-1})$

If $a \leftrightharpoons b$, it follows $G(y_b) > F(x_{a-1})$, which then yields $a - 1 \leftrightharpoons b$. We obtain

$$F(x_{a-2}) = F(x_{a-1}) - p_{a-1} \leq F(x_{a-1}) - q_b < G(y_b) - q_b = G(y_{b-1}).$$

If $a-1 \fallingdotseq b-1$, it follows directly that $F(x_{a-2}) < G(y_{b-1})$. Thus, the implication $F(x_{a-1}) = \min(F(D_F) \cap [G(y_{b-1}), 1)) \in \mathrm{NN}_G^F(b)$ generally holds.

Case 3: $F(x_{a-1}) = G(y_{b-1})$

Similarly to part a), it immediately follows that

$$\begin{aligned} F(x_{a-1}) &= \min(F(D_F) \cap [G(y_{b-1}), 1)) \\ &= \max(F(D_F) \cap (0, G(y_{b-1})]) \in \mathrm{NN}_G^F(b). \end{aligned}$$

■

Since Proposition 7.19 provides a new way of determining which pairs of indices are to be compared, it implicitly also provides new characterizations of the orders $\leq_{D-supp}^{\wedge-disc}$ and $\leq_{D-supp}^{\vee-disc}$. Therefore, let $F, G \in \mathcal{D}_0$ satisfy $F \leq_{D-pm}^{disc} G$. For $\leq_{D-supp}^{\wedge-disc}$, this new characterization is then given by

$$\begin{aligned} F \leq_{D-supp}^{\wedge-disc} G &\iff x_a - x_{a-1} \leq y_b - y_{b-1} \quad \forall (a,b) \in R(\fallingdotseq_\wedge) \\ &\iff \lambda^1(F^{-1}(\{F(x_{a-1})\})) \leq \lambda^1(G^{-1}(\{G(y_{b-1})\})) \quad \forall (a,b) \in R(\fallingdotseq_\wedge) \\ &\iff \lambda^1(F^{-1}(\{F(x_{a-1})\})) \leq \lambda^1(G^{-1}(\{q\})) \quad \forall q \in \mathrm{NN}_F^G(a)\ \forall a \in \underline{A} \\ &\iff \lambda^1(F^{-1}(\{p\})) \leq \lambda^1(G^{-1}(\{\sup(G(D_G) \cap (0,p])\})) \quad \text{and} \\ &\qquad \lambda^1(F^{-1}(\{p\})) \leq \lambda^1(G^{-1}(\{\inf(G(D_G) \cap [p,1))\})) \quad \forall p \in F(D_F). \end{aligned} \tag{7.13}$$

The third equivalence is true because of Proposition 7.19a). The last equivalence is true because of the first preliminary of the proof of said proposition, which states that

$$\mathrm{NN}_F^G(a) = \{\sup\left(G(D_G) \cap (0, F(x_{a-1})]\right), \inf\left(G(D_G) \cap [F(x_{a-1}), 1)\right)\}$$

holds for all $a \in \underline{A}$, if $F \leq_{D-pm}^{disc} G$. The new characterization of $\leq_{D-supp}^{\vee-disc}$ is similarly given by

$$\begin{aligned} F \leq_{D-supp}^{\vee-disc} G &\iff x_a - x_{a-1} \leq y_b - y_{b-1} \quad \forall (a,b) \in R(\fallingdotseq_\vee) \\ &\iff \lambda^1(F^{-1}(\{F(x_{a-1})\})) \leq \lambda^1(G^{-1}(\{G(y_{b-1})\})) \quad \forall (a,b) \in R(\fallingdotseq_\vee) \\ &\iff \lambda^1(F^{-1}(\{p\})) \leq \lambda^1(G^{-1}(\{G(y_{b-1})\})) \quad \forall p \in \mathrm{NN}_G^F(b)\ \forall b \in \underline{B} \\ &\iff \forall q \in G(D_G): \\ &\lambda^1(F^{-1}(\{\sup(F(D_F) \cap (0,q])\})) \leq \lambda^1(G^{-1}(\{q\})), \text{ if } F(D_F) \cap (0,q] \neq \emptyset, \quad \text{and} \\ &\lambda^1(F^{-1}(\{\inf(F(D_F) \cap [p,1))\})) \leq \lambda^1(G^{-1}(\{q\})), \text{ if } F(D_F) \cap [q,1) \neq \emptyset. \end{aligned} \tag{7.14}$$

These equivalent characterizations (7.13) and (7.14) further strengthen the heuristic idea that $\leq_{D-supp}^{\wedge-disc}$ (and thereby $\leq_{disp}^{\wedge-disc}$) employs a pairwise comparison with respect to F whereas $\leq_{D-supp}^{\vee-disc}$ (and thereby $\leq_{disp}^{\vee-disc}$) employs a pairwise comparison with respect to G. Since Lemma 7.18 states that $G(D_G)$ partitions the interval $(0,1)$ in a finer way than $F(D_F)$, this also provides a heuristic explanation of why $\leq_{D-supp}^{\wedge-disc}$ is a weakening of $\leq_{D-supp}^{\vee-disc}$.

We now come back to Example 7.16 in order to further explain the meaning of Proposition 7.19 and to show that it coincides with our hypothesis based on the example.

Example 7.20 (Continuation of Example 7.16). a) Consider Example 7.16a) with $A = \{1, 2, 3\}$ and $B = \{1, 2, 3, 4\}$. The sets on the right sides of the identities in Proposition 7.19 are given by

$$\bigcup_{a \in \underline{A}} \Big(\{a\} \times \{\beta \in \underline{B} : G(y_{\beta-1}) \in \mathrm{NN}_F^G(a)\} \Big) = \{(2,2),(2,3)\} \cup \{(3,3),(3,4)\},$$
$$\bigcup_{b \in \underline{B}} \Big(\{\alpha \in \underline{A} : F(x_{\alpha-1}) \in \mathrm{NN}_G^F(b)\} \times \{b\} \Big) = \{(2,2)\} \cup \{(2,3),(3,3)\} \cup \{(3,4)\}.$$

Note that the two sets are equal (as expected by combining the results of Example 7.16a) and Proposition 7.19) in spite of the differing cardinal numbers $|\underline{A}| = 2$ and $|\underline{B}| = 3$. This can be explained as follows: because of $F \leq_{D-pm}^{disc} G$, the proof of Proposition 7.19a) states that $|\mathrm{NN}_F^G(a)| = 2$ holds for all $a \in \underline{A}$ with $F(x_{a-1}) \notin G(D_G)$. However, this is not true if the roles of F and G are reversed. In our specific example, $|\mathrm{NN}_G^F(2)| = |\mathrm{NN}_G^F(4)| = 1$ since the constant intervals of G with indices $2, 4 \in \underline{B}$ are smaller / larger than every element of $F(D_F)$.

b) In Example 7.16b), we have $A = \{1, 2, 3\}$ and $B = \{1, \ldots, 8\}$ and the sets from Proposition 7.19 are given by

$$\bigcup_{a \in \underline{A}} \Big(\{a\} \times \{\beta \in \underline{B} : G(y_{\beta-1}) \in \mathrm{NN}_F^G(a)\} \Big) = \{(2,3),(2,4)\} \cup \{(3,6),(3,7)\},$$
$$\bigcup_{b \in \underline{B}} \Big(\{\alpha \in \underline{A} : F(x_{\alpha-1}) \in \mathrm{NN}_G^F(b)\} \times \{b\} \Big) = \{(2,2)\} \cup \{(2,3)\}$$
$$\cup \bigcup_{b \in \{4,5,6\}} \{(2,b),(3,b)\} \cup \{(3,7)\} \cup \{(3,8)\}.$$

As postulated by Proposition 7.19, the first set is equal to $R(\leftrightharpoons_\wedge)$ and the second one is equal to $R(\leftrightharpoons_\vee)$. Since the sets $F(D_F)$ and $G(D_G)$ do not share any elements, we have $|\mathrm{NN}_F^G(2)| = |\mathrm{NN}_F^G(3)| = 2$. For the set $R(\leftrightharpoons_\vee)$, we have $G(D_G) \cap (0, \min F(D_F)] = \{\frac{1}{4}, \frac{5}{16}\} = \{G(y_1), G(y_2)\}$. Therefore, $F(D_F) \cap (0, G(y_1)] = \emptyset = F(D_F) \cap (0, G(y_2)]$ and $|\mathrm{NN}_G^F(2)| = |\mathrm{NN}_G^F(3)| = 1$. This means that $\leq_{D-supp}^{\vee-disc}$ compares the first and second constant interval of G (connected to the indices $2, 3 \in \underline{B}$) each to one constant interval of F and not two. Similarly, because of $G(D_G) \cap [\max F(D_F), 1) = \{\frac{11}{16}, \frac{3}{4}\} = \{G(y_6), G(y_7)\}$, the constant intervals of G connected to the indices $7, 8 \in \underline{B}$ are also each compared to only one constant interval of F by $\leq_{D-supp}^{\vee-disc}$.

7.2. First Properties of the Discrete Dispersive Orders

It would be desirable for a discrete generalization of the dispersive order to satisfy two main properties. First, it should bridge the gap between the dispersive order $\leq_{disp}$ in the discrete setting (so for $F(D_F) \subseteq G(D_G)$) and the discrete dispersive order $\leq_{disp}^{disc}$ defined in Definition 7.6. Ideally, this means that the new discrete order is equivalent to $\leq_{disp}$ and $\leq_{disp}^{disc}$ on their respective areas of applicability. The second desirable main property is that a discrete dispersive order should be reflexive and transitive. We start by proving the following result concerning the first property.

Theorem 7.21. *Let $F, G \in \mathcal{D}_0$.*

a) If condition (7.3) is satisfied, then

$$F \leq_{disp}^{disc} G \Longleftrightarrow F \leq_{disp}^{\wedge-disc} G \Longleftrightarrow F \leq_{disp}^{\vee-disc} G.$$

b) If $F(D_F) \subseteq G(D_G)$, then

$$F \leq_{disp}^{\vee-disc} G \Longrightarrow F \leq_{disp}^{\wedge-disc} G \Longleftrightarrow F \leq_{disp} G.$$

Proof. a) Proposition 7.9 states that $F \leq_{disp}^{disc} G$ is equivalent to $F \leq_{D-pm}^{disc} G$ and (7.3). Therefore, by Definition 7.15, it only remains to be shown that condition (7.3) already implies $F \leq_{D-supp}^{\vee-disc} G$, which, in turn, implies $F \leq_{D-supp}^{\wedge-disc} G$. The central inequality to be satisfied is the same in all three conditions and is given by $x_a - x_{a-1} \leq y_b - y_{b-1}$. The only difference lies in the required supports of that inequality, i.e. the pairs of indices (a, b), for which it is required to hold. This support is given by $\underline{A} \times \underline{B}$ for condition (7.3), by $R(\leftrightharpoons_\vee)$ for $F \leq_{D-supp}^{\vee-disc} G$ and by $R(\leftrightharpoons_\wedge)$ for $F \leq_{D-supp}^{\wedge-disc} G$. The assertion now follows from the chain of inclusions $R(\leftrightharpoons_\wedge) \subseteq R(\leftrightharpoons_\vee) \subseteq \underline{A} \times \underline{B}$.

b) Since the implication holds in a more general setting, only the equivalence must be proven.

We start out by proving that $F(D_F) \subseteq G(D_G)$ already implies $F \leq_{D-pm}^{disc} G$. To this end, let $(a, b) \in R(\leftrightharpoons)$, i.e.

$$(F(x_{a-1}), F(x_a)) \cap (G(y_{b-1}), G(y_b)) \neq \emptyset. \tag{7.15}$$

By assumption, $G(y_b) > F(x_a)$ implies $F(x_a) = G(y_{b-1})$ and $G(y_{b-1}) < F(x_{a-1})$ implies $F(x_{a-1}) = G(y_b)$, so both cases contradict (7.15). This yields

$$(G(y_{b-1}), G(y_b)) \subseteq (F(x_{a-1}), F(x_a)). \tag{7.16}$$

Again by assumption, there exist $b_u \in B, b_\ell \in B \cup \{-\infty\}, b_\ell < b_u$ such that $F(x_a) =$

$G(y_{b_u})$ and $F(x_{a-1}) = G(y_{b_\ell})$. By combining this with (7.16), we obtain $b, b-1 \in \{b_\ell, \dots, b_u\}$ or, equivalently, $b \in \{b_\ell + 1, \dots, b_u\}$. It follows

$$p_a = F(x_a) - F(x_{a-1}) = G(y_{b_u}) - G(y_{b_\ell}) = \sum_{j=b_\ell+1}^{b_u} q_j \geq q_b,$$

thus proving $F \leq_{D-pm}^{disc} G$.

It remains to be shown that $F \leq_{disp} G$ is equivalent to $F \leq_{D-supp}^{\wedge-disc} G$. Note that $F(D_F) = F(\mathrm{supp}(F)) \setminus \{1\} = F(\mathrm{supp}(F) \setminus \{\max(\mathrm{supp}(F))\})$ holds as well as the analogous identity for G. The following equivalences hold:

$$\begin{aligned}
F \leq_{disp} G &\Leftrightarrow \lambda^1(F^{-1}(\{r\})) \leq \lambda^1(G^{-1}(\{r\})) \quad \forall r \in F(D_F) \\
&\Leftrightarrow \lambda^1(F^{-1}(\{F(x)\})) \leq \lambda^1(G^{-1}(\{G(y)\})) \\
&\qquad\qquad\qquad\qquad \forall x \in \mathrm{supp}(F) \setminus \{\max(\mathrm{supp}(F))\}, \\
&\qquad\qquad\qquad\qquad y \in \mathrm{supp}(G) \setminus \{\max(\mathrm{supp}(G))\} : F(x) = G(y) \\
&\Leftrightarrow \lambda^1(F^{-1}(\{F(x_a)\})) \leq \lambda^1(G^{-1}(\{G(y_b)\})) \ \forall a \in \overline{A}, b \in \overline{B} : F(x_a) = G(y_b) \\
&\Leftrightarrow \lambda^1([x_a, x_{a+1})) \leq \lambda^1([y_b, y_{b+1})) \qquad\qquad \forall a \in \overline{A}, b \in \overline{B} : F(x_a) = G(y_b) \\
&\Leftrightarrow x_a - x_{a-1} \leq y_b - y_{b-1} \qquad\qquad\qquad \forall a \in \underline{A}, b \in \underline{B} : F(x_{a-1}) = G(y_{b-1}),
\end{aligned}$$

where the first equivalence holds because of Proposition 7.4. Comparing the last equivalent characterization with the definition of $F \leq_{D-supp}^{\wedge-disc} G$ yields, that only the equivalence of $F(x_{a-1}) = G(y_{b-1})$ and $a \leftrightharpoons_\wedge b$ is left to prove for $a \in \underline{A}, b \in \underline{B}$. To this end, we use Proposition 7.19a) to obtain that, since $a \in \underline{A}$, $a \leftrightharpoons_\wedge b$ is equivalent to

$$\begin{aligned}
G(y_{b-1}) \in \mathrm{NN}_F^G(a) &= \{\sup\left(G(D_G) \cap (0, F(x_{a-1})]\right), \inf\left(G(D_G) \cap [F(x_{a-1}), 1)\right)\} \\
&= \{F(x_{a-1})\}.
\end{aligned}$$

The equivalence in Theorem 7.21a) is particularly relevant for the class $\mathcal{LD}$ of all lattice distribution. As discussed in Example 7.7b), the comparison of the supports boils down to a comparison of the defining distances in this case. Therefore, the comparison of two lattice distributions is greatly simplified since, essentially, the involved cdf's only need to be compared with respect to $\leq_{D-pm}^{disc}$. This observation is formalized in Corollary 7.23a).

Note that the implication in Theorem 7.21b) is strict, i.e. that the reverse implication $F \leq_{disp}^{\wedge-disc} G \Rightarrow F \leq_{disp}^{\vee-disc} G$ does not hold in general under the assumption $F(D_F) \subseteq G(D_G)$. For this, consider $X \sim \mathrm{Bin}(1, \frac{1}{2})$ and $\mathbb{P}(Y = 0) = \frac{1}{2}$, $\mathbb{P}(Y = 1) = \frac{1}{4}$ and $\mathbb{P}(Y = \frac{3}{2}) = \frac{1}{4}$, which yields $F, G \in \mathcal{D}_0$. Then we have $F(D_F) = \{\frac{1}{2}\} \subseteq \{\frac{1}{2}, \frac{3}{4}\} = G(D_G)$, which, as noted in the proof of Theorem 7.21b), implies $F \leq_{D-pm}^{disc} G$. The reduced indexing sets are given by

$\underline{A} = \{2\}$ and $\underline{B} = \{2, 3\}$. The corresponding sets of nearest neighbours are

$$\mathrm{NN}_F^G(2) = \left\{\sup\left(G(D_G) \cap (0, \tfrac{1}{2}]\right), \inf\left(G(D_G) \cap [\tfrac{1}{2}, 1)\right)\right\} = \{\tfrac{1}{2}\} = \{G(y_1)\},$$
$$\mathrm{NN}_G^F(2) = \left\{\sup\left(F(D_F) \cap (0, \tfrac{1}{2}]\right), \inf\left(F(D_F) \cap [\tfrac{1}{2}, 1)\right)\right\} = \{\tfrac{1}{2}\} = \{F(x_1)\},$$
$$\mathrm{NN}_G^F(3) = \left\{\sup\left(F(D_F) \cap (0, \tfrac{3}{4}]\right)\right\} = \{\tfrac{1}{2}\} = \{F(x_1)\},$$

where the structure of the set $\mathrm{NN}_G^F(3)$ is due to the fact that $F(D_F) \cap [\frac{3}{4}, 1) = \emptyset$. It follows that $R(\leftrightharpoons_\wedge) = \{(2, 2)\}$ and $R(\leftrightharpoons_\vee) = \{(2, 2), (2, 3)\}$. For the first element of both sets, we obtain $x_2 - x_1 = 1 = y_2 - y_1$, already implying $F \leq_{disp}^{\wedge-disc} G$. However, for the second element of $R(\leftrightharpoons_\vee)$, $x_2 - x_1 = 1 > \frac{1}{2} = y_3 - y_2$ holds, yielding $F \not\leq_{disp}^{\vee-disc} G$.

Generally, a counterexample can be constructed out of any pair of distributions $F, G \in \mathcal{D}_0$ with $F(D_F) \subseteq G(D_G)$ but $F(D_F) \neq G(D_G)$ by choosing the supports of F and G accordingly. This is due to the fact that the constant intervals of G with values in $G(D_G) \setminus F(D_F)$ are used for comparisons by $\leq_{D\text{-}supp}^{\vee-disc}$, but not by $\leq_{D\text{-}supp}^{\wedge-disc}$.

The next result concerns itself with the second desirable main property of a discrete generalization of the dispersive order, i.e. that it is transitive. It is also shown that both discrete orders are reflexive.

Theorem 7.22. *Let $F, G, H \in \mathcal{D}_0$.*

a) The orders $\leq_{disp}^{\wedge-disc}$ and $\leq_{disp}^{\vee-disc}$ are both reflexive, i.e. $F \leq_{disp}^{\wedge-disc} F$ and $F \leq_{disp}^{\vee-disc} F$.

b) The order $\leq_{disp}^{\vee-disc}$ is transitive, i.e. $F \leq_{disp}^{\vee-disc} G$ and $G \leq_{disp}^{\vee-disc} H$ implies $F \leq_{disp}^{\vee-disc} H$. However, in general, the order $\leq_{disp}^{\wedge-disc}$ is not transitive.

Proof. a) For $a, b \in A$, we obviously have $a \leftrightharpoons b$, if and only if $a = b$. Consequently, for all $a, b \in \underline{A}$, we have

$$a \leftrightharpoons_\wedge b \iff a = b \text{ and } a - 1 = b - 1 \iff a = b,$$
$$a \leftrightharpoons_\vee b \iff a = b \;\text{ or }\; a - 1 = b - 1 \iff a = b.$$

Therefore, $F \leq_{D-pm}^{disc} F$ is equivalent to $p_a \leq p_a$ for all $a \in A$; and $F \leq_{D-supp}^{\wedge-disc} F$ is equivalent to $F \leq_{D-supp}^{\vee-disc} F$, which, in turn, is equivalent to $x_a - x_{a-1} \leq x_a - x_{a-1}$ for all $a \in \underline{A}$. Hence, $F \leq_{disp}^{\wedge-disc} F$ and $F \leq_{disp}^{\vee-disc} F$ both hold.

b) Let $H \mathrel{\hat{=}} (C, (z_j, r_j)_{j \in C})$. First, let $(a, c) \in R(\overset{F,H}{\leftrightharpoons})$, i.e.

$$(F(x_{a-1}), F(x_a)) \cap (H(z_{c-1}), H(z_c)) \neq \emptyset.$$

Furthermore, note that $(\bigcup_{b \in B}(G(y_{b-1}), G(y_b))) \cup G(D_G) = (0, 1)$ holds and the union is disjoint. Since $(F(x_{a-1}), F(x_a)) \cap (H(z_{c-1}), H(z_c))$ is a non-empty open sub-interval of

$(0,1)$, there exists a $b \in B$ such that

$$(F(x_{a-1}), F(x_a)) \cap (G(y_{b-1}), G(y_b)) \cap (H(z_{c-1}), H(z_c)) \neq \emptyset.$$

It follows that $a \stackrel{F,G}{\leftrightharpoons} b$ as well as $b \stackrel{G,H}{\leftrightharpoons} c$. By assumption, this yields $r_c \leq q_b \leq p_a$ and since $(a,c) \in R(\stackrel{F,H}{\leftrightharpoons})$ was arbitrary, the transitivity of $\leq_{D-pm}^{disc}$ follows.

Now let $(a,c) \in R(\stackrel{F,H}{\leftrightharpoons}_{\vee})$. We prove the existence of a $b \in \underline{B}$ such that $a \stackrel{F,G}{\leftrightharpoons}_{\vee} b$ and $b \stackrel{G,H}{\leftrightharpoons}_{\vee} c$ via case distinction.

Case 1: $a \leftrightharpoons c$

From the first part of the proof of b), we already know that there exists $b \in B$ such that $a \stackrel{F,G}{\leftrightharpoons} b$ and $b \stackrel{G,H}{\leftrightharpoons} c$. Note that $F(x_{a-1}) < G(y_b)$ follows from $a \stackrel{F,G}{\leftrightharpoons} b$. Assume now $b = \min B$ along with the existence of that minimum. This implies $G(y_{b-1}) = 0$ and therefore $\alpha \stackrel{F,G}{\leftrightharpoons} b$ for all $\alpha \in A \cap (-\infty, a]$. That latter set is not empty and includes at least the element $a-1$ since $a > \min A$. We obtain

$$G(y_b) = G(y_b) - G(y_{b-1}) = q_b \leq p_{a-1} = F(x_{a-1}) - F(x_{a-2}) \leq F(x_{a-1}),$$

a contradiction, and therefore $b \neq \min B$ or, equivalently, $b \in \underline{B}$. Combining $a \stackrel{F,G}{\leftrightharpoons} b$ with $(a,b) \in \underline{A} \times \underline{B}$ yields $a \stackrel{F,G}{\leftrightharpoons}_{\vee} b$ and in the same way, combining $b \stackrel{G,H}{\leftrightharpoons} c$ with $(b,c) \in \underline{B} \times \underline{C}$ yields $b \stackrel{G,H}{\leftrightharpoons}_{\vee} c$.

Case 2: $a-1 \leftrightharpoons c-1$

Analogously to Case 1, there exists a $b \in \underline{B}$ such that $a-1 \stackrel{F,G}{\leftrightharpoons} b-1$ and $b-1 \stackrel{G,H}{\leftrightharpoons} c-1$. It follows directly that $a \stackrel{F,G}{\leftrightharpoons}_{\vee} b$ and $b \stackrel{G,H}{\leftrightharpoons}_{\vee} c$.

It follows that

$$x_a - x_{a-1} \leq y_b - y_{b-1} \leq z_c - z_{c-1},$$

which yields the transitivity of $\leq_{D-supp}^{\vee-disc}$ since $(a,c) \in R(\stackrel{F,H}{\leftrightharpoons}_{\vee})$ was arbitrary. Combined with the fact that $\leq_{D-pm}^{disc}$ is transitive, the transitivity of $\leq_{disp}^{\vee-disc}$ follows.

It remains to give a counterexample for the transitivity of $\leq_{disp}^{\wedge-disc}$. Since we already proved the transitivity of $\leq_{D-pm}^{disc}$, it needs to be a counterexample for the transitivity of $\leq_{D-supp}^{\wedge-disc}$. Let G be defined as F in Example 7.16b) and H be defined as G in that example. Let $F = \text{Bin}(1, \frac{1}{2})$. Just like in the latter part of Example 7.16b), let

$$\begin{aligned} c_y &= y_2 - y_1 = y_3 - y_2, \\ c_z &= z_3 - z_2 = z_4 - z_3 = z_6 - z_5 = z_7 - z_6, \\ \delta &= z_2 - z_1 = z_5 - z_4 = z_8 - z_7. \end{aligned}$$

Set $c_y = 2$, $c_z = 3$ and $\delta = \frac{1}{2}$. We easily obtain

$$R(\overset{F,G}{\leftrightharpoons_\wedge}) = \{(2,2),(2,3)\}, \quad R(\overset{G,H}{\leftrightharpoons_\wedge}) = \{(2,3),(2,4),(3,6),(3,7)\}, \quad R(\overset{F,H}{\leftrightharpoons_\wedge}) = \{(2,5)\}.$$

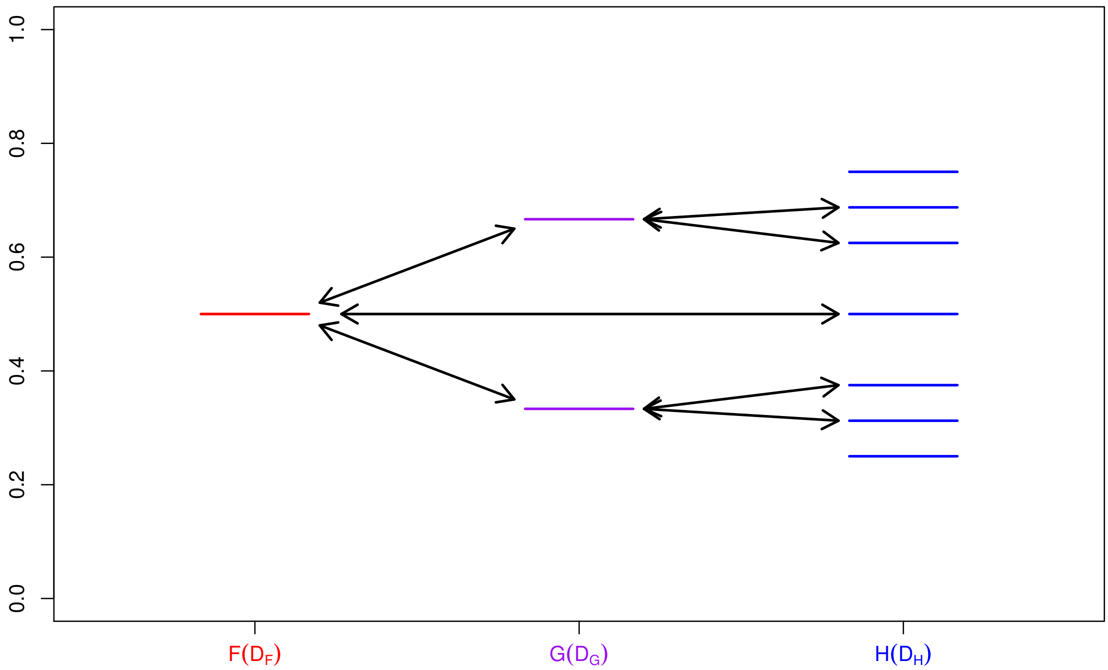

Figure 7.4.: Illustration of the counterexample for the transitivity of $\leq_{D-supp}^{\wedge-disc}$ and $\leq_{disp}^{\wedge-disc}$ (as given in the proof of Theorem 7.22b)). The pairs of constant intervals, of which the lengths are to be compared with respect to $\leftrightharpoons_\wedge$, are identified by double-sided arrows. Each constant interval is represented by the value that the corresponding cdf takes on there.

An illustration of which pairs of constant intervals are compared among the three cdf's is given in Figure 7.4. Because of

$$x_2 - x_1 = 1 \leq 2 = c_y = y_2 - y_1 = y_3 - y_2,$$

$F \leq_{D-supp}^{\wedge-disc} G$ holds, and since

$$\begin{aligned} y_2 - y_1 &= c_y = 2 \leq 3 = c_z = z_3 - z_2 = z_4 - z_3 \quad \text{and} \\ y_3 - y_2 &= c_y = 2 \leq 3 = c_z = z_6 - z_5 = z_7 - z_6, \end{aligned}$$

$G \leq_{D-supp}^{\wedge-disc} H$ holds. However,

$$x_2 - x_1 = 1 > \tfrac{1}{2} = \delta = z_5 - z_4$$

contradicts $F \leq_{D-supp}^{\wedge-disc} H$ and thereby contradicts the transitivity of $\leq_{D-supp}^{\wedge-disc}$ and of $\leq_{disp}^{\wedge-disc}$.

∎

It turns out that both proposed discrete dispersion orders only satisfy one of the desired main properties. $\leq_{disp}^{\vee-disc}$ is transitive but strictly stronger than the original dispersive order on their joint area of applicability. $\leq_{disp}^{\wedge-disc}$ is equivalent to $\leq_{disp}$ on that area but is not transitive. This begs the question whether there even exists a discrete dispersion order that satisfies both properties. While this question cannot be answered rigorously here, we give a heuristic explanation suggesting that such an order does not exist.

Due to Proposition 7.4, any discrete dispersion order $\leq_D$ that is equivalent to $\leq_{disp}$ if $F(D_F) \subseteq G(D_G)$ compares the lengths of constant intervals in a pointwise fashion that is dictated by the candidate for the less dispersed cdf F. This, however, makes it possible to construct a counterexample for the transitivity of $\leq_D$ similar to the one in the proof of Theorem 7.22b). The general approach is to choose $F, G \in \mathcal{D}_0$ with $F(D_F) \cap G(D_G) = \emptyset$ and both sets being small. Then, $H \in \mathcal{D}_0$ is chosen in such a way that $H(D_H)$ is a disjoint union of one set of points close to the elements of $F(D_F)$ and another set of points close to the elements of $G(D_G)$. As a result, the constant intervals of H to be compared with those of F are not compared with any constant intervals of G because of this disjoint split. The transitivity of $\leq_D$ can be contradicted by choosing the lengths of the involved intervals accordingly. Graphically, the situation should be similar to what is depicted in Figure 7.4.

A notable exception to this problem is given by the class of all distributions in $\mathcal{D}_0$ that satisfy (7.3), or, more specifically, by the class $\mathcal{LD}$ of all lattice distributions.

Corollary 7.23. *a) Let $F, G \in \mathcal{LD}$ with defining distances $c_F, c_G > 0$. Then, the following equivalences hold:*

$$F \leq_{disp}^{\vee-disc} G \Leftrightarrow F \leq_{disp}^{\wedge-disc} G \Leftrightarrow F \leq_{disp}^{disc} G \Leftrightarrow F \leq_{D-pm}^{disc} G \quad \text{and} \quad c_F \leq c_G.$$

b) The orders $\leq_{disp}^{disc}$, $\leq_{disp}^{\wedge-disc}$ and $\leq_{disp}^{\vee-disc}$ are transitive on the set $\mathcal{LD}$.

Proof. a) According to Example 7.7b), condition (7.3) is equivalent to $c_F \leq c_G$. However, because of $x_a - x_{a-1} = c_F$ and $y_b - y_{b-1} = c_G$ for all $a \in \underline{A}, b \in \underline{B}$, $c_F \leq c_G$ is also equivalent to both $F \leq_{disp}^{\vee-disc} G$ and $F \leq_{disp}^{\wedge-disc} G$.

b) Let $F, G, H \in \mathcal{LD}$ with defining distances $c_F, c_G, c_H > 0$. Since $\leq_{D-pm}^{disc}$ is transitive as shown in the proof of Theorem 7.22b) and $c_F \leq c_G$ combined with $c_G \leq c_H$ implies $c_F \leq c_H$, the order $\leq_{disp}^{disc}$ is transitive on the set $\mathcal{LD}$, and the assertion follows from part a).

∎

Corollary 7.23 shows that the limitations of our approach to define a discrete dispersive order are not relevant for lattice distributions, which is one of the two most important classes of discrete distributions; the other one being the class of empirical distributions. First, one does not have to choose one of the three given options for discrete dispersive orders because they all coincide. And second, the one remaining discrete dispersive order fulfils both desirable properties stated at the beginning of Section 7.2.

For the remainder of this section, we analyze whether the orders we derived in Section 7.1 change, if we alter the definition of the quantile function. The fact that the quantile function of a discrete distribution is generally not unambiguous is noted in Section 2.1. There, the p-quantile is characterized in a more general sense in (2.3); in order to differentiate this from the specifically defined $F^{-1}(p)$, the notation Q_F^p is introduced. For $p \in (0,1)$, any value within the interval $[\inf\{t \in \mathbb{R} : F(t) \geq p\}, \sup\{t \in \mathbb{R} : F(t) \leq p\}]$ is valid as a p-quantile Q_F^p of a cdf F since $\mathbb{P}(X \leq Q_F^p) \geq p$ and $\mathbb{P}(X \geq Q_F^p) \geq 1-p$ holds for all of these choices. Note that, for $p \notin F(\mathrm{supp}(F))$, $\inf\{t \in \mathbb{R} : F(t) \geq p\} = \sup\{t \in \mathbb{R} : F(t) \leq p\}$ holds and, consequently, all of the above quantile definitions coincide. The definition of the discrete dispersive orders does not involve any definition of the quantile function, neither does the characterization of the original dispersive order used for Theorem 7.21b). This characterization for purposive discrete distributions, given in Proposition 7.4, along with the equivalent condition for the relation $\fallingdotseq$ in Proposition 7.10 are the crucial points, where the definition of the quantile function influences the definitions of the discrete orders and their properties. The proofs of these two results only need to be changed slightly in order to also be valid for any alternate quantile definition from before. This is due to the fact that, as mentioned already, the different quantile definitions all coincide for $p \notin F(\mathrm{supp}(F))$. As an example, in (7.1), which proves one implication of Proposition 7.4, the notation of the utilized limits is changed to

$$\lambda^1(F^{-1}(\{p\})) = \lim_{r \searrow p}(Q_F^r - Q_F^{2p-r}) \leq \lim_{r \searrow p}(Q_G^r - Q_G^{2p-r}) = \lambda^1(G^{-1}(\{p\})).$$

In conclusion, the derivations of our discrete dispersive orders does not depend on the specific definition of the quantile function that is used.

7.3. Further Properties of the Discrete Dispersive Orders

In the following, a number of properties and results concerning the original dispersive order $\leq_{disp}$ are transferred to the discrete setting. First, we consider the equivalence classes of the relation $=_{disp}$, which denotes equivalence with respect to the order $\leq_{disp}$. Note that $=_{disp}$ inherits the properties of reflexivity and transitivity from $\leq_{disp}$ (see Proposition 3.2), and it is symmetric by definition. Thus, $=_{disp}$ is an equivalence relation. The equivalence class of any $F \in \mathcal{D}_0$ with respect to $=_{disp}$ is given by all real shifts of F, i.e. $\{F(\cdot - \lambda) : \lambda \in \mathbb{R}\}$ (see Proposition 3.3b)). The following are the discrete versions of that result.

Theorem 7.24. *Let $F, G \in \mathcal{D}_0$. Then, $F =_{disp}^{\wedge-disc} G$ holds, if and only if there exists a $\lambda \in \mathbb{R}$ such that $G(t) = F(t-\lambda)$ for all $t \in \mathbb{R}$.*

Proof. '⇒': Let $r \in (0,1) \setminus (F(D_F) \cup G(D_G))$, which is possible since $F(D_F) \cup G(D_G)$ is at most countable and $(0,1)$ is uncountable. Now let $a_r = \min\{a \in A : F(x_a) \geq r\}$ (which exists since $r < 1 = \sup F(\mathrm{supp}(F))$ holds and thus, there exists an $a \in A$ such that $r \in (F(x_{a-1}), F(x_a)))$ and $b_r = \min\{b \in B : G(y_b) \geq r\}$. Then, $a_r \leftrightharpoons b_r$ and, by assumption, $F(x_{a_r}) - F(x_{a_r-1}) = p_{a_r} = q_{b_r} = G(y_{b_r}) - G(y_{b_r-1})$ follows. Rearranging yields

$$F(x_{a_r}) - G(y_{b_r}) = F(x_{a_r-1}) - G(y_{b_r-1}).$$

Assume $F(x_{a_r}) > G(y_{b_r})$. It follows that $a_r \leftrightharpoons b_r + 1$, yielding $F(x_{a_r}) - F(x_{a_r-1}) = p_{a_r} = q_{b_r+1} = G(y_{b_r+1}) - G(y_{b_r})$. Note that $b_r + 1 \in B$ since $G(y_{b_r}) < F(x_{a_r}) \leq 1 = \sup G(\mathrm{supp}(G))$. Because of $G(y_{b_r}) > F(x_{a_r-1})$, it follows that $G(y_{b_r+1}) > F(x_{a_r})$, thus yielding $a_r + 1 \leftrightharpoons b_r + 1$. Now the same line of reasoning applied to a_r and b_r can also be applied to $a_r + 1$ and $b_r + 1$. Inductively, it follows that $p_{a_r} = p_\alpha = q_\beta$ for all $\alpha \in A \cap [a_r, \infty)$ and all $\beta \in B \cap [b_r, \infty)$. We now know that there exist $c_A, c_B \in \mathbb{N}_0$ such that

$$F(x_{a_r}) + c_A \cdot p_{a_r} = F(x_{a_r}) + \sum_{a_r+1}^{\sup A} p_\alpha = 1 = G(y_{b_r}) + \sum_{\beta=b_r+1}^{\sup B} p_\beta = G(y_{b_r}) + c_B \cdot p_{a_r},$$

which, since $0 < F(x_{a_r}) - G(y_{b_r}) < p_{a_r}$ (otherwise $a_r \leftrightharpoons b_r$ would not hold), yields

$$0 = F(x_{a_r}) - G(y_{b_r}) + p_{a_r}(c_A - c_B) \neq 0,$$

a contradiction. By symmetry, the case $G(y_{b_r}) > F(x_{a_r})$ also yields a contradiction, leaving only $F(x_{a_r}) = G(y_{b_r})$. It immediately follows that $a_r \leftrightharpoons b_r$ and $a_r + 1 \leftrightharpoons b_r + 1$ (see (7.11) and (7.12)), yielding $p_{a_r} = q_{b_r}$ and $p_{a_r+1} = q_{b_r+1}$. This also yields $F(x_{a_r-1}) = G(y_{b_r-1})$ and $F(x_{a_r+1}) = G(y_{b_r+1})$. Inductively, we obtain $p_{a_r+d} = q_{b_r+d}$ and $F(x_{a_r+d}) = G(y_{b_r+d})$ for all $d \in \mathbb{Z}$ such that $a_r + d \in A$ and $b_r + d \in B$. Furthermore, $a_r + d = \min A$ is equivalent to $b_r + d = \min B$ for all $d \in \mathbb{Z}$ and the same is true for the maximums of A and B. It follows that $F(\mathrm{supp}(F)) = G(\mathrm{supp}(G))$ and, since the indexing sets are uniquely determined by the supports, $A = B$ follows along with $a_r = b_r$. The sets of pairs of indices to be compared are given by $R(\leftrightharpoons) = \{(a,a) : a \in A\}$ and $R(\leftrightharpoons_\wedge) = \{(a,a) : a \in \underline{A}\}$.

Now define $\lambda = y_{a_r} - x_{a_r}$. Let $\alpha \in A$ and, without restriction, let $\alpha \geq a_r$. Then,

$$y_\alpha - x_\alpha = y_{a_r} - x_{a_r} + \sum_{j=a_r+1}^{\alpha} ((y_j - y_{j-1}) - (x_j - x_{j-1})) = y_{a_r} - x_{a_r} = \lambda$$

follows from $F =_{disp}^{\wedge-disc} G$. Overall, we obtain $(B, (y_j, q_j)_{j\in B}) = (A, (x_j + \lambda, p_j)_{j\in A})$ and since the indexing set and the identifying sequence uniquely identify the corresponding cdf, the assertion follows.

'$\Leftarrow$': The assumption directly implies $B = A$, $y_j = x_j + \lambda$ and $q_j = p_j$ for all $j \in A$. The latter observation then implies $F(\mathrm{supp}(F)) = G(\mathrm{supp}(G))$ as well as $R(\leftrightharpoons) = \{(a, a) : a \in A\}$ and $R(\leftrightharpoons_\wedge) = \{(a, a) : a \in \underline{A}\}$. While $F =_{D-pm}^{disc} G$ is now trivial, $F =_{D-supp}^{\wedge-disc} G$ follows from $(x_a + \lambda) - (x_{a-1} + \lambda) = x_a - x_{a-1}$ for all $a \in \underline{A}$, thus concluding the proof.

Corollary 7.25. *Let $F, G \in \mathcal{D}_0$. Then, $F =_{disp}^{\vee-disc} G$ holds, if and only if there exists a $\lambda \in \mathbb{R}$ such that $G(t) = F(t - \lambda)$ for all $t \in \mathbb{R}$.*

Proof. Since $\leq_{disp}^{\wedge-disc}$ is a weakening of $\leq_{disp}^{\vee-disc}$, the implication from left to right follows directly from Theorem 7.24. For the other implication, it is sufficient to show $F \leq_{disp}^{\wedge-disc} G \Leftrightarrow F \leq_{disp}^{\vee-disc} G$ for all $F, G \in \mathcal{D}_0$ with $F(\mathrm{supp}(F)) = G(\mathrm{supp}(G))$. Assuming $F(\mathrm{supp}(F)) = G(\mathrm{supp}(G))$ directly implies $R(\leftrightharpoons) = \{(a, a) : a \in A\}$, where A is the indexing set of either cdf. This yields

$$R(\leftrightharpoons_\wedge) = \{(a, a) : a \in \underline{A}\} = R(\leftrightharpoons_\vee),$$

thus ensuring the equivalence of $\leq_{disp}^{\wedge-disc}$ and $\leq_{disp}^{\vee-disc}$ and concluding the proof.

Theorem 7.24 and Corollary 7.25 state that $F =_{disp}^{\wedge-disc} G$ is equivalent to $F =_{disp}^{\vee-disc} G$ for all $F, G \in \mathcal{D}_0$. Since $=_{disp}^{\vee-disc}$ inherits reflexivity and transitivity as its properties from $\leq_{disp}^{\vee-disc}$ and is obviously symmetric, it is an equivalence relation. Its equivalence classes are, as for $\leq_{disp}$, of the form $\{F(\cdot - \lambda) : \lambda \in \mathbb{R}\}$ for $F \in \mathcal{D}_0$. We can now consider the quotient set of $\mathcal{D}_0$ by $=_{disp}^{\vee-disc}$, denoted by $\mathcal{D}_0/ =_{disp}^{\vee-disc}$, and also define both discrete dispersion orders on that set. For all $\mathcal{F}, \mathcal{G} \in \mathcal{D}_0/ =_{disp}^{\vee-disc}$, the orders are defined by

$$\mathcal{F} \leq_{disp}^{\vee-disc} \mathcal{G}, \quad \text{if and only if} \quad \exists F \in \mathcal{F}, G \in \mathcal{G} : F \leq_{disp}^{\vee-disc} G,$$

and analogously for $\leq_{disp}^{\wedge-disc}$. The consideration of these equivalence classes is relevant to the following result.

Proposition 7.26. *Let $F, G \in \mathcal{D}_0$ not belong to the same equivalence class of $\mathcal{D}_0$ by $=_{disp}^{\wedge-disc}$. Then, it follows from $F \leq_{disp}^{\wedge-disc} G$ that either $\lambda^1(D_F) = \lambda^1(D_G) = \infty$ or $\lambda^1(D_F) < \lambda^1(D_G)$ holds.*

Proof. If $F(D_F) \subseteq G(D_G)$, the order $\leq_{disp}^{\wedge-disc}$ is equivalent to $\leq_{disp}$ and the assertion follows. Otherwise, let $a \in \underline{A}$ such that $F(x_{a-1}) \in F(D_F) \setminus G(D_G) \neq \emptyset$. It follows that

$$\mathrm{NN}_F^G(a) \ni \sup(G(D_G) \cap (0, F(x_{a-1})]) < F(x_{a-1}) < \inf(G(D_G) \cap [F(x_{a-1}), 1)) \in \mathrm{NN}_F^G(a).$$

Hence, $|\,\mathrm{NN}_F^G(a)| = 2$, and we choose $b \in \overline{B}$ such that $\mathrm{NN}_F^G(a) = \{G(y_b), G(y_{b-1})\}$. (Note that $\overline{B} \neq \emptyset$ since $|B| \geq 3$ follows from Lemma 7.18.) Particularly, this means that $a \leftrightharpoons_\wedge b+1$ and, therefore, $x_a - x_{a-1} \leq y_{b+1} - y_b$. Due to Lemma 7.18 and the beginning of the proof of Proposition 7.19, for all $j \in \underline{A}, j \geq a$, there exists a $k_j \in \underline{B}, k_j \geq k_a = b+1$ such that $G(y_{k_j-1}) = \inf(G(D_G) \cap [F(x_{j-1}), 1)) \in \mathrm{NN}_F^G(j)$. It follows that $j \leftrightharpoons_\wedge k_j$ and, particularly, $j-1 \leftrightharpoons k_j - 1$ for all $j \in \underline{A}, j \geq a$. Moreover, the k_j's are pairwise distinct. To see this, let $i, j \in \underline{A} \cap [a, \infty)$ with $i > j$; then we have

$$F(x_{j-1}) \leq F(x_{i-2}) = F(x_{i-1}) - p_{i-1} \leq G(y_{k_i-1}) - q_{k_i-1} = G(y_{k_i-2}),$$

which implies $G(y_{k_j-1}) \leq G({}_{k_i-1})$ and $k_j \leq k_i - 1 < k_i$. It follows that

$$\sum_{j=a}^{\sup \underline{A}} (x_j - x_{j-1}) \leq \sum_{j=a}^{\sup \underline{A}} (y_{k_j} - y_{k_j-1}) \leq \sum_{k=b+1}^{\sup \underline{B}} (y_k - y_{k-1}). \tag{7.17}$$

Instead of using the index a as a starting point upwards, we can also use it as a starting point downwards. It follows from the structure of the set $\mathrm{NN}_F^G(a)$ that $a \leftrightharpoons_\wedge b$ holds and, subsequently, $x_a - x_{a-1} \leq y_b - y_{b-1}$. By Lemma 7.18, we obtain that, for all $j \in \underline{A}, j \leq a$, there exists a $\ell_j \in \underline{B}, \ell_j \leq \ell_a = b$ such that $G(y_{\ell_j-1}) = \sup(G(D_G) \cap (0, F(x_{j-1})]) \in \mathrm{NN}_F^G(j)$. As before, it follows that $j \leftrightharpoons_\wedge \ell_j$ and, particularly, $j \leftrightharpoons \ell_j$ for all $j \in \underline{A}, j \leq a$. To see that the ℓ_j's are also pairwise distinct, let $i, j \in \underline{A} \cap (-\infty, a]$ with $i > j$, yielding

$$F(x_{i-1}) \geq F(x_j) = F(x_{j-1}) + p_j \geq G(y_{\ell_j-1}) + q_{\ell_j} = G(y_{\ell_j})$$

and, therefore, $\ell_i \geq \ell_j + 1 > \ell_j$. It follows that

$$\sum_{j=\inf \underline{A}}^{a} (x_j - x_{j-1}) \leq \sum_{j=\inf \underline{A}}^{a} (y_{\ell_j} - y_{\ell_j-1}) \leq \sum_{\ell=\inf \underline{B}}^{b} (y_\ell - y_{\ell-1}). \tag{7.18}$$

By combining (7.17) and (7.18), we obtain

$$\lambda^1(D_F) = \sum_{j=\inf \underline{A}}^{\sup \underline{A}} (x_j - x_{j-1}) < \sum_{j=\inf \underline{A}}^{\sup \underline{A}} (x_j - x_{j-1}) + (x_a - x_{a-1}) \leq \sum_{k=\inf \underline{B}}^{\sup \underline{B}} (y_k - y_{k-1}) = \lambda^1(D_G),$$

if $\lambda^1(D_F) < \infty$. Otherwise, we obtain that D_G also has infinite Lebesgue measure, since then the leftmost sum in (7.17) or (7.18) is already infinite.

The same result for $\leq_{disp}^{\vee-disc}$ instead of $\leq_{disp}^{\wedge-disc}$ follows directly.

Corollary 7.27. *Let $F, G \in \mathcal{D}_0$ not belong to the same equivalence class of $\mathcal{D}_0$ by $=_{disp}^{\vee-disc}$. Then, it follows from $F \leq_{disp}^{\vee-disc} G$ that either $\lambda^1(D_F) = \lambda^1(D_G) = \infty$ or $\lambda^1(D_F) < \lambda^1(D_G)$ holds.*

An analogous result also holds for the the original dispersive order. It directly follows from its definition in (2.9) and from $\lambda^1(D_F) = \lim_{\alpha\searrow 0}(F^{-1}(1-\alpha) - F^{-1}(\alpha))$, $\lambda^1(D_G) = \lim_{\alpha\searrow 0}(G^{-1}(1-\alpha) - G^{-1}(\alpha))$.

Another result that relates the dispersive order to the supports of the involved distributions is given in Müller and Stoyan (2002, p. 42, Theorem 1.7.6a)) and can also be reproduced for both discrete dispersive orders.

Proposition 7.28. *Let $F, G \in \mathcal{D}_0$. If $F \leq_{disp}^{\wedge-disc} G$ and $\min(\operatorname{supp}(F)) \leq \min(\operatorname{supp}(G))$ with both minimums existing, then $F \leq_{st} G$.*

Proof. Similarly as in the proof of Proposition 7.26, Lemma 7.18 states that for all $a \in \underline{A}$, there exists a $b_a \in \underline{B}$ such that $G(y_{b_a-1}) = \sup(G(D_G) \cap (0, F(x_{a-1})]) \in \mathrm{NN}_F^G(a)$, implying $a \leftrightharpoons_\wedge b_a$. As shown for Proposition 7.26, these b_a's are pairwise distinct. It follows for all $a \in A$ that

$$
\begin{aligned}
x_a = \min(\operatorname{supp}(F)) + \sum_{j=\min \underline{A}}^{a} (x_j - x_{j-1}) &\leq \min(\operatorname{supp}(G)) + \sum_{j=\min \underline{A}}^{a} (y_{b_j} - y_{b_j-1}) \\
&\leq \min(\operatorname{supp}(G)) + \sum_{k=\min \underline{B}}^{b_a} (y_k - y_{k-1}) = y_{b_a}.
\end{aligned}
$$

Note that for all $(a, b) \in R(\leftrightharpoons)$, $b \geq b_a$ holds because of $G(y_b) > F(x_{a-1}) \geq G(y_{b_a-1})$ and the maximality of b_a. This means that, for all $(a, b) \in R(\leftrightharpoons)$, we obtain $x_a \leq y_{b_a} \leq y_b$. According to Lemma A.12, this is equivalent to $F \leq_{st} G$.

Corollary 7.29. *Let $F, G \in \mathcal{D}_0$. If $F \leq_{disp}^{\vee-disc} G$ and $\min(\operatorname{supp}(F)) \leq \min(\operatorname{supp}(G))$ with both minimums existing, then $F \leq_{st} G$.*

Analogously, if F is less dispersed than G with respect to either order, and $\max(\operatorname{supp}(F)) \geq \max(\operatorname{supp}(G))$ holds with both maximums existing, $F \leq_{st} G$ also follows.

It is worth noting that Proposition 7.28 and Corollary 7.29 relate the respective discrete dispersive order to the usual stochastic order in the same way as the original dispersive order is related to the stochastic order. Since both the stochastic order and the dispersive order are orders of convex characteristics, these results somewhat legitimize $\leq_{disp}^{\wedge-disc}$ and $\leq_{disp}^{\vee-disc}$ in their asserted roles as discrete orders of the first convex characteristic.

The same can be said about the next results, which relate the discrete dispersive orders to the so-called weak dispersive order, which is also based on the usual stochastic order. Note that the weak dispersive order considered here is different from the order with the same name from Definition 3.21a) in Part I of this thesis. The order discussed here in Part II was introduced and noted to be weaker than $\leq_{disp}$ by Giovagnoli and Wynn (1995, p. 326), who used it as a starting point for a multivariate dispersion order. They said that F precedes G in the weak dispersive order, if $|X - X'| \leq_{st} |Y - Y'|$ holds for $X, X' \sim F$ independent and $Y, Y' \sim G$ independent.

Theorem 7.30. *Let $F, G \in \mathcal{D}_0$ with $X, X' \sim F$ independent and $Y, Y' \sim G$ independent. Then, $F \leq_{disp}^{\wedge-disc} G$ implies $|X - X'| \leq_{st} |Y - Y'|$.*

Proof. Let $(a, b), (\alpha, \beta) \in R(\leftrightharpoons)$. For better cross-reference, we divide the proof into three parts.

Part 1: In this part, we show by contradiction that $|x_\alpha - x_a| > |y_\beta - y_b|$ implies $|\alpha - a| = |\beta - b| + 1$. To this end, we first assume $|\alpha - a| \leq |\beta - b|$. Without restriction, let $\alpha \geq a$. If $\alpha = a$, then $|x_\alpha - x_a| = 0 \leq |y_\beta - y_b|$ follows, contradicting the assumption. Hence, it remains to consider the case $\alpha > a$. Because of $a \leftrightharpoons b$ and $\alpha \leftrightharpoons \beta$, we obtain

$$G(y_{b-1}) < F(x_a) \leq F(x_{\alpha-1}) < G(y_\beta), \tag{7.19}$$

yielding $\beta \geq b$. We either have $G(y_{\beta-1}) \leq F(x_{\alpha-1})$, yielding $G(y_{\beta-1}) = \sup(G(D_G) \cap (0, F(x_{\alpha-1})]) \in \mathrm{NN}_F^G(\alpha)$; in this case we define $k_\alpha = \beta$. Or we have $G(y_{\beta-1}) > F(x_{\alpha-1})$, implying that there exists a $k_\alpha \leq \beta$ such that $G(y_{k_\alpha-1}) = \inf(G(D_G) \cap [F(x_{\alpha-1}), 1)) \in \mathrm{NN}_F^G(\alpha)$. Note that $k_\alpha > b$ holds because of $G(y_{k_\alpha-1}) \geq F(x_{\alpha-1})$ and (7.19). So, considering both cases as well as Proposition 7.19a), there exists a $k_\alpha \in \{b+1, \ldots, \beta\}$ such that $\alpha \leftrightharpoons_\wedge k_\alpha$.

It follows that $\alpha - 1 \leftrightharpoons k_\alpha - 1$ and we can repeat the line of argument from above with $\alpha - 1$ taking the role of α and $k_\alpha - 1$ taking the role of β, as long as $\alpha - 1 > a$. We then obtain a $k_{\alpha-1} \in \{b+1, \ldots, k_\alpha - 1\}$ such that $\alpha - 1 \leftrightharpoons_\wedge k_{\alpha-1}$ and can repeat the line of argument again, starting with $\alpha - 2 \leftrightharpoons k_{\alpha-1} - 1$ as long as $\alpha - 2 > a$. Iteratively, we obtain the following statement:

$$\forall j \in \{a+1, \ldots, \alpha\} \; \exists \text{ pairwise distinct } k_j \in \{b+1, \ldots, \beta\} \text{ such that } j \leftrightharpoons_\wedge k_j.$$

Since $F \leq_{D-supp}^{\wedge-disc} G$ holds, it follows

$$\begin{aligned} |x_\alpha - x_a| = x_\alpha - x_a &= \sum_{j=a+1}^{\alpha} (x_j - x_{j-1}) \\ &\leq \sum_{j=a+1}^{\alpha} (y_{k_j} - y_{k_j-1}) \\ &\leq \sum_{k=b+1}^{\beta} (y_k - y_{k-1}) = y_\beta - y_b = |y_\beta - y_b|, \end{aligned}$$

a contradiction to the assumption $|x_\alpha - x_a| > |y_\beta - y_b|$. This closes the case $|\alpha - a| \leq |\beta - b|$.

Second, we assume $|\alpha - a| \geq |\beta - b| + 2$. Again, let $\alpha \geq a$. Because of $\alpha - a = |\alpha - a| \geq |\beta - b| + 2 \geq 2$, we have $\alpha > a$. By combining $a \leftrightharpoons b$ and $\alpha \leftrightharpoons \beta$, we obtain (7.19) and

$\beta \geq b$, as before. It follows from (7.19) that, for every $j \in \{a+1, \ldots, \alpha-1\}$, there exists a $k \in \{b, \ldots, \beta\}$ such that $j \leftrightharpoons k$. More specifically, we define $k_j = \min\{k \in \{b, \ldots \beta\} : j \leftrightharpoons k\}$ for every $j \in \{a+1, \ldots, \alpha-1\}$. These k_j are pairwise distinct; otherwise there would exist indices $i, j \in \{a+1, \ldots, \alpha-1\}, i \neq j$ (without restriction $i < j$) such that $k_i = k_j$. This implies $G(y_{k_i-1}) \leq F(x_{i-1})$ because of the minimality of k_i. We obtain

$$G(y_{k_i}) = G(y_{k_i-1}) + q_{k_i} \leq F(x_{i-1}) + p_i = F(x_i) \leq F(x_{j-1}) < G(y_{k_j}),$$

yielding a contradiction, and thereby proving that the mapping $j \mapsto k_j$ is injective. The cardinal number of the domain $\{a+1, \ldots, \alpha-1\}$ of that mapping is $|\alpha - a| - 1$ and by assumption larger than or equal to the cardinal number $|\beta - b| + 1$ of its codomain $\{b, \ldots, \beta\}$. For $|\alpha - a| > |\beta - b| + 2$, this directly contradicts the injectivity of the mapping. For $|\alpha - a| = |\beta - b| + 2$, it follows that the mapping is bijective and we obtain

$$F(x_{\alpha-1}) - F(x_a) = \sum_{j=a+1}^{\alpha-1} p_j \geq \sum_{j=a+1}^{\alpha-1} q_{k_j} = \sum_{k=b}^{\beta} q_k = G(y_\beta) - G(y_{b-1}).$$

This, however, contradicts (7.19). Thus, we have shown the implication

$$|x_\alpha - x_a| > |y_\beta - y_b| \Longrightarrow |\alpha - a| = |\beta - b| + 1 \quad \forall (a,b), (\alpha, \beta) \in R(\leftrightharpoons) \tag{7.20}$$

by excluding all other possibilities.

Part 2: It becomes apparent in Part 3 of the proof that pairs $(a,b), (\alpha, \beta) \in R(\leftrightharpoons)$ satisfying $|x_\alpha - x_a| \leq |y_\beta - y_b|$ are easy to deal with. Therefore, the critical situation $|x_\alpha - x_a| > |y_\beta - y_b|$ is of particular interest. Considering in the final result (7.20) of Part 1, this implies $|\alpha - a| = |\beta - b| + 1$. It is the purpose of Part 2 to analyze the situation for these kinds of pairs more closely, so let $(a,b), (\alpha, \beta) \in R(\leftrightharpoons)$ with $|\alpha - a| = |\beta - b| + 1$. Furthermore, let $\alpha > a$. As in Part 1, we obtain (7.19) and thereby $\beta \geq b$. It follows from (7.19) and $a \leftrightharpoons b$ that, for every $j \in \{a, \ldots, \alpha - 1\}$, there exists a $k \in \{b, \ldots, \beta\}$ such that $j \leftrightharpoons k$. We consider the mapping

$$\varphi_1 : \{a, \ldots, \alpha - 1\} \to \{b, \ldots, \beta\}, \quad j \mapsto k_j = \max\{k \in \{b, \ldots \beta\} : j \leftrightharpoons k\}.$$

To see that φ_1 is strictly increasing, let $j, j+1 \in \{a, \ldots, \alpha - 1\}$. It follows from the maximality of k_{j+1} that $G(y_{k_{j+1}}) \geq F(x_{j+1})$, yielding

$$G(y_{k_{j+1}-1}) = G(y_{k_{j+1}}) - q_{k_{j+1}} \geq F(x_{j+1}) - p_{j+1} = F(x_j) > G(y_{k_j-1})$$

and, thereby, $k_{j+1} > k_j$. Thus, φ_1 is strictly increasing and therefore also injective. Since $\alpha - a - 1 = \beta - b$ means that the domain and the codomain of φ_1 have the same cardinal number, the mapping is bijective. This means that

$$a \leftrightharpoons b,\ a+1 \ \leftrightharpoons\ b+1,\ \dots,\ \alpha-2 \ \leftrightharpoons\ \beta-1,\ \alpha-1 \ \leftrightharpoons\ \beta, \tag{7.21}$$

$$a+1 \leftrightharpoons_\wedge b+1,\ \dots,\ \alpha-2 \leftrightharpoons_\wedge \beta-1,\ \alpha-1 \leftrightharpoons_\wedge \beta. \tag{7.22}$$

Combined with (7.19), this has the following two implications:

$$\begin{aligned} G(y_{b-1}) - F(x_{a-1}) &= \left(G(y_\beta) - \sum_{k=b}^{\beta} q_k \right) - \left(F(x_{\alpha-1}) - \sum_{j=a}^{\alpha-1} p_j \right) \\ &= G(y_\beta) - F(x_{\alpha-1}) + \sum_{j=a}^{\alpha-1} (p_j - q_{k_j}) \geq G(y_\beta) - F(x_{\alpha-1}) > 0, \end{aligned} \tag{7.23}$$

$$x_{\alpha-1} - x_a = \sum_{j=a+1}^{\alpha-1} (x_j - x_{j-1}) \leq \sum_{j=a+1}^{\alpha-1} (y_{k_j} - y_{k_j-1}) = \sum_{k=b+1}^{\beta} (y_k - y_{k-1}) = y_\beta - y_b. \tag{7.24}$$

It follows from (7.19), $a \leftrightharpoons b$ and $\alpha \leftrightharpoons \beta$ that, for every $j \in \{a, \dots, \alpha\}$, there exists a $k \in \{b-1, \dots, \beta\}$ such that $j \leftrightharpoons k$. Note that $b-1 \in B$ because of (7.23), which implies $G(y_{b-1}) > F(x_{a-1}) \geq 0$. We now consider the mapping

$$\varphi_2 : \{a, \dots, \alpha\} \to \{b-1, \dots, \beta\}, \quad j \mapsto k_j = \min\{k \in \{b-1, \dots \beta\} : j \leftrightharpoons k\}.$$

To see that φ_2 is strictly increasing, let $j, j+1 \in \{a, \dots, \alpha\}$. It follows from the minimality of k_j that $G(y_{k_j-1}) \leq F(x_{j-1})$, yielding

$$G(y_{k_j}) = G(y_{k_j-1}) + q_{k_j} \leq F(x_{j-1}) + p_j = F(x_j) = F(x_{(j+1)-1}) < G(y_{k_{j+1}})$$

and, thereby, $k_j < k_{j+1}$. Thus, φ_2 is strictly increasing and therefore also injective. Since $\alpha - a = \beta - b + 1$ means that the domain and the codomain of φ_2 have the same cardinal number, the mapping is bijective. This means that

$$a \leftrightharpoons b-1,\ a+1 \ \leftrightharpoons\ b,\ \dots,\ \alpha-1 \ \leftrightharpoons\ \beta-1,\ \alpha \ \leftrightharpoons\ \beta, \tag{7.25}$$

$$a+1 \leftrightharpoons_\wedge b,\ \dots,\ \alpha-1 \leftrightharpoons_\wedge \beta-1,\ \alpha \leftrightharpoons_\wedge \beta. \tag{7.26}$$

Combined with (7.19), this has the following two implications:

$$F(x_{\alpha-1}) - G(y_{\beta-1}) = \left(F(x_a) + \sum_{j=a+1}^{\alpha-1} p_j \right) - \left(G(y_{b-1}) + \sum_{k=b}^{\beta-1} q_k \right)$$

$$= F(x_a) - G(y_{b-1}) + \sum_{j=a+1}^{\alpha-1} (p_j - q_{k_j}) \geq F(x_a) - G(y_{b-1}) > 0, \tag{7.27}$$

$$x_\alpha - x_a = \sum_{j=a+1}^{\alpha} (x_j - x_{j-1}) \leq \sum_{j=a+1}^{\alpha} (y_{k_j} - y_{k_j-1}) = \sum_{k=b}^{\beta} (y_k - y_{k-1}) = y_\beta - y_{b-1}. \tag{7.28}$$

Part 3: Recall the definition of $r_{(a,b)}$ in (7.6) to see that the following holds true for all $t \in \mathbb{R}$:

$$\begin{aligned}
H_{|X-X'|}(t) &= \mathbb{P}(|X - X'| \leq t) \\
&= \sum_{x,x' \in \operatorname{supp}(F)} \mathbb{P}(X = x)\mathbb{P}(X' = x')\mathbb{1}\{|x - x'| \leq t\} \\
&= \sum_{a,\alpha \in A} p_a p_\alpha \mathbb{1}\{|x_a - x_\alpha| \leq t\} \\
&= \sum_{a,\alpha \in A} \left(\sum_{b \in B: a \leftrightharpoons b} r_{(a,b)} \right) \left(\sum_{\beta \in B: \alpha \leftrightharpoons \beta} r_{(\alpha,\beta)} \right) \mathbb{1}\{|x_a - x_\alpha| \leq t\} \\
&= \sum_{(a,b),(\alpha,\beta) \in R(\leftrightharpoons)} r_{(a,b)} r_{(\alpha,\beta)} \mathbb{1}\{|x_a - x_\alpha| \leq t\}.
\end{aligned} \tag{7.29}$$

Analogously, we obtain

$$H_{|Y-Y'|}(t) = \mathbb{P}(|Y - Y'| \leq t) = \sum_{(a,b),(\alpha,\beta) \in R(\leftrightharpoons)} r_{(a,b)} r_{(\alpha,\beta)} \mathbb{1}\{|y_b - y_\beta| \leq t\} \tag{7.30}$$

for all $t \in \mathbb{R}$. The claim of the theorem is equivalent to

$$\begin{aligned}
&0 \leq H_{|X-X'|}(t) - H_{|Y-Y'|}(t) \\
&= \sum_{(a,b),(\alpha,\beta) \in R(\leftrightharpoons)} r_{(a,b)} r_{(\alpha,\beta)} \big(\mathbb{1}\{|x_\alpha - x_a| \leq t\} - \mathbb{1}\{|y_\beta - y_b| \leq t\}\big)
\end{aligned} \tag{7.31}$$

for all $t \in \mathbb{R}$. By Part 1, we have for all $(a, b), (\alpha, \beta) \in R(\leftrightharpoons)$ that

$$\begin{aligned}
|\alpha - a| \neq |\beta - b| + 1 \Longrightarrow\ & |x_\alpha - x_a| \leq |y_\beta - y_b| \\
\iff\ & (|y_\beta - y_b| \leq t \Longrightarrow |x_\alpha - x_a| \leq t) \quad \forall t \in \mathbb{R} \\
\iff\ & \mathbb{1}\{|x_\alpha - x_a| \leq t\} - \mathbb{1}\{|y_\beta - y_b| \leq t\} \geq 0 \quad \forall t \in \mathbb{R}.
\end{aligned}$$

The sums in (7.29) and (7.30) can each be split into three separate sums, one with $\alpha > a$, one with $\alpha = a$ and one with $\alpha < a$. Since all the summands in (7.29) and (7.30) are symmetric in (a, b) and (α, β), the sum with $\alpha > a$ is equal to the sum with $\alpha < a$.

Because $\alpha = a$ and $|\alpha - a| = |\beta - b| + 1$ are not possible simultaneously, we obtain for all $t \in \mathbb{R}$ that

$$\begin{aligned}
&H_{|X-X'|}(t) - H_{|Y-Y'|}(t) \\
&\quad = 2 \sum_{\substack{(a,b),(\alpha,\beta)\in R(\Leftarrow\!\!\!\!\Rightarrow):\\ \alpha>a}} r_{(a,b)}r_{(\alpha,\beta)}\big(\mathbb{1}\{|x_\alpha - x_a| \le t\} - \mathbb{1}\{|y_\beta - y_b| \le t\}\big) \\
&\qquad + \sum_{\substack{(a,b),(\alpha,\beta)\in R(\Leftarrow\!\!\!\!\Rightarrow):\\ \alpha=a}} r_{(a,b)}r_{(\alpha,\beta)}\big(\mathbb{1}\{|x_\alpha - x_a| \le t\} - \mathbb{1}\{|y_\beta - y_b| \le t\}\big) \\
&\quad \ge 2 \sum_{\substack{(a,b),(\alpha,\beta)\in R(\Leftarrow\!\!\!\!\Rightarrow):\\ |\alpha-a|=|\beta-b|+1,\\ \alpha>a}} \Big[r_{(a,b)}r_{(\alpha,\beta)}\big(\mathbb{1}\{x_\alpha - x_a \le t\} - \mathbb{1}\{y_\beta - y_b \le t\}\big) \\
&\qquad + r_{(a,b-1)}r_{(\alpha-1,\beta)}\big(\mathbb{1}\{x_{\alpha-1} - x_a \le t\} - \mathbb{1}\{y_\beta - y_{b-1} \le t\}\big)\Big]. \qquad (7.32)
\end{aligned}$$

For the validity of the inequality we use that, under the assumptions $(a,b),(\alpha,\beta) \in R(\Leftarrow\!\!\!\!\Rightarrow)$, $|\alpha - a| = |\beta - b| + 1$ and $\alpha > a$, it follows that $(a, b-1), (\alpha-1, \beta) \in R(\Leftarrow\!\!\!\!\Rightarrow)$ according to (7.21) and (7.25). Note that $|(\alpha - 1) - a| \ne |\beta - (b-1)| + 1$ holds true in all summands of the last sum, so no summand is used twice in that sum. For the differences of the indicator functions, it holds that

$$\begin{aligned}
\mathbb{1}\{x_\alpha - x_a \le t\} - \mathbb{1}\{y_\beta - y_b \le t\} = -1 &\iff t \in [y_\beta - y_b, x_\alpha - x_a) \quad \text{and} \\
\mathbb{1}\{x_{\alpha-1} - x_a \le t\} - \mathbb{1}\{y_\beta - y_{b-1} \le t\} = 1 &\iff t \in [x_{\alpha-1} - x_a, y_\beta - y_{b-1})
\end{aligned}$$

for all $(a,b),(\alpha,\beta) \in R(\Leftarrow\!\!\!\!\Rightarrow)$ with $|\alpha - a| = |\beta - b| + 1$ and $\alpha > a$. Because of (7.24) and (7.28), $[y_\beta - y_b, x_\alpha - x_a) \subseteq [x_{\alpha-1} - x_a, y_\beta - y_{b-1})$ follows, yielding

$$\mathbb{1}\{x_\alpha - x_a \le t\} - \mathbb{1}\{y_\beta - y_b \le t\} = -1 \Longrightarrow \mathbb{1}\{x_{\alpha-1} - x_a \le t\} - \mathbb{1}\{y_\beta - y_{b-1} \le t\} = 1.$$

This means that in order to prove that (7.32) is larger than or equal to zero and thereby complete the proof, it remains to show that $r_{(a,b)}r_{(\alpha,\beta)} \le r_{(a,b-1)}r_{(\alpha-1,\beta)}$ holds for all $(a,b),(\alpha,\beta) \in R(\Leftarrow\!\!\!\!\Rightarrow)$ with $|\alpha - a| = |\beta - b| + 1$ and $\alpha > a$. In that setting, by combining (7.23) and (7.27), we obtain

$$\begin{aligned}
r_{(a,b)}r_{(\alpha,\beta)} &= \big(F(x_a) - G(y_{b-1})\big)\big(G(y_\beta) - F(y_{\alpha-1})\big) \\
&\le \big(F(x_{\alpha-1}) - G(y_{\beta-1})\big)\big(G(y_{b-1}) - F(x_{a-1})\big) = r_{(\alpha-1,\beta)}r_{(a,b-1)}.
\end{aligned}$$

▨

The following corollary directly follows from above theorem and the fact that $\le_{disp}^{\wedge-disc}$ is a weakening of $\le_{disp}^{\vee-disc}$.

Corollary 7.31. *Let $F, G \in \mathcal{D}_0$ with $X, X' \sim F$ independent and $Y, Y' \sim G$ independent. Then, $F \leq_{disp}^{\vee-disc} G$ implies $|X - X'| \leq_{st} |Y - Y'|$.*

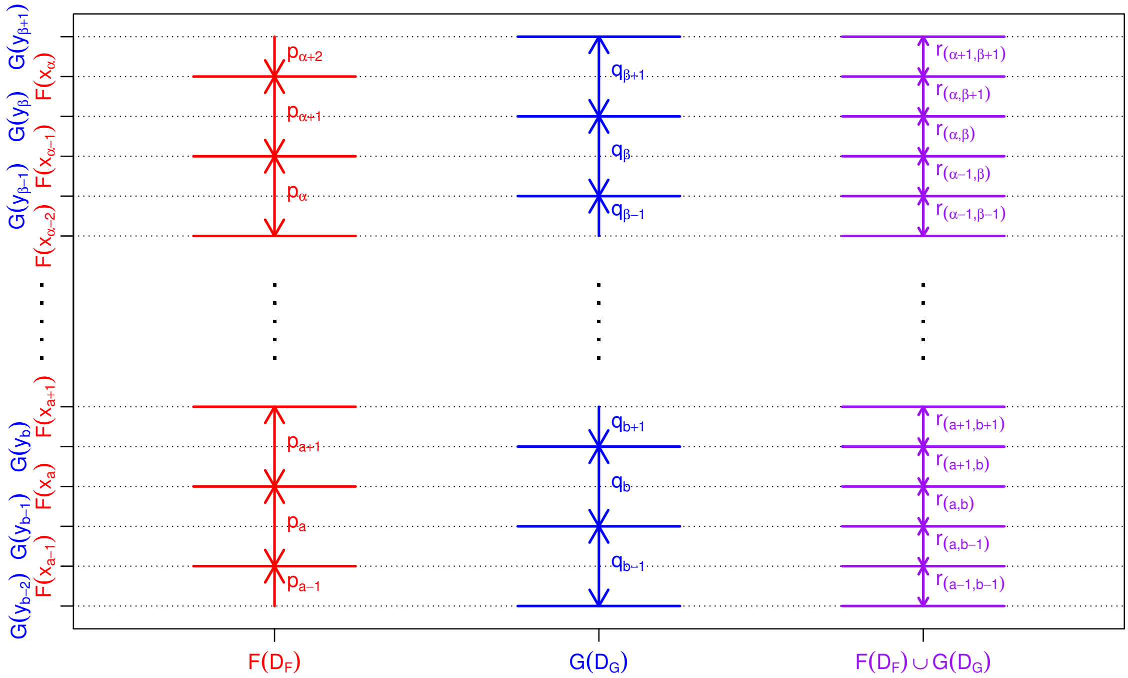

Figure 7.5.: Exemplary visualization of crucial situation in the proof of Theorem 7.30. The horizontal lines represent the elements of the sets given on the x-axis. The double-sided arrows and the variable names next to them represent the distances between those horizontal lines.

The starting point of the proof of Theorem 7.30 is the representation of the cdf of $|X - X'|$ given in (7.29) and the subsequent equivalent characterization of the assertion given in (7.31). Because of $r_{(a,b)} \geq 0$ for all $(a, b) \in R(\leftrightharpoons)$, each summand in (7.31) is negative, if and only if the difference of indicator functions is negative. The implication (7.20), which is shown in Part 1 of the proof, states that this difference of indicator functions can only be negative if $|\alpha - a| = |\beta - b| + 1$, which is a crucial restriction. In Part 3 of the proof, it is shown that, for every two pairs of indices $(a, b), (\alpha, \beta)$ for which the difference of indicator functions is equal to -1, there exists another set of indices $(a, b-1), (\alpha-1, \beta)$ for which the difference of indicator functions is equal to 1. Additionally, it is shown that $r_{(a,b-1)} r_{(\alpha-1,\beta)}$, the coefficient for the latter set, is always larger than $r_{(a,b)} r_{(\alpha,\beta)}$, the coefficient for the former set. So for every negative summand in (7.31), there exists a corresponding positive summand with equal or larger absolute value. The situation is illustrated in Figure 7.5. The fact that the values of $F(D_F)$ and $G(D_G)$ in question are alternating, follows from $|\alpha - a| = |\beta - b| + 1$ and is derived in Part 2 of the proof. The plot also hints at the fact that, for reasons of symmetry, $(a+1, b), (\alpha, \beta+1)$ would have been a viable alternative of $(a, b-1), (\alpha-1, \beta)$ in its role as compensatory set of indices.

Since $\leq_{disp}^{\wedge-disc}$ is a strictly weaker order than $\leq_{disp}^{\vee-disc}$, it is obvious that the statement of Corollary 7.31 is a strict implication and no equivalence. The same can be proved for Theorem 7.30 by making use of the fact that $\leq_{disp}^{\wedge-disc}$ is not transitive. For that, let $F, G, H \in \mathcal{D}_0$ be defined as in the part of the proof of Theorem 7.22b), where the counterexample for the transitivity of $\leq_{disp}^{\wedge-disc}$ was constructed. It is shown there that $F \leq_{disp}^{\wedge-disc} G$ and $G \leq_{disp}^{\wedge-disc} H$, but $F \not\leq_{disp}^{\wedge-disc} H$. If we now let $X, X' \sim F$ independent, $Y, Y' \sim G$ independent and $Z, Z' \sim H$ independent, Theorem 7.30 yields $|X - X'| \leq_{st} |Y - Y'|$ as well as $|Y - Y'| \leq_{st} |Z - Z'|$. Since the stochastic order is transitive (see Proposition 3.2), we have now shown $|X - X'| \leq_{st} |Z - Z'|$ while $F \not\leq_{disp}^{\wedge-disc} H$ holds and, thus, that the statement of Theorem 7.30 is indeed a strict implication.

The implications given in Theorem 7.30 and Corollary 7.31 are also useful in Section 7.4, where the compatibility of the discrete dispersive orders with popular dispersion measures in analyzed. If a measure of dispersion can be written as a location measure applied to a term of the form $|X - X'|$, it obviously preserves the weak dispersion order. Because of Theorem 7.30 and Corollary 7.31, the dispersion measure then also preserves the discrete dispersive orders $\leq_{disp}^{\wedge-disc}$ and $\leq_{disp}^{\vee-disc}$. A popular measure of this type is Gini's mean difference, which is given by $\mathbb{E}[|X - X'|]$.

Another term used in well-established dispersion measures like the mean absolute deviation from the median is $|X - F^{-1}(\frac{1}{2})|$ instead of $|X - X'|$, where, as before, $F \in \mathcal{D}_0$ and $X, X' \sim F$ are independent. So if, analogously to Theorem 7.30 and Corollary 7.31, we could prove that $|X - F^{-1}(\frac{1}{2})| \leq_{st} |Y - G^{-1}(\frac{1}{2})|$ follows from $F \leq_{disp}^{\wedge-disc} G$ and/or $F \leq_{disp}^{\vee-disc} G$ for all $F, G \in \mathcal{D}_0$, this would heavily simplify the analysis of these dispersion measures. However, the following counterexample proves that this implication does not hold. Note that this does not necessarily imply that any of the dispersion measures mentioned above do not preserve the discrete dispersive orders.

Let $F, G \in \mathcal{D}_0$ be defined by

$$\mathbb{P}(X = 0) = \frac{1}{2},\ \mathbb{P}(X = 1) = \mathbb{P}(X = 2) = \frac{1}{4},$$
$$\mathbb{P}(Y = 0) = \frac{3}{8},\ \mathbb{P}(Y = 1) = \mathbb{P}(Y = 2) = \frac{1}{4},\ \mathbb{P}(Y = 3) = \frac{1}{8}$$

(see upper panel of Figure 7.6). It follows $F(D_F) = \{\frac{1}{2}, \frac{3}{4}\}, G(D_G) = \{\frac{3}{8}, \frac{5}{8}, \frac{7}{8}\}$ and the indexing sets of F and G are $A = \{1, \ldots, 3\}$ and $B = \{1, \ldots, 4\}$, respectively. The medians are given by $F^{-1}(\frac{1}{2}) = 0$ and $G^{-1}(\frac{1}{2}) = 1$ with $F(F^{-1}(\frac{1}{2})) = \frac{1}{2}$ and $G(G^{-1}(\frac{1}{2})) = \frac{5}{8}$. We obtain

$$R(\leftrightharpoons) = \{(1,1), (1,2), (2,2), (2,3), (3,3), (3,4)\}.$$

By combining this with the jump heights given in the definition of F and G, it is easy to see that $F \leq_{D-pm}^{disc} G$ holds. Since, furthermore, $x_a - x_{a-1} = 1 = y_b - y_{b-1}$ for all $(a, b) \in \underline{A} \times \underline{B}$, i.e. $F, G \in \mathcal{LD}$ with defining distance 1, condition (7.3) is satisfied. This implies $F \leq_{D-supp}^{\wedge-disc} G$

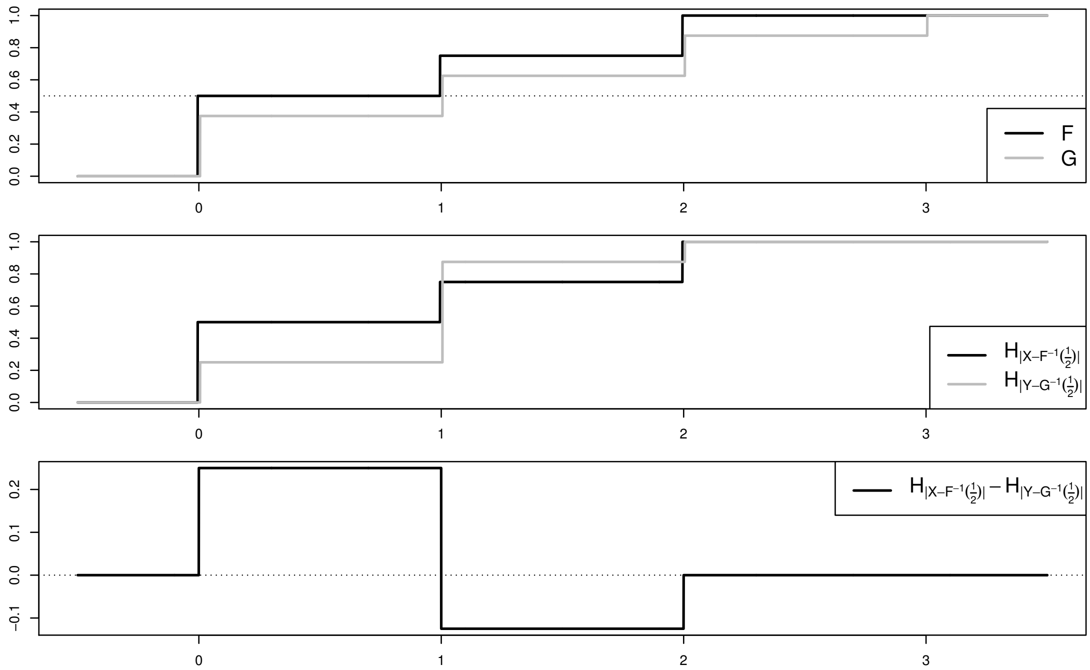

Figure 7.6.: Illustration of the given counterexample for the implication $F \leq_{disp}^{\vee-disc} G \Rightarrow |X - F^{-1}(\frac{1}{2})| \leq_{st} |Y - G^{-1}(\frac{1}{2})|$. Upper panel: Cdf's. Central panel: Cdf's of the absolute deviations from the medians. Lower panel: Difference of the cdf's in the central panel.

as well as $F \leq_{D-supp}^{\vee-disc} G$, which, in turn, implies $F \leq_{disp}^{\wedge-disc} G$ and $F \leq_{disp}^{\vee-disc} G$. Now let $t \in [1, 2)$. It follows

$$\begin{aligned} H_{|X-F^{-1}(\frac{1}{2})|}(t) &= \mathbb{P}(|X - 0| \leq t) = \mathbb{P}(X \leq t) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \frac{3}{4} \\ &< \frac{7}{8} = \mathbb{P}(Y = 0) + \mathbb{P}(Y = 1) + \mathbb{P}(Y = 2) \\ &= \mathbb{P}(1 - t \leq Y \leq 1 + t) = \mathbb{P}(|Y - 1| \leq t) = H_{|Y-G^{-1}(\frac{1}{2})|}(t) \end{aligned}$$

(see lower tow panels of Figure 7.6). This proves that neither $F \leq_{disp}^{\wedge-disc} G$ nor $F \leq_{disp}^{\vee-disc} G$ implies $|X - F^{-1}(\frac{1}{2})| \leq_{st} |Y - G^{-1}(\frac{1}{2})|$.

Another dispersion order, which is weaker than $\leq_{disp}$ for continuous distributions, is the dilation order $\leq_{dil}$. It is introduced in Definition 2.18 and Example 2.21b), and it is shown to be a weakening of $\leq_{disp}$ in Proposition 2.20b) and Example 2.21b). Throughout Chapter 3, the dilation order is often used as an intermediate step in showing that a certain dispersion measure (e.g. the standard deviation) preserves the dispersive order $\leq_{disp}$. Hence, proving that the discrete dispersive orders also imply the dilation order is very helpful for Section 7.4. For its use in the following results, recall that the dilation order is characterized by the corresponding stop-loss transforms via (2.12).

The proof of the implication $F \leq_{disp} G \Rightarrow F \leq_{dil} G$ given in Proposition 2.20b) makes use of the intersection criteria for the dispersive order. Specifically, it is first proved that, under the assumption of equal means, $F \leq_{disp} G$ implies that F and G intersect exactly once with F being smaller than G before the intersection and larger afterwards (see Proposition 2.17b)).

The fact that the discrete dispersive orders also imply the dilation order is proved in a similar way. However, the intersection criterion in this case is not as simple as for $\leq_{disp}$. The following lemma gives the corresponding result, which is the discrete analogue of Proposition 2.17b).

Lemma 7.32. *Let $F, G \in \mathcal{D}_0$ with $F \neq G$ have finite and coinciding means and satisfy $F \leq_{disp}^{\wedge-disc} G$. Then:*

a) $\exists (a,b) \in A \times \underline{B} : F(x_{a-1}) \leq G(y_{b-1}) \leq F(x_a), y_{b-1} < x_a \leq y_b$.

b) One of the following two statements is true:

(i) $\exists (a,b) \in \overline{A} \times \overline{B} : F(x_a) = G(y_b), y_b < x_a, x_{a+1} \leq y_{b+1}$ *or*

(ii) $\exists (a,b) \in A \times \underline{B} : F(x_{a-1}) < G(y_{b-1}) < F(x_a), y_{b-1} < x_a \leq y_b$.

Proof. Let $X \sim F$, $Y \sim G$ and, as assumed $\mathbb{E}[X] = \mathbb{E}[Y]$ with $F \neq G$.

a) For all $a \in A$, define the set $B_a = \{b \in \underline{B} : F(x_{a-1}) \leq G(y_{b-1}) \leq F(x_a)\}$. Note that, for all $a \in \underline{A}$, $B_a \neq \emptyset$ holds. Otherwise, there would exist a $b \in \underline{B}$ such that $G(y_{b-1}) < F(x_{a-1}) < F(x_a) < G(y_b)$, yielding $q_b > p_a$ in spite of $a \leftrightharpoons b$ and thus contradicting $F \leq_{disp}^{\wedge-disc} G$. With similar reasoning, $B_{\min A} \neq \emptyset$ follows (provided that the minimum exists). We now prove part a) by contradiction and therefore assume that $x_a \leq y_{b-1}$ or $y_b < x_a$ holds for all $b \in B_a$ and all $a \in A$. This is contradicted by case distinction.

Case 1: $x_a \leq y_{b-1}\ \forall (a,b) \in \bigcup_{a \in A} (\{a\} \times B_a)$

Let $a \in \underline{A}$. Obviously, $a \leftrightharpoons b$ holds for all $b \in B_a$, except if $G(y_{b-1}) = F(x_a)$. More precisely, if there exists a $b \in B$ such that $F(x_{a-1}) = G(y_{b-1})$, then $\{b \in B : a \leftrightharpoons b\} \subseteq B_a$. Otherwise, we have $\{b \in B : a \leftrightharpoons b\} \subseteq B_a \cup \{\min B_a - 1\}$. Overall, $x_a \leq y_b$ follows for all $(a,b) \in R(\leftrightharpoons) \cap (\underline{A} \times B)$. (This is because, for all $a \in \underline{A}$, we have $x_a \leq y_{b-1} < y_b$ for $b \in B_a$ and $x_a \leq y_{\min B_a - 1}$ by assumption.)

Now, let $a = \min A$ and assume that this minimum exists. Similarly as before, we have

$$\{b \in B : a \leftrightharpoons b\} \subseteq B_a \cup \{\min B_a - 1\} = B_a \cup \{\min B\}$$

and we can infer $x_a \leq y_b$ for all $(a,b) \in R(\leftrightharpoons) \cap (\{\min A\} \times B)$. Combined with the results for $a \in \underline{A}$, it follows that $x_a \leq y_b$ holds for all $(a,b) \in R(\leftrightharpoons)$. Furthermore, at least one of these inequalities is strict since equality for all these pairs would imply $F = G$ (see Lemma A.12 and Proposition 3.3a)). It follows

$$0 = \mathbb{E}[Y] - \mathbb{E}[X] = \sum_{(a,b)\in R(\leftrightharpoons)} r_{(a,b)}(y_b - x_a) > 0,$$

a contradiction.

Case 2: $y_b < x_a \ \forall (a,b) \in \bigcup_{a\in A}\{a\} \times B_a$

Let $a \in A$. Analogously to Case 1, it can be shown that $\{b \in B : a \leftrightharpoons b\} \subseteq B_a \cup \{\min B_a - 1\}$. If $b \in B_a$, $y_b < x_a$ holds by assumption; if $b = \min B_a - 1$, $y_b < y_{\min B_a} < x_a$ holds. Overall, we have $y_b < x_a$ for all $(a,b) \in R(\leftrightharpoons)$, yielding

$$0 = \mathbb{E}[X] - \mathbb{E}[Y] = \sum_{(a,b)\in R(\leftrightharpoons)} r_{(a,b)}(x_a - y_b) > 0, \tag{7.33}$$

a contradiction.

The remaining cases all consist of $x_a \leq y_{b-1}$ holding for some pairs $(a,b) \in \bigcup_{a\in A}\{a\} \times B_a$ and $y_b < x_a$ holding for others. For that, we order all these pairs from low to high, or, in other words, primarily by $a \in A$ and secondarily by $b \in B_a$, i.e.

$$\ldots, (a-1, \min B_{a-1}), \ldots, (a-1, \max B_{a-1}), (a, \min B_a), \ldots, (a, \max B_a), \ldots.$$

This gives us three possible kinds of successive pairs. The first one is $(a,b), (a, b+1)$, where $a \in A$ and $b, b+1 \in B_a$ (denoted by (P1)). Both the second and the third kind are of the form $(a, \max B_a), (a+1, \min B_{a+1})$ for an $a \in \overline{A}$. If $G(y_{\max B_a - 1}) \in F(D_F)$, then $\max B_a = \min B_{a+1}$ holds by definition of B_a. This gives the second kind of successive pairs, which is of the form $(a,b), (a+1, b)$, where $a \in \overline{A}$ and $G(y_{b-1}) = F(x_a)$ (denoted by (P2)). (Note that $G(y_{\max B_a - 1}) = F(x_{a-1})$ is not possible because $G(y_{\max B_a}) > F(x_a)$ then contradicts $q_{\max B_a} \leq p_a$ and therefore $a \leftrightharpoons \max B_a$, and $G(y_{\max B_a}) \leq F(x_a)$ then contradicts the maximality of $\max B_a$.) If $G(y_{\max B_a - 1}) \notin F(D_F)$, then $\max B_a = \min B_{a+1} - 1$ holds. This gives us the third kind of successive pairs, which is of the form $(a,b), (a+1, b+1)$, where $a \in \overline{A}$ and $b = \max B_a$ (denoted by (P3)).

For each of these kinds of successive pairs, $x_a \leq y_{b-1}$ can hold for the former and $y_b < x_a$ can hold for the latter or vice versa. Overall, this gives us six cases that remain to be considered.

Case 3: (P1) with $x_a \leq y_{b-1}$ and $y_{b+1} < x_a$

This directly gives $y_{b+1} < x_a \leq y_{b-1}$, a contradiction.

Case 4: (P1) with $y_b < x_a$ and $x_a \leq y_b$

These two statements directly contradict each other.

Case 5: (P2) with $x_a \leq y_{b-1}$ and $y_b < x_{a+1}$

Since $G(y_{b-1}) = F(x_a)$ holds for any successive pair of index pairs of the second kind, it follows $a+1 \leftrightharpoons_\wedge b$. Then,

$$x_{a+1} = x_a + (x_{a+1} - x_a) \leq y_{b-1} + (x_{a+1} - x_a) \leq y_{b-1} + (y_b - y_{b-1}) = y_b, \quad (7.34)$$

which contradicts the assumption $y_b < x_{a+1}$.

Case 6: (P2) with $y_b < x_a$ and $x_{a+1} \leq y_{b-1}$
This directly gives $x_{a+1} \leq y_{b-1} < y_b < x_a$, a contradiction.

Case 7: (P3) with $x_a \leq y_{b-1}$ and $y_{b+1} < x_{a+1}$
Since $G(y_{b-1}) < F(x_a) < G(y_b)$ holds for any successive pair of pairs of the third kind, it follows from Proposition 7.19a) that $a+1 \leftrightharpoons_\wedge b$. Then, (7.34) again holds, contradicting the assumption $y_b < y_{b+1} < x_{a+1}$.

Case 8: (P3) with $y_b < x_a$ and $x_{a+1} \leq y_b$
This directly gives $x_{a+1} \leq y_b < x_a$, a contradiction.

Now, we have considered all relevant cases and have thereby proved part a).

b) We prove this part by showing that (i) follows, if (ii) does not hold. This is done in several steps.

Step 1: If (ii) does not hold, part a) yields that there exists a pair $(a,b) \in A \times \overline{B}$ such that $F(x_a) = G(x_b)$ and $y_b < x_a$ or there exists a pair $(a,b) \in \overline{A} \times \overline{B}$ such that $F(x_a) = G(x_b)$ and $x_{a+1} \leq y_{b+1}$.

Step 2: In the first case from Step 1, it directly follows from $b \in \overline{B}$ and $F(x_a) = G(y_b)$ that $a \in \overline{A}$. Then, either $x_{a+1} \leq y_{b+1}$ holds (in which case we are finished) or $y_{b+1} < x_{a+1}$. Hence, assume $y_{b+1} < x_{a+1}$. We proceed by showing $b+1 \in \overline{B}$. We prove this by contradiction and therefore assume $G(y_{b+1}) = 1$, from which $1 = F(x_{a+1}) \geq G(y_{b+1})$ directly follows because of $a+1 \leftrightharpoons b+1$. Thus, $p_{a+1} = q_{b+1}$ holds, which yields $R(\leftrightharpoons) \cap (\{a+1\} \times B) = \{(a+1, b+1)\}$. We now seek to show $y_\beta < x_\alpha$ for all $(\alpha, \beta) \in R(\leftrightharpoons)$, which then implies $\mathbb{E}[Y] < \mathbb{E}[X]$ in the same way as in (7.33) and thereby concludes the proof of $b+1 \in \overline{B}$ by contradiction. Obviously, $y_\beta < x_a$ is true for all $\beta \in B$ such that $a \leftrightharpoons \beta$ because of $y_b < x_a$ and since all such β's are no larger than b. From Proposition 7.19a), it follows that there exists a $\beta_0 \leq b-1$ such that $a \leftrightharpoons_\wedge \beta_0 + 1$ with β_0 being the largest element of B, for which $a-1 \leftrightharpoons \beta_0$ holds. (This can be achieved by defining $\beta_0 = \min \mathrm{NN}_F^G(a) - 1$.) From this,

$$\begin{aligned} y_{\tilde{\beta}_0} < y_{\beta_0} &= y_{\beta_0+1} - (y_{\beta_0+1} - y_{\beta_0}) \leq y_{\beta_0+1} - (x_a - x_{a-1}) \\ &\leq y_b - (x_a - x_{a-1}) < x_a - (x_a - x_{a-1}) = x_{a-1} \end{aligned}$$

follows for all $\tilde{\beta}_0 \in B$ such that $a-1 \leftrightharpoons \tilde{\beta}_0$. This can be recursively continued for $a-2, a-3, \ldots$ as long as these indices are in A. Overall, this proves that $y_\beta < x_\alpha$

holds for all $(\alpha, \beta) \in R(\leftrightharpoons)$.

Step 3: Still assuming $y_{b+1} < x_{a+1}$, it follows from Step 2 that $b+1 \in \overline{B}$. Now, if $F(x_{a+1}) = 1$, along with $y_\beta < x_{a+1}$ for all $\beta > b+1$, $\mathbb{E}[Y] < \mathbb{E}[X]$ follows similarly as before, leading to a contradiction. Assuming $F(x_{a+1}) = 1$ along with the existence of a $\beta > b+1$ such that $y_{\beta-1} < x_{a+1} \leq y_\beta$ yields that $(a+1, \beta)$ satisfies condition (ii), which also poses a contradiction. Hence, $F(x_{a+1}) < 1$ and $a+1 \in \overline{A}$.

Step 4: Since (ii) is assumed to not hold, $y_{\beta+1} < x_{a+1}$ follows for all $\beta \in \overline{B}$ with $F(x_a) < G(y_\beta) < F(x_{a+1})$. Hence, there exists a $\beta_1 \in \overline{B}$ with $F(x_{a+1}) \leq G(y_{\beta_1}) < F(x_{a+2})$ and $y_{\beta_1} < x_{a+1} < x_{a+2}$. Now we can use our earlier procedure recursively in the following sense. If $G(y_{\beta_1}) > F(x_{a+1})$, $y_{\beta_1+1} < x_{a+2}$ follows since (ii) does not hold. This gives us the setting from the beginning of Step 4 with (a, β) replaced by $(a+1, \beta_1)$. Here, $a+2 \in \overline{A}$ is guaranteed since otherwise, $\mathbb{E}[Y] < \mathbb{E}[X]$ would follow. This can be repeated recursively as long as the strict inequality analogous to $G(y_{\beta_1}) > F(x_{a+1})$ is satisfied. The recursion has to stop at some point (in the sense that equality holds) since otherwise, $\mathbb{E}[Y] < \mathbb{E}[X]$ would follow.

If $G(y_{\beta_1}) = F(x_{a+1})$ holds (or an analogous inequality later in the recursion), we can reuse the proof from Step 2 onwards, replacing the pair (a, b) by $(a+1, \beta_1)$. With this new recursion, we end up at some point with a pair that satisfies condition (i) since, otherwise, the contradiction $\mathbb{E}[Y] < \mathbb{E}[X]$ would follow again.

Step 5: It remains to consider the second case in Step 1, so that there exists a pair $(a, b) \in \overline{A} \times \overline{B}$ such that $x_{a+1} \leq y_{b+1}$. Here, we can use the exact same procedure as in Steps 2–4, only going through the cdf's F and G in the opposite direction. The fact that the inequality $x_{a+1} \leq y_{b+1}$ is not a strict inequality does not change anything, because if equality held every time, we would end up with $F = G$, a contradiction. (In Steps 2–4, the contradiction $\mathbb{E}[Y] < \mathbb{E}[X]$ always follows from $y_\beta < x_\alpha$ holding for all $(\alpha, \beta) \in R(\leftrightharpoons)$. Hence, if the inequality is not strict, $\mathbb{E}[X] = \mathbb{E}[Y]$ can only occur if $y_\beta = x_\alpha$ holds for all $(\alpha, \beta) \in R(\leftrightharpoons)$. However, in that case, Lemma A.12 states that $F =_{st} G$ would follow, which is equivalent to $F = G$.)

$\blacksquare$

As mentioned before, the statement of Lemma 7.32 is the discrete analogue of an intersection of F and G in the continuous case, which is the transition from G being larger to F being larger. Note that part b) is just a more refined version of part a) and distinguishes between two kinds of intersection equivalents, both of which are depicted schematically in Figure 7.7.

If condition (i) is fulfilled, the images of the two (standardized) cdf's share a common element and the generalized intersection occurs as both cdf take that value. This means that (the more dispersed cdf) G is larger than F before said constant interval. After the constant

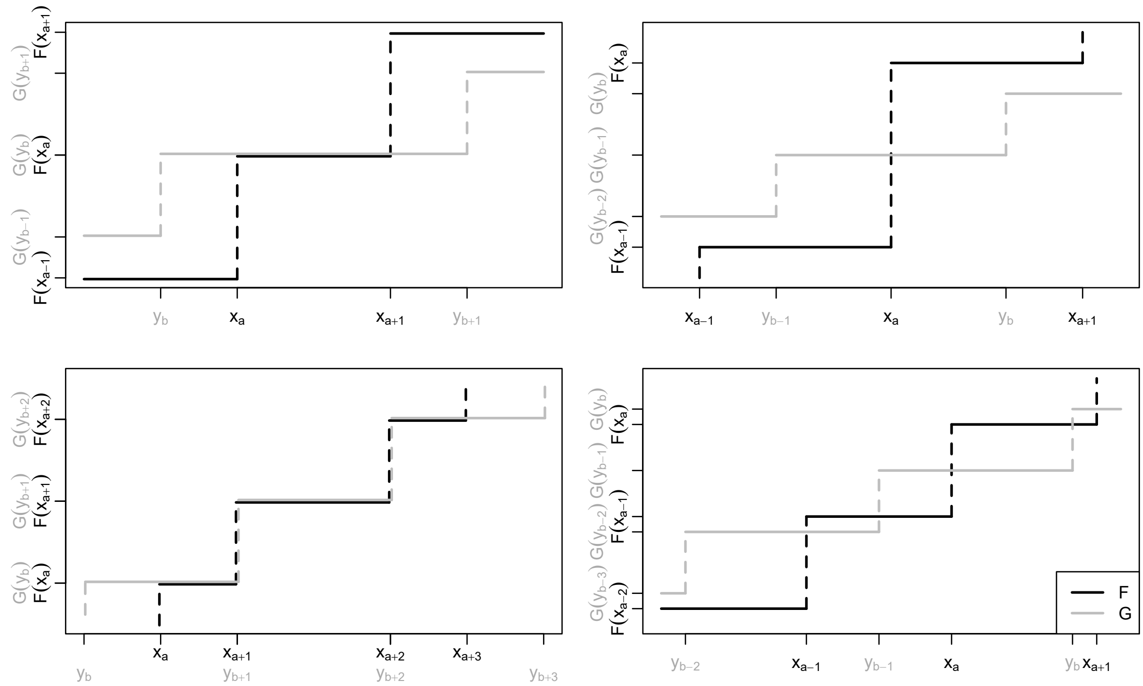

Figure 7.7.: Different variations of 'generalized intersections' between two standardized (w.r.t. the mean) cdf's F and G with $F \leq_{disp}^{\wedge-disc} G$, as specified in Lemma 7.32b). Left panels: condition (i). Right panels: condition (ii).

interval, G is either smaller than F right away (as in the upper left panel of Figure 7.7) or the two cdf's coincide for a while before G eventually becomes smaller (as in the lower left panel). The specific formulation of condition (i) (and also condition (ii)) that allows equality on the right side but not on the left is somewhat arbitrary in the sense that one could swap the requirements without invalidating the result. Although this potentially results in a different pair (a, b) being picked, that pair could be used in the same way going forward. Note that if equality was disallowed on both sides, Lemma 7.32 would no longer be true.

If condition (ii) is fulfilled, a jump of F and a constant interval of G form a cross (or a kind of degenerated cross if equality holds on the right side as discussed in the previous paragraph). G is either larger than F before that cross and smaller after (as exemplified in the upper right panel of Figure 7.7) or this kind of situation can occur repeatedly (as exemplified in the lower right panel). The latter situation is the main difficulty in the proof of the following theorem.

Theorem 7.33. *Let $F, G \in \mathcal{D}_0$ have finite means. Then, $F \leq_{disp}^{\wedge-disc} G$ implies $F \leq_{dil} G$.*

Proof. Assume without restriction that $\mathbb{E}[X] = \mathbb{E}[Y]$ and $F \neq G$. We proceed by showing $F \leq_{dil} G$ in cases (i) and (ii) of Lemma 7.32b), adopting the notation from there. According to the assumption $\mathbb{E}[X] = \mathbb{E}[Y]$, the characterization (2.12) of the dilation order and the identity

$$\mathbb{E}[(X-t)_+] = \mathbb{E}[X-t] + \mathbb{E}[(X-t)_-] = \mathbb{E}[X] - t + \mathbb{E}[(X-t)_-]$$

for all $t \in \mathbb{R}$, $F \leq_{dil} G$ is equivalent to

$$\mathbb{E}[(X-t)_+] \leq \mathbb{E}[(Y-t)_+] \quad \forall t \geq t_0 \quad \text{and} \quad \mathbb{E}[(X-t)_-] \leq \mathbb{E}[(Y-t)_-] \quad \forall t \leq t_0 \tag{7.35}$$

for any $t_0 \in \mathbb{R}$.

(i) Let (a,b) be the pair, the existence of which is guaranteed by (i) in Lemma 7.32b) and choose $t_0 = \frac{x_a + x_{a+1}}{2}$. The first inequality in (7.35) is now equivalent to

$$\sum_{(\alpha,\beta)\in R(\leftrightharpoons)} r_{(\alpha,\beta)}((y_\beta - t)_+ - (x_\alpha - t)_+) \geq 0. \tag{7.36}$$

for all $t \geq t_0$. The summand corresponding to any pair $(\alpha,\beta) \in R(\leftrightharpoons)$ with $\alpha \leq a$ is zero because $x_\alpha < t_0$ and $y_\beta < t_0$ hold in that case. So, let $\alpha \geq a+1$, also yielding $\beta \geq b+1$. For all $k > a+1$, there exists exactly one $\ell_k > b+1$ with $G(y_{\ell_k - 1}) \leq F(x_{k-1})$ and $k \leftrightharpoons_\wedge \ell_k$, according to Proposition 7.19a). ($\ell_k > b+1$ holds because $a+1$ is the only element of A, for which $a+1 \leftrightharpoons_\wedge b+1$ holds, since $F(x_a) = G(x_b)$.) Furthermore, $a+1 < k-1 < k$ implies $b+1 < \ell_{k-1} < \ell_k$ because of $G(y_{\ell_{k-1}-1}) \leq F(x_{k-2}) < G(y_{\ell_k - 1})$ (see Lemma 7.18a)); thus, the ℓ_k's are pairwise distinct. Hence, for all $\alpha > a+1$, there exists a $\beta_0 > b+1$ with $G(y_{\beta_0 - 1}) \leq F(x_{\alpha-1})$ and

$$x_\alpha - x_{a+1} = \sum_{k=a+2}^{\alpha} (x_k - x_{k-1}) \leq \sum_{k=a+2}^{\alpha} (y_{\ell_k} - y_{\ell_k - 1}) \leq \sum_{\ell=b+2}^{\beta_0} (y_\ell - y_{\ell-1}) = y_{\beta_0} - y_{b+1}.$$

Because of $G(y_{\beta_0 - 1}) \leq F(x_{\alpha-1})$, $\beta \geq \beta_0$ holds for all $\beta \in B$ such that $\alpha \leftrightharpoons \beta$. Overall, all such β satisfy

$$y_\beta \geq y_{\beta_0} = (y_{\beta_0} - y_{b+1}) + y_{b+1} \geq (x_\alpha - x_{a+1}) + y_{b+1} \geq (x_\alpha - x_{a+1}) + x_{a+1} = x_\alpha,$$

yielding $(y_\beta - t)_+ \geq (x_\alpha - t)_+$ for all $t \geq t_0$. Hence, (7.36) is true. The second inequality in (7.35) is shown completely analogous by extending the inequality $y_b < x_a$ down the cdf's F and G instead of extending the inequality $x_{a+1} \leq y_{b+1}$ up the cdf's.

(ii) Let $(a,b), (\alpha,\beta) \in A \times \underline{B}$ satisfy condition (ii) with $\alpha - 1 > a$, which yields $\beta - 1 > b$. This implies $(a, b-1), (\alpha, \beta-1) \in R(\leftrightharpoons)$ as well as $y_{b-1} < x_a \leq y_b < y_{\beta-1} < x_\alpha \leq y_\beta$, which yields $x_\alpha - x_a > y_{\beta-1} - y_b$. According to (7.20) in Part 1 of the proof of Theorem 7.30, $\alpha - a = (\beta - 1) - b + 1 = \beta - b$ follows. Part 2 of the same proof then implies

$$\begin{aligned} R(\leftrightharpoons_\wedge) \supseteq \{&(a+1,b),(a+1,b+1),(a+2,b+1),(a+2,b+2),\ldots \\ &\ldots,(\alpha-2,\beta-2),(\alpha-1,\beta-2),(\alpha-1,\beta-1),(\alpha,\beta-1)\} \end{aligned}$$

(see (7.21), (7.22), (7.25) and (7.26)). Obviously, $\alpha \leftrightharpoons \beta$ also holds. Thus, for all $k \in \{1,\ldots,\alpha-a\} = \{1,\ldots,\beta-b\}$, we have $a+k \leftrightharpoons_\wedge b+k-1$ and $a+k \leftrightharpoons_\wedge b+k$.

Let $k \in \{1, \dots, \alpha - a - 1\}$. Now $G(y_{b+k-1}) \in (F(x_{a+k-1}), F(x_{a+k}))$ follows from $a + k \leftrightharpoons b + k - 1$ and $a + k \leftrightharpoons b + k$. Furthermore,

$$\begin{aligned} x_{a+k} &= x_a + \sum_{j=1}^{k}(x_{a+j} - x_{a+j-1}) \leq y_b + \sum_{j=1}^{k}(x_{a+j} - x_{a+j-1}) \\ &\leq y_b + \sum_{j=1}^{k}(y_{b+j} - y_{b+j-1}) = y_{b+k} \end{aligned}$$

and

$$\begin{aligned} x_{a+k} &= x_\alpha - \sum_{j=k+1}^{\alpha-a}(x_{a+j} - x_{a+j-1}) > y_{\beta-1} - \sum_{j=k+1}^{\alpha-a}(x_{a+j} - x_{a+j-1}) \\ &\geq y_{\beta-1} - \sum_{j=k+1}^{\beta-b}(y_{b+j-1} - y_{b+j-2}) = y_{b+k-1}. \end{aligned}$$

Overall, the pair $(a + k, b + k)$ satisfies condition (ii) for all $k \in \{1, \dots, \alpha - a - 1\}$.

The entirety of the proof for case (ii) so far states that, if there are multiple pairs of indices satisfying condition (ii), they are of the form $\dots, (a, b), (a+1, b+1), \dots$. Assume now that this chain of pairs has a lower end, so that there exists a pair $(a, b) \in \underline{A} \times \underline{B}$ that satisfies condition (ii), but the pair $(a - 1, b - 1)$ does not. Additionally, let $(\alpha, \beta) \in R(\leftrightharpoons)$ be a lower pair than (a, b) in the sense that the interval measured by $r_{(\alpha,\beta)}$ is lower on the unit interval than the interval measured by $r_{(a,b)}$. We seek to prove $y_\beta < x_\alpha$ for all such pairs. If $x_a - x_\alpha \leq y_{b-1} - y_\beta$ holds, then $y_\beta < x_\alpha$ directly follows because of $y_{b-1} < x_a$. According to (7.20) in Part 1 of the proof of Theorem 7.30, $a - \alpha = b - \beta = (b - 1) - \beta + 1$ is a necessary condition for $x_a - x_\alpha > y_{b-1} - y_\beta$. In that case, Part 2 of the same proof implies that the values of F and G alternate in the sense of $F(x_{\alpha-1}) < G(y_{\beta-1}) < F(x_\alpha)$. If $y_\beta < x_\alpha$ already holds, we are done; if $y_\beta \geq x_\alpha$ holds along with $x_\alpha > y_{\beta-1}$, the pair (α, β) satisfies condition (ii), posing a contradiction. Therefore, assume $x_\alpha \leq y_{\beta-1}$. Again invoking (7.20), this is only possible if $a - \alpha = (b - 1) - (\beta - 1) + 1 = b - \beta + 1$, which contradicts $a - \alpha = b - \beta$. Overall, it follows that $y_\beta < x_\alpha$ holds for all pairs $(\alpha, \beta) \in R(\leftrightharpoons)$ lower than the chain of pairs satisfying condition (ii). Analogously, it can be shown that $x_\alpha \leq y_\beta$ holds for all pairs $(\alpha, \beta) \in R(\leftrightharpoons)$ higher than the chain of pairs satisfying condition (ii).

Due to the proven structure of this chain, we can find a unique indexing set $C \in \{\{0, \dots, n\} : n \in \mathbb{N}_0\} \cup \{\mathbb{N}_0, -\mathbb{N}_0, \mathbb{Z}\}$ and an initial pair $(a, b) \in A \times \underline{B}$ such that $((a + c, b + c))_{c \in C}$ is the sequence of all pairs satisfying condition (ii). We now define the following two mappings:

$$\tilde{m}_{pm} : C \to (0, 1), \quad c \mapsto \frac{G(y_{b+c-1}) - F(x_{a+c-1})}{F(x_{a+c}) - F(x_{a+c-1})},$$

$$\tilde{m}_{supp} : C \to (0,1], \quad c \mapsto \frac{x_{a+c} - y_{b+c-1}}{y_{b+c} - y_{b+c-1}}$$

Both mappings can be extended to $\mathbb{Z}$ by assigning the value 1 (to both), if the argument is smaller than the minimum of C, and assigning the value 0 (to both), if the argument is larger than the maximum of C. We now show that both mappings are decreasing on C. For that, let $c, c+1 \in C$ and define $d_F = \min\{F(x_{a+c}) - F(x_{a+c-1}), F(x_{a+c+1}) - F(x_{a+c})\}$ as well as $d_y = \min\{y_{b+c} - y_{b+c-1}, y_{b+c+1} - y_{b+c}\}$. Note that, because of

$$F(x_{a+c-1}) < G(y_{b+c-1}) < F(x_{a+c}) < G(y_{b+c}) < F(x_{a+c+1}),$$

$a+c \leftrightharpoons b+c, a+c+1 \leftrightharpoons b+c$ follows as well as $a+c+1 \leftrightharpoons_{\wedge} b+c, a+c+1 \leftrightharpoons_{\wedge} b+c+1$. Then,

$$\begin{aligned}
\tilde{m}_{pm}(c+1) - \tilde{m}_{pm}(c) &= \frac{G(y_{b+c}) - F(x_{a+c})}{F(x_{a+c+1}) - F(x_{a+c})} + \frac{F(x_{a+c}) - G(y_{b+c-1})}{F(x_{a+c}) - F(x_{a+c-1})} - 1 \\
&\leq \frac{G(y_{b+c}) - G(y_{b+c-1})}{d_F} - 1 \leq 0, \qquad (7.37) \\
\tilde{m}_{supp}(c+1) - \tilde{m}_{supp}(c) &= \frac{x_{a+c+1} - y_{b+c}}{y_{b+c+1} - y_{b+c}} + \frac{y_{b+c} - x_{a+c}}{y_{b+c} - y_{b+c-1}} - 1 \\
&\leq \frac{x_{a+c+1} - x_{a+c}}{d_y} - 1 \leq 0.
\end{aligned}$$

It follows that the mappings

$$\begin{aligned}
m_{pm} &: C \to (0,\infty), \quad c \mapsto \left(\tilde{m}_{pm}(c)^{-1} - 1\right)^{-1} = \frac{G(y_{b+c-1}) - F(x_{a+c-1})}{F(x_{a+c}) - G(y_{b+c-1})}, \\
m_{supp} &: C \to (0,\infty], \quad c \mapsto \left(\tilde{m}_{supp}(c)^{-1} - 1\right)^{-1} = \frac{x_{a+c} - y_{b+c-1}}{y_{b+c} - x_{a+c}}
\end{aligned}$$

are also decreasing. Now, define the mapping

$$m : C \to (0,\infty], \quad c \mapsto m_{pm}(c) \cdot m_{supp}(c) = \frac{(G(y_{b+c-1}) - F(x_{a+c-1}))(x_{a+c} - y_{b+c-1})}{(F(x_{a+c}) - G(y_{b+c-1}))(y_{b+c} - x_{a+c})}.$$

m can also be extended to the domain $\mathbb{Z}$ by assigning $m(c) = \infty$ if $c < \min C$ and $m(c) = 0$ if $c > \max C$. Furthermore, m inherits from m_{pm} and m_{supp} that it is decreasing. Define $c_0 = \min\{c \in \mathbb{Z} : m(c) \leq 1\}$. Since m is monotone, the existence of this minimum follows, if we can show that the range of m is neither a subset of $[0,1]$ nor a subset of $(1,\infty]$. To prove this, first assume that $m(c) \leq 1$ for all $c \in \mathbb{Z}$. It follows $\inf C = -\infty$, which implies $C \in \{-\mathbb{N}_0, \mathbb{Z}\}$. If $C = -\mathbb{N}_0$, $y_\beta \geq x_\alpha$ holds for all pairs $(\alpha, \beta) \in R(\leftrightharpoons)$ higher than (a,b). Hence, for $C \in \{-\mathbb{N}_0, \mathbb{Z}\}$,

$$\mathbb{E}[Y] - \mathbb{E}[X] \geq \sum_{c=\inf C}^{\sup C} r_{(a+c,b+c)}(y_{b+c} - x_{b+c}) + r_{(a+c,b+c-1)}(y_{b+c-1} - x_{a+c})$$

$$= \sum_{c=\inf C}^{\sup C} (F(x_{a+c}) - G(y_{b+c-1}))(y_{b+c} - x_{a+c}) \tag{7.38}$$

$$- (G(y_{b+c-1}) - F(x_{a+c-1}))(x_{a+c} - y_{b+c-1}).$$

The non-negativity of the c-th summand in (7.38) is equivalent to $m(c) \leq 1$, which was assumed for all $c \in C$. The assumption $\mathbb{E}[X] = \mathbb{E}[Y]$ now implies $m(c) = 1$ for all $c \in C$. Because of $\inf C = -\infty$, the sequence $(F(x_{a+c}) - F(x_{a+c-1}))_{c \in C}$ converges to zero as $c \to -\infty$. This means that there exists a $c \in C$ such that $F(x_{a+c-1}) - F(x_{a+c-2}) < F(x_{a+c}) - F(x_{a+c-1})$, meaning that the first inequality in (7.37) is strict. This contradicts that $\tilde{m}_{pm}$ is constant on C and therefore also that $m(c) = 1$ holds for all $c \in C$. Overall, $m(c) \leq 1$ cannot hold for all $c \in \mathbb{Z}$. Since assuming $m(c) > 1$ for all $c \in \mathbb{Z}$ analogously leads to a contradiction, we have proved that the minimum in the definition of c_0 exists.

Define $(a_0, b_0) = (a + c_0, b + c_0)$. We now prove $F \leq_{dil} G$ by showing (7.35), which is equivalent to

$$\sum_{(\alpha,\beta) \in R(\leftrightharpoons)} r_{(\alpha,\beta)}((y_\beta - t)_+ - (x_\alpha - t)_+) \geq 0 \quad \forall t \geq t_0 \quad \text{and} \tag{7.39}$$

$$\sum_{(\alpha,\beta) \in R(\leftrightharpoons)} r_{(\alpha,\beta)}((y_\beta - t)_- - (x_\alpha - t)_-) \geq 0 \quad \forall t \leq t_0. \tag{7.40}$$

We proceed via case distinction.

Case 1: $c_0 - 1, c_0 \in C$

In this case, we choose $t_0 = y_{b_0-1}$. It holds that

$$y_{b_0-2} < x_{a_0-1} \leq y_{b_0-1} < x_{a_0} \leq y_{b_0},$$
$$F(x_{a_0-2}) < G(y_{b_0-2}) < F(x_{a_0-1}) < G(y_{b_0-1}) < F(x_{a_0}).$$

Hence, $(a_0, b_0 - 1)$ is the lowest pair in $R(\leftrightharpoons)$ such that the corresponding summand in (7.39) is not zero. The next pairs in $R(\leftrightharpoons)$ are then $(a_0, b_0), (a_0 + 1, b_0), (a_0 + 1, b_0 + 1), \ldots$ as long as the pairs of the form $(a_0 + k, b_0 + k), k \in \mathbb{N}$, still satisfy condition (ii). If $\max C$ exists, $y_\beta \geq x_\alpha$ holds for any pair $(x_\alpha, y_\beta) \in R(\leftrightharpoons)$ higher than $(a + \max C, b + \max C)$ (as was shown earlier). Therefore,

$$0 \leq \sum_{c=c_0}^{\max C} \Big[r_{(a+c,b+c)}((y_{b+c} - t)_+ - (x_{a+c} - t)_+) \tag{7.41}$$

$$+ r_{(a+c,b+c-1)}((y_{b+c-1} - t)_+ - (x_{a+c} - t)_+) \Big]$$

$$= \sum_{c=c_0}^{\max C} \Big[(F(x_{a+c}) - G(y_{b+c-1}))\,((y_{b+c} - t)_+ - (x_{a+c} - t)_+) \tag{7.42}$$

$$- \left(G(y_{b+c-1}) - F(x_{a+c-1})\right)\left((x_{a+c} - t)_+ - (y_{b+c-1} - t)_+\right)\Big]$$

for all $t \geq t_0$ is sufficient to show (7.39). We now consider each summand in (7.42) separately. First, in the case $t < y_{b+c-1}$, the summand is equal to

$$\left(F(x_{a+c}) - G(y_{b+c-1})\right)(y_{b+c} - x_{a+c}) - \left(G(y_{b+c-1}) - F(x_{a+c-1})\right)(x_{a+c} - y_{b+c-1})\,. \tag{7.43}$$

The non-negativity of this is equivalent to $m(c) \leq 1$, which is true since $c \geq c_0$. In the case $y_{b+c-1} \leq t \leq x_{a+c}$, the summand is equal to

$$\left(F(x_{a+c}) - G(y_{b+c-1})\right)(y_{b+c} - x_{a+c}) - \left(G(y_{b+c-1}) - F(x_{a+c-1})\right)(x_{a+c} - t)\,,$$

which is no smaller than (7.43) and therefore also non-negative. In the case $x_{a+c} \leq t \leq y_{b+c}$, the summand is equal to $\left(F(x_{a+c}) - G(y_{b+c-1})\right)(y_{b+c} - t)$, which is non-negative because both factors are. Finally, in the case $t > y_{b+c}$, the summand is zero. Overall, inequality (7.41) is satisfied, leaving (7.40) to be shown. For that, proceeding similarly to before, it is sufficient to show

$$\begin{aligned} 0 \leq \sum_{c=\min C}^{c_0-1} \Big[&\left(F(x_{a+c}) - G(y_{b+c-1})\right)\left((y_{b+c} - t)_- - (x_{a+c} - t)_-\right) \\ &- \left(G(y_{b+c-1}) - F(x_{a+c-1})\right)\left((x_{a+c} - t)_- - (y_{b+c-1} - t)_-\right)\Big] \end{aligned} \tag{7.44}$$

for all $t \leq t_0$. We again only consider the c-th summand, beginning with the case $t > y_{b+c}$, in which it is equal to negative (7.43). The non-negativity of that term is equivalent to $m(c) \geq 1$, which is true since $c < c_0$. The non-negativity of the c-th summand in the remaining cases follows in a similar fashion. Hence, both (7.39) and (7.40) are satisfied, yielding $F \leq_{dil} G$.

Case 2: $c_0 - 1 = \max C$

In this case, $(a_0 - 1, b_0 - 1) \in A \times \underline{B}$ obviously holds and we choose $t_0 = x_{a_0-1}$. Because of $y_{b_0-2} < x_{a_0-1} \leq y_{b_0-1}$, only $(a_0 - 1, b_0 - 1)$ and pairs $(\alpha, \beta) \in R(\leftrightharpoons)$ higher than $(a_0 - 1, b_0 - 1)$ have non-zero summands in (7.39). However, since $x_\alpha \leq y_\beta$ and, consequently, $(x_\alpha - t)_+ \leq (y_\beta - t)_+$ holds for all those pairs, (7.39) is true.

Because of $y_{b_0-2} < x_{a_0-1} \leq y_{b_0-1}$, the highest pair in $R(\leftrightharpoons)$ with a non-zero summand in (7.40) is $(a_0 - 1, b_0 - 2)$. As before, all pairs $(\alpha, \beta) \in R(\leftrightharpoons)$ below the chain of pairs in $R(\leftrightharpoons)$ satisfying condition (ii) can be disregarded as they satisfy $y_\beta < x_\alpha$. It follows that (7.44) is a sufficient condition for (7.40) and it can be shown analogously to Case 1 as the chain indexed by C has the same properties

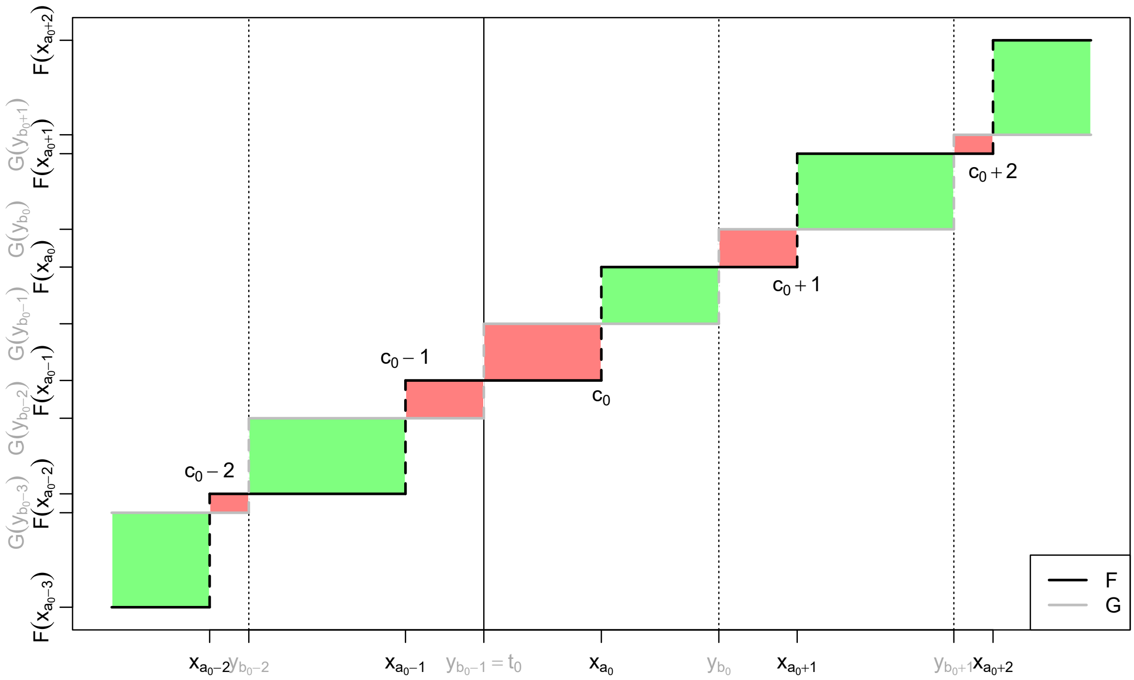

Figure 7.8.: Exemplary situation in the proof of Theorem 7.33, part (ii), Case 1. (F and G are assumed to be standardized w.r.t. the mean.)

from $c_0 - 1$ downwards. (In fact, condition (7.44) can even be weakened since it includes the non-positive summand associated with the pair $(a_0 - 1, b_0 - 1)$, which is not relevant here.)

Case 3: $c_0 = \min C$

For this case, we can proceed analogously to Case 2 after choosing $t_0 = x_{a_0}$ and by flipping the procedure upside-down.

▨

Corollary 7.34. *Let $F, G \in \mathcal{D}_0$ have finite means. Then, $F \leq_{disp}^{\vee-disc} G$ implies $F \leq_{dil} G$.*

The crucial inequalities (7.35) the proof of Theorem 7.33 are easily shown to be satisfied, if G is larger than (or equal to) F for $t \leq t_0$ and G is smaller than F for $t \geq t_0$. For all situations depicted in Figure 7.7 except for the lower right panel, t_0 can be chosen in such a way that this holds. In particular, this is always possible in case (i), making the proof in this case rather easy.

The only situation left to be considered is the one shown in the lower right panel of Figure 7.7, where F and G intersect multiple times. The structure of those intersections can be further narrowed down before then being divided in three cases in the proof. The line of reasoning in Case 1 is illustrated in Figure 7.8. The main focus lies on the ratio between the two areas between each pair of vertical lines. If the left area is divided by the right area, the

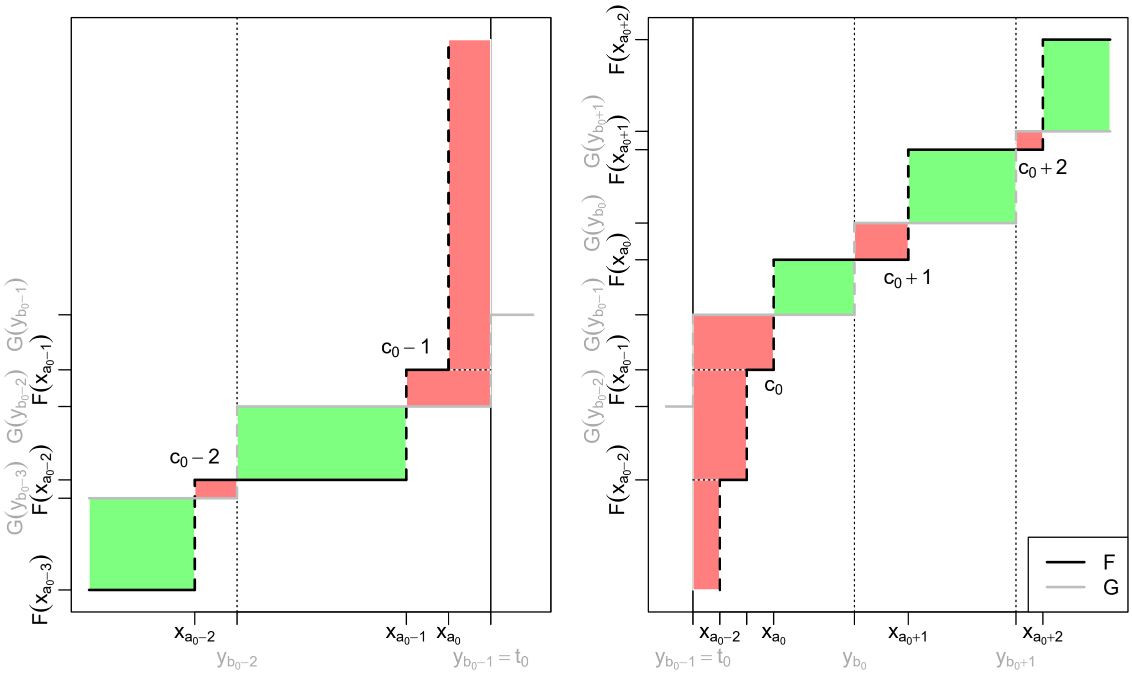

Figure 7.9.: Exemplary situation in the proof of Theorem 7.33, part (ii), Cases 2 and 3, if $t_0 = y_{b_0-1}$ was chosen instead of $t_0 = x_{a_0-1}$ and $t_0 = x_{a_0}$. (F and G are assumed to be standardized w.r.t. the mean.)

resulting ratio decreases as the corresponding index $(c_0-2, c_0-1, c_0, \ldots)$ increases. (In the proof, the ratio for an index c is denoted by $m(c)$.) Now, t_0 is chosen such that this ratio is smaller than 1 on the left side of t_0 and larger than or equal to 1 on the right side of t_0. In Figure 7.8, the smaller area of each pair is filled in red and the larger one is filled in green. As noted in (7.42), for all $t \geq t_0$, the sum of the green areas on the right side of t minus the sum of the red areas on the right side of t is required to be non-negative. For $t = t_0$, this is obviously true since the ratio between the green and the red areas is no smaller than 1 for each index. If t is increased from t_0 towards ∞, the red areas are always reduced before the corresponding green areas are. Thus, the non-negativity of (7.42) stays intact. The situation on the left side of t_0, so for $t \leq t_0$, is analogous.

This begs the question why Cases 2 and 3 in part (ii) of the proof of Theorem 7.33 need to be considered separately and cannot be included in the line of reasoning illustrated by Figure 7.8. This is illustrated in Figure 7.9, where the left panel is exemplary for Case 2 (as the chain of intersections of F and G ends at index c_0, which is chosen as before) and the right panel is exemplary for Case 3 (as the chain of intersections of F and G begins at index c_0). In Case 3, if $t_0 = y_{b_0-1}$ is chosen as before, it is possible for F to exhibit more steep jumps on the right side of t_0, thus expanding the corresponding red area to be larger than the corresponding green area. This is possible because the chain of intersections of F and G ends at index c_0 and does not extend any further below that index. The problem can be solved by choosing $t_0 = x_{a_0}$ instead. On the left side of t_0, F is then smaller than G; on the right side of t_0, there is an additional green area before the first full pair of red and green areas, once again ensuring the non-negativity required in (7.42). The line of reasoning in Case 2 is analogous, just going down instead of up.

7.4. Compatibility with Popular Dispersion Measures

In the literature, applying popular dispersion measures to discrete distributions is commonplace. Even though dispersion was measured long before an axiomatic framework was created (see, e.g., Kourkoulos and Tzanakis, 2010), such a widely used concept should definitely not lack a rigorous foundation.

Throughout Chapter 3, a number of dispersion measures are introduced in a continuous setting, based on the order-based approach given in Definition 3.1b). Part of this definition is that a dispersion measure preserves the dispersive order $\leq_{disp}$. This order is chosen for this role because of its basic interpretation of dispersion that does not favour certain kinds of measures. The discrete dispersive orders $\leq_{disp}^{\wedge-disc}$ and $\leq_{disp}^{\vee-disc}$ are constructed in such a way that they are also the strongest reasonable dispersion orders and do not favour any specific type of dispersion measure. Furthermore, the original and discrete dispersive orders compare two distribution in a pointwise way while $\leq_{dil}$, as an example of another dispersion measure, only considers means. This allows for probability mass to be smeared and for deviations in the wrong direction to be compensated. Hence, the orders $\leq_{disp}^{\wedge-disc}$ and $\leq_{disp}^{\vee-disc}$ are the canonical choice as basic tools for the assessment of discrete dispersion measures. This yields the following definition.

Definition 7.35. Let $\mathcal{Q} \subseteq \mathcal{D}_0$. Then, a mapping $\tau : \mathcal{Q} \to [0, \infty]$ is said to be a *discrete dispersion measure*, if

(DD1) $\tau(aX + b) = |a| \cdot \tau(X)$ for all $a, b \in \mathbb{R}$ and $X \sim \mathcal{Q}$,

(DD2) $\tau(F) \leq \tau(G)$ for all $F, G \in \mathcal{Q}$ with $F \leq_{disp}^{\wedge-disc} G$.

Note that any discrete dispersion measure according to Definition 7.35 is also a discrete dispersion measure with respect to $\leq_{disp}^{\vee-disc}$ as that order is stronger than $\leq_{disp}^{\wedge-disc}$. Next, we define a number of potential discrete dispersion measures, all of which are also introduced in Chapter 3 under different names.

Definition 7.36. For $F \in \mathcal{P}$, let $X, X' \sim F$ be independent and let $e_X : (0, 1) \to \mathbb{R}$ be the corresponding expectile function in the case $F \in \mathcal{L}^1$.

a) The mapping

$$\mathrm{SD} : \mathcal{L}^2 \to [0, \infty), \quad F \mapsto \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}$$

is said to be the *standard deviation* of the corresponding distribution. On $\mathcal{P}_I$, it coincides with τ_M (see Theorem 3.6b)).

b) The mapping
$$\mathrm{GMD} : \mathcal{L}^1 \to [0, \infty), \quad F \mapsto \mathbb{E}[|X - X'|]$$
is said to be the *Gini mean difference* of the corresponding distribution. On $\mathcal{P}_I$, it coincides with $2\tau_{LM}$ (see Theorem 3.7b)).

c) The mapping
$$\mathrm{MAD} : \mathcal{L}^1 \to [0, \infty), \quad F \mapsto \mathbb{E}[|X - \mathbb{E}[X]|]$$
is said to be the *mean absolute deviation from the mean* of the corresponding distribution. On $\mathcal{P}_I$, it coincides with τ_{EL} (see Corollary 3.25).

d) The mapping
$$\mathrm{MDMAD} : \mathcal{L}^1 \to [0, \infty), \quad F \mapsto \mathbb{E}[|X - F^{-1}(\tfrac{1}{2})|]$$
is said to be the *mean absolute deviation from the median* of the corresponding distribution. On $\mathcal{P}_I$, it coincides with τ_{IQ} (see (3.4)).

e) For $0 < \alpha < \beta < 1$, the mapping
$$\mathrm{IQR}(\alpha, \beta) : \mathcal{P} \to [0, \infty), \quad F \mapsto F^{-1}(\beta) - F^{-1}(\alpha)$$
is said to be the (α, β)-*interquantile range* of the corresponding distribution. For $\alpha \in (0, \frac{1}{2})$, $\beta = 1 - \alpha$ and on $\mathcal{P}_I$, it coincides with τ_Q^α (see Theorem 3.8b)). $\mathrm{IQR} = \mathrm{IQR}(\frac{1}{4}, \frac{3}{4})$ is called the *interquartile range*.

f) For $0 < \alpha < \beta < 1$, the mapping
$$\mathrm{IER}(\alpha, \beta) : \mathcal{L}^1 \to [0, \infty), \quad F \mapsto e_X(\beta) - e_X(\alpha)$$
is said to be the (α, β)-*interexpectile range* of the corresponding distribution. For $\alpha \in (0, \frac{1}{2})$, $\beta = 1 - \alpha$ and on $\mathcal{P}_I$, it coincides with τ_E^α (see Corollary 3.24b)). Furthermore, we define $\mathrm{IER} = \mathrm{IER}(\frac{1}{4}, \frac{3}{4})$.

For the remainder of Section 7.4, we consider the restrictions of these measures to $\mathcal{D}_0$. Before considering the crucial property (DD2) for the above mappings, we first briefly discuss whether they satisfy (DD1). While the requirement in (DD1) coincides with that in (D1), the different underlying sets of distributions make a difference for a number of mappings. This is made more precise by the following result. The proof borrows an idea from the proof of Theorem 2 in Eberl and Klar (2019, pp. 268–269).

Proposition 7.37. *Let* $\alpha \in (0, \frac{1}{2})$.

a) The mappings SD*,* GMD*,* MAD *and* $\mathrm{IER}(\alpha, 1 - \alpha)$ *all satisfy property (DD1).*

b) *On the set of all cdf's* $F \in \mathcal{D}_0$ *satisfying* $F(t) > \frac{1}{2}$ *for all* $t > F^{-1}(\frac{1}{2})$*, the mapping* MDMAD *satisfies property (DD1).*

c) *On the set of all cdf's* $F \in \mathcal{D}_0$ *satisfying* $F(t) > p$ *for all* $t > F^{-1}(p)$ *for* $p \in \{\alpha, 1-\alpha\}$*, the mapping* $\mathrm{IQR}(\alpha, 1-\alpha)$ *satisfies property (DD1).*

Proof. First, the necessity of the requirement $\beta = 1-\alpha$ for the interquantile and interexpectile ranges is obvious for reasons of symmetry. Part a) now follows from the linearity of the expected value and from Proposition 2.22a), b), e). If τ denotes either mapping from parts b) and c), $\tau(aX+b) = a \cdot \tau(X)$ for all $a > 0, b \in \mathbb{R}$ follows from the linearity of the quantile function.

This leaves $\tau(-X) = \tau(X)$ to be shown for the mappings in parts b) and c). This follows from the fact that, for $p \in (0,1)$, $F(t) > p$ for all $t > F^{-1}(p)$ implies $H_{-X}^{-1}(p) = -F^{-1}(1-p)$ (see Eberl and Klar, 2019, p. 269). ▨

The additional requirements in parts b) and c) of Proposition 7.37 yield that $\inf\{t \in \mathbb{R} : F(t) \geq p\} = \sup\{t \in \mathbb{R} : F(t) \leq p\}$ holds for the corresponding $p \in \{\alpha, \frac{1}{2}, 1-\alpha\}$. This means that all possible definitions of the p-quantile Q_F^p coincide. The additional requirements would not be necessary, if the p-quantile was instead defined by

$$\tfrac{1}{2}\big(\inf\{t \in \mathbb{R} : F(t) \geq p\} + \sup\{t \in \mathbb{R} : F(t) \leq p\}\big),$$

which is often done for empirical quantiles.

The fact that most of the mappings from Definition 7.35 also satisfy property (DD2) follows directly from Theorems 7.30 and 7.33.

Corollary 7.38. *The mappings* SD*,* MAD *and* GMD *all satisfy property (DD2).*

For SD and MAD, this follows from Theorem 7.33 since both $t \mapsto t^2$ and $t \mapsto |t|$ are convex functions on the real numbers. For GMD, (DD2) follows from Theorem 7.30 since the expected value is a measure of (central) location, which preserves the usual stochastic order $\leq_{st}$. We can also use Theorem 7.33 to show that GMD satisfies (DD2) since $F \leq_{dil} G$ implies $\mathrm{GMD}(F) \leq \mathrm{GMD}(G)$, as proved in Ramos and Sordo (2003, p. 126, Thm. 2.2) and pointed out by Sordo et al. (2016, p. 65). Conversely, the fact that SD satisfies (DD2) also follows from Theorem 7.30 because of

$$\mathrm{SD}(F)^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \tfrac{1}{2}\left(\mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X'] + \mathbb{E}[X'^2]\right) = \tfrac{1}{2}\mathbb{E}[(X-X')^2]$$

for $X, X' \sim F$ independent.

It can be shown in a similar way that $\mathrm{IER}(\alpha, \beta)$ satisfies property (DD2) for $0 < \alpha < \frac{1}{2} < \beta < 1$, which includes all cases relevant for applications. This again holds due to Theorem 7.33, combined with the fact that $F \leq_{dil} G$ implies $\mathrm{IER}(\alpha,\beta)(F) \leq \mathrm{IER}(\alpha,\beta)(G)$ for all

$0 < \alpha < \frac{1}{2} < \beta < 1$. This implication was first shown by Bellini (2012, p. 2020, Thm. 3(b)); a more elementary proof that also includes the reverse implication is given in Theorem 3.23 and Corollary 3.24 of this thesis. The assumptions can be weakened to include all distributions in $\mathcal{D}_0 \cap \mathcal{L}^1$ without changing the proof.

Corollary 7.39. *If $0 < \alpha < \frac{1}{2} < \beta < 1$, the mapping* $\mathrm{IER}(\alpha, \beta)$ *satisfies property (DD2).*

It remains to be determined whether the mappings $\mathrm{IQR}(\alpha, \beta)$ and MDMAD satisfy (DD2). Both mappings are based on quantiles, which are well-known to not be as useful for discrete distributions as they are for continuous distributions. This is partly due to the fact that they are not unique in the discrete case, which also proved to be problematic for (DD1). Furthermore, quantiles only evaluate a distribution in a very local sense, which also explains their popularity in robust measures and estimators. However, for discrete distributions, where the probability mass in very sparse, this leads to a lack of information that is conveyed by single evaluations of quantile functions. In accordance with these observations, the interquantile range $\mathrm{IQR}(\alpha, \beta)$ does generally not preserve the discrete dispersive orders, as noted in the following result.

Theorem 7.40. *For all choices $0 < \alpha < \beta < 1$, the mapping* $\mathrm{IQR}(\alpha, \beta)$ *does not satisfy property (DD2).*

Proof. We prove this by constructing a counterexample for arbitrary values of α and β. First, define $n = \left\lfloor \frac{\beta-\alpha}{\min(\alpha, 1-\beta)} \right\rfloor + 1 \in \mathbb{N}$ and choose $\delta \in \left(\frac{\beta-\alpha}{n}, \min(\alpha, 1-\beta) \right) \subseteq (0,1)$. The latter interval is not empty since

$$n > \frac{\beta - \alpha}{\min(\alpha, 1-\beta)} \Leftrightarrow \min(\alpha, 1-\beta) > \frac{\beta - \alpha}{n}$$

holds. Note that, by definition, $0 \leq (n-1)\delta < \beta - \alpha < n\delta < 1$ is true. The last inequality specifically follows from

$$\begin{aligned} n\delta &< \left(\left\lfloor \frac{\beta - \alpha}{\min(\alpha, 1-\beta)} \right\rfloor + 1 \right) \cdot \min(\alpha, 1-\beta) \\ &< (\beta - \alpha) + \min(\alpha, 1-\beta) = 1 - \max(\alpha, 1-\beta) < 1. \end{aligned}$$

Further, let $\varepsilon_p = \frac{(\beta-\alpha)-(n-1)\delta}{2} > 0$, $\varepsilon_q = \frac{n\delta - (\beta-\alpha)}{2} > 0$. Now we define $F, G \in \mathcal{D}_0$ by $F \mathrel{\hat{=}} (A, (x_j, p_j)_{j \in A})$ and $G \mathrel{\hat{=}} (B, (y_j, q_j)_{j \in B})$, where $A = \{1, \ldots, n+1\}$ with $x_a = a$ for all $a \in A$ and $B = \{1, \ldots, n+2\}$ with $y_b = b$ for all $b \in B$. Furthermore, let

$$\begin{aligned} p_1 &= \alpha + \varepsilon_p, \quad p_j = \delta \ \forall j \in \{2, \ldots, n\}, \qquad p_{n+1} = (1-\beta) + \varepsilon_p, \\ q_1 &= \alpha - \varepsilon_q, \quad q_j = \delta \ \forall j \in \{2, \ldots, n+1\}, \quad q_{n+2} = (1-\beta) - \varepsilon_q. \end{aligned}$$

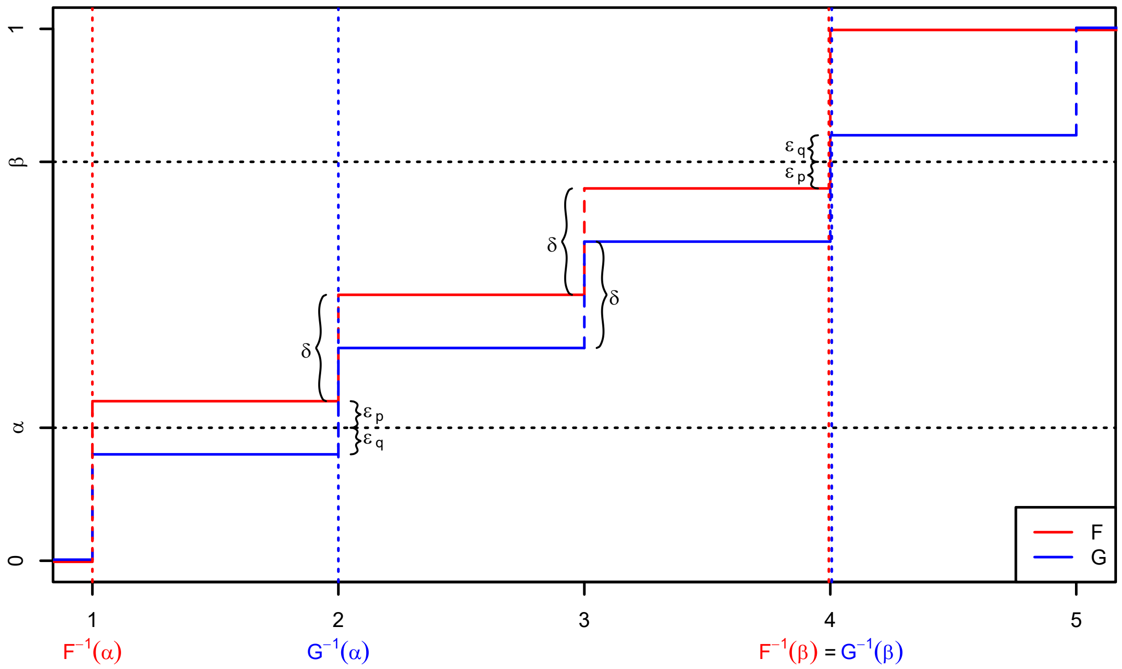

Figure 7.10.: Illustration of the counterexample in the proof of Theorem 7.40 with $\alpha = \frac{1}{4}$, $\beta = \frac{3}{4}$, $n = 3$ and $\delta = \frac{1}{5} \in (\frac{1}{6}, \frac{1}{4})$. Hence, $\varepsilon_p = \varepsilon_q = \frac{1}{20}$.

For all $j \in \{1, \dots, n\}$, it holds that

$$\begin{aligned} G(y_j) &= \alpha - \varepsilon_q + (j-1)\delta = \frac{\alpha+\beta}{2} + \delta\left(j - 1 - \frac{n}{2}\right) \\ &< \frac{\alpha+\beta}{2} + \delta\left(j - \frac{n}{2} - \frac{1}{2}\right) = \alpha + \varepsilon_p + (j-1)\delta = F(x_j) \\ &< \frac{\alpha+\beta}{2} + \delta\left(j - \frac{n}{2}\right) = \alpha + \varepsilon_p + j\delta = G(y_{j+1}), \end{aligned}$$

yielding $R(\leftrightharpoons) = \bigcup_{j=1}^{n+1}\{(j,j),(j,j+1)\}$. Hence, $F \leq_{D-pm}^{disc} G$. Since F and G are both lattice distributions with the same defining distance, it follows that $F \leq_{disp}^{\wedge-disc} G$. However,

$$\text{IQR}(\alpha,\beta)(F) = (n+1) - 1 = n > n - 1 = (n+1) - 2 = \text{IQR}(\alpha,\beta)(G).$$

▨

An illustration of the counterexample in the proof of Theorem 7.40 is given in Figure 7.10. The statement of Theorem 7.40 also holds, if we replace $\leq_{disp}^{\wedge-disc}$ by $\leq_{disp}^{\vee-disc}$ or $\leq_{disp}^{disc}$ in the definition of (DD2). This is due to the fact that the distributions used in the proof of Theorem 7.40 are lattice distributions, for which all discrete dispersive orders are equivalent (see Corollary 7.23a)).

Theorem 7.40 implies that the interquantile range is not fit to be used as a dispersion measure for discrete distributions. This statement is similar to Bellini et al. (2018b, p. 1852) suggesting that the interquantile range is not a 'true measure of variability' because it does not preserve the dilation order. However, as noted in Theorem 3.8b), the interquantile range indeed measures dispersion in a meaningful way for distributions in $\mathcal{P}_I$. The requirement that dispersion measures should preserve the dilation order is simply too strong. Neither the interquantile range nor the mean absolute deviation from the median could be proved to meet this requirement.

Concerning the latter mapping MDMAD, it is shown in Section 7.1 (around Figure 7.6) that the mapping $F \mapsto \nu(|X - F^{-1}(\frac{1}{2})|)$, where ν is an arbitrary measure of central location, generally does not preserve the order $\leq_{disp}^{\wedge-disc}$. Hence, the desired result is not a simple corollary as for some of the other dispersion measure contenders. However, the implication still holds, as shown in the following.

Theorem 7.41. *The mapping* MDMAD *satisfies property (DD2).*

Proof. Without restriction, let $F^{-1}(\frac{1}{2}) = G^{-1}(\frac{1}{2}) = 0$. Furthermore, let $a_0 \in A$ and $b_0 \in B$ be the unique indices that satisfy $x_{a_0} = F^{-1}(\frac{1}{2}) = 0 = G^{-1}(\frac{1}{2}) = y_{b_0}$. Because of $F(x_{a_0-1}) < \frac{1}{2} \leq G(y_{b_0})$ and $G(y_{b_0-1}) < \frac{1}{2} \leq F(x_{a_0})$, $a_0 \leftrightharpoons b_0$ holds. Hence, $F(x_{a_0}) - F(x_{a_0-1}) = p_{a_0} \geq q_{b_0} = G(y_{b_0}) - G(y_{b_0-1})$ and it follows that $G(y_{b_0}) \leq F(x_{a_0})$ or $G(y_{b_0-1}) \geq F(x_{a_0-1})$.

We begin by assuming $G(y_{b_0}) \leq F(x_{a_0})$ and show that, consequently, $y_\beta \geq x_\alpha$ holds for all pairs $(\alpha, \beta) \in R(\leftrightharpoons)$ higher than (a_0, b_0). Choosing $G(y_{b_k-1}) = \min \mathrm{NN}_F^G(a_0 + k)$ for all $k \in \mathbb{N}$ such that $a_0 + k \in A$ gives pairwise distinct indices $b_k \in \underline{B}$ such that $a_0 + k \leftrightharpoons_\wedge b_k$. This is guaranteed by Lemma 7.18. Note that, for all $k \in \mathbb{N}$, b_k is also the smallest index in B such that $a_0 + k \leftrightharpoons b_k$. Note also that $b_0 < b_1$. Hence, for all pairs $(\alpha, \beta) \in R(\leftrightharpoons)$ higher than (or equal to) (a_0, b_0), there exists a $k \in \mathbb{N}_0$ such that

$$
\begin{aligned}
x_\alpha = x_{a_0+k} &= x_{a_0} + \sum_{j=1}^{k}(x_{a_0+j} - x_{a_0+j-1}) = \sum_{j=1}^{k}(x_{a_0+j} - x_{a_0+j-1}) \\
&\leq \sum_{j=1}^{k}(y_{b_j} - y_{b_{j-1}}) = y_{b_0} + \sum_{j=1}^{k}(y_{b_j} - y_{b_{j-1}}) \leq y_{b_k} \leq y_\beta.
\end{aligned}
$$

Analogously, the assumption $G(y_{b_0-1}) \geq F(x_{a_0-1})$ yields $y_\beta \leq x_\alpha$ holds for all pairs $(\alpha, \beta) \in R(\leftrightharpoons)$ lower than (a_0, b_0). With this in mind, we prove the assertion via case distinction.

Case 1: $F(x_{a_0-1}) \leq G(y_{b_0-1}) < G(y_{b_0}) \leq F(x_{a_0})$

It holds that

$$
\begin{aligned}
\mathrm{MDMAD}(G) - \mathrm{MDMAD}(F) &= \sum_{(a,b)\in R(\leftrightharpoons)} r_{(a,b)}(|y_b - y_{b_0}| - |x_a - x_{a_0}|) \\
&= \sum_{(a,b)\in R(\leftrightharpoons)} r_{(a,b)}(|y_b| - |x_a|)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{\substack{(a,b)\in R(\leftrightharpoons)\\ \text{lower than}(a_0,b_0)}} r_{(a,b)}(x_a - y_b)\\
&\quad + \sum_{\substack{(a,b)\in R(\leftrightharpoons)\\ \text{higher than}(a_0,b_0)}} r_{(a,b)}(y_b - x_a). \qquad (7.45)
\end{aligned}
$$

Since $x_a \geq y_b$ holds for all pairs $(a,b) \in R(\leftrightharpoons)$ lower than (a_0, b_0) and $y_b \geq x_a$ holds for all pairs $(a,b) \in R(\leftrightharpoons)$ higher than (a_0, b_0), (7.45) is non-negative. Thus, the proof in this case is completed.

Case 2: $F(x_{a_0-1}) < G(y_{b_0-1}) < F(x_{a_0}) < G(y_{b_0})$

We start out by considering pairs $(a,b) \in R(\leftrightharpoons)$ that are higher than (a_0, b_0). According to (7.20), $x_a > y_b$ is only possible for any such pair, if $a - a_0 = b - b_0 + 1$ holds. Hence, let $(a,b) \in R(\leftrightharpoons)$ be higher than (a_0, b_0) and satisfy $a - a_0 = b - b_0 + 1$. Part 2 of the proof of Theorem 7.30 states that, in this case, the values of F and G alternate for index pairs between (a_0, b_0) and (a,b). In that area, the pairs in $R(\leftrightharpoons)$ are of the form

$$
\begin{aligned}
&(a_0,b_0),(a_0+1,b_0),(a_0+1,b_0+1),(a_0+2,b_0+1),\dots\\
&\qquad \dots,(a-2,b-1),(a-1,b-1),(a-1,b),(a,b). \qquad (7.46)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\sum_{\substack{(a,b)\in R(\leftrightharpoons)\\ \text{higher than}(a_0,b_0)}} r_{(a,b)}(y_b - x_a)\\
&\geq - \sum_{\substack{c\in\mathbb{N}_0:\\ (a_0+1+c,b_0+c)\in R(\leftrightharpoons)}} r_{(a_0+1+c,b_0+c)}(x_{a_0+1+c} - y_{b_0+c})\\
&= - \sum_{\substack{c\in\mathbb{N}_0:\\ F(x_{a_0+c})<G(y_{b_0+c})}} (G(y_{b_0+c}) - F(x_{a_0+c}))(x_{a_0+1+c} - y_{b_0+c}). \qquad (7.47)
\end{aligned}
$$

Note that, due to the structure of $R(\leftrightharpoons)$ given in (7.46), $F(x_{a_0+c}) < G(y_{b_0+c})$ is equivalent to $(a_0+1+c, b_0+c) \in R(\leftrightharpoons)$ for any $c \in \mathbb{N}_0$. Since $(a_0+1+c) - (a_0+1) < (b_0+c) - b_0 + 1$ holds for all $c \in \mathbb{N}_0$ such that $(a_0+1+c, b_0+c) \in R(\leftrightharpoons)$, (7.20) implies $x_{a_0+1+c} - y_{b_0+c} \leq x_{a_0+1} - y_{b_0}$. Combined with (7.47), this yields

$$
\sum_{\substack{(a,b)\in R(\leftrightharpoons)\\ \text{higher than}(a_0,b_0)}} r_{(a,b)}(y_b - x_a) \geq -(x_{a_0+1} - y_{b_0}) \cdot \sum_{\substack{c\in\mathbb{N}_0:\\ F(x_{a_0+c})<G(y_{b_0+c})}} (G(y_{b_0+c}) - F(x_{a_0+c})). \qquad (7.48)
$$

For any $c \in \mathbb{N}_0$ with $F(x_{a_0+c}) < G(y_{b_0+c})$ and $F(x_{a_0+c+1}) < G(y_{b_0+c+1})$, it can be shown analogously to (7.37) that

$$\frac{G(y_{b_0+c+1}) - F(x_{a_0+c+1})}{F(x_{a_0+c+2}) - F(x_{a_0+c+1})} \le \frac{G(y_{b_0+c}) - F(x_{a_0+c})}{F(x_{a_0+c+1}) - F(x_{a_0+c})}.$$

By applying this inequality recursively, we obtain

$$\frac{G(y_{b_0}) - F(x_{a_0})}{F(x_{a_0+1}) - F(x_{a_0})} \ge \frac{G(y_{b_0+c}) - F(x_{a_0+c})}{F(x_{a_0+c+1}) - F(x_{a_0+c})}$$

for all $c \in \mathbb{N}_0$ with $F(x_{a_0+c}) < G(y_{b_0+c})$. With this, we can continue inequality (7.48) by writing

$$\begin{aligned}
&\sum_{\substack{(a,b)\in R(\leftrightharpoons)\\ \text{higher than}(a_0,b_0)}} r_{(a,b)}(y_b - x_a)\\
\ge\ & -(x_{a_0+1} - y_{b_0}) \cdot \sum_{\substack{c\in\mathbb{N}_0:\\ F(x_{a_0+c})<G(y_{b_0+c})}} (G(y_{b_0+c}) - F(x_{a_0+c}))\\
=\ & -(x_{a_0+1} - y_{b_0}) \cdot \sum_{\substack{c\in\mathbb{N}_0:\\ F(x_{a_0+c})<G(y_{b_0+c})}} \Bigg(\frac{G(y_{b_0+c}) - F(x_{a_0+c})}{F(x_{a_0+c+1}) - F(x_{a_0+c})} \cdot\\
&\qquad (F(x_{a_0+c+1}) - F(x_{a_0+c}))\Bigg)\\
\ge\ & -(x_{a_0+1} - y_{b_0}) \cdot \frac{G(y_{b_0}) - F(x_{a_0})}{F(x_{a_0+1}) - F(x_{a_0})} \sum_{\substack{c\ge 0\\ F(x_{a_0+c})<G(y_{b_0+c})}} (F(x_{a_0+c+1}) - F(x_{a_0+c}))\\
\ge\ & -(x_{a_0+1} - y_{b_0}) \cdot \frac{G(y_{b_0}) - F(x_{a_0})}{F(x_{a_0+1}) - F(x_{a_0})} \cdot (1 - F(x_{a_0}))\\
\ge\ & -\frac{x_{a_0+1} - y_{b_0}}{2} \cdot \frac{G(y_{b_0}) - F(x_{a_0})}{F(x_{a_0+1}) - F(x_{a_0})}.
\end{aligned} \tag{7.49}$$

Now we consider the pairs $(a, b) \in R(\leftrightharpoons)$ lower than (a_0, b_0). According to (7.20), $x_a - y_b < x_{a_0+1} - y_{b_0}$ is only possible for this kind of pair if $a_0 - a = b_0 - b$ holds. In this case, the values of F and G between these pairs of indices once again alternate so that the corresponding elements of $R(\leftrightharpoons)$ are of the form

$$(a_0, b_0), (a_0, b_0 - 1), (a_0 - 1, b_0 - 1), (a_0 - 1, b_0 - 2), \ldots$$
$$\ldots, (a + 2, b + 1), (a + 1, b + 1), (a + 1, b), (a, b).$$

It follows

$$\begin{aligned}
&\sum_{\substack{(a,b)\in R(\Leftarrow)\\ \text{lower than}(a_0,b_0)}} r_{(a,b)}(x_a - y_b)\\
&\geq (x_{a_0+1} - y_{b_0}) \cdot \left(F(x_{a_0}) - \sum_{\substack{c\in\mathbb{N}_0:\\ (a_0-c,b_0-c)\in R(\Leftarrow)}} r_{(a_0-c,b_0-c)} \right)\\
&= (x_{a_0+1} - y_{b_0}) \cdot \left(F(x_{a_0}) - \sum_{\substack{c\in\mathbb{N}_0:\\ F(x_{a_0-c})>G(y_{b_0-c-1})}} (F(x_{a_0-c}) - G(y_{b_0-c-1})) \right).
\end{aligned} \tag{7.50}$$

Note for the inequality that $x_a - y_b < 0 = x_{a_0} - y_{b_0}$ would imply $a_0 - a = b_0 - b + 1$, which contradicts $a_0 - a = b_0 - b$. For any $c \in \mathbb{N}_0$ with $F(x_{a_0-c}) > G(y_{b_0-c-1})$ and $F(x_{a_0-c-1}) > G(y_{b_0-c-2})$, one can now again show similarly to (7.37) that

$$\frac{F(x_{a_0-c}) - G(y_{b_0-c-1})}{F(x_{a_0-c}) - F(x_{a_0-c-1})} \geq \frac{F(x_{a_0-c-1}) - G(y_{b_0-c-2})}{F(x_{a_0-c-1}) - F(x_{a_0-c-2})}$$

holds. Inductively, it follows that

$$\frac{F(x_{a_0}) - G(y_{b_0-1})}{F(x_{a_0}) - F(x_{a_0-1})} \geq \frac{F(x_{a_0-c}) - G(y_{b_0-c-1})}{F(x_{a_0-c}) - F(x_{a_0-c-1})}$$

is true for all $c \in \mathbb{N}_0$ with $F(x_{a_0-c}) > G(y_{b_0-c-1})$. Continuing from (7.50), we obtain

$$\begin{aligned}
&\sum_{\substack{(a,b)\in R(\Leftarrow)\\ \text{lower than}(a_0,b_0)}} r_{(a,b)}(x_a - y_b)\\
&\geq (x_{a_0+1} - y_{b_0}) \cdot \left(F(x_{a_0}) - \sum_{\substack{c\in\mathbb{N}_0:\\ F(x_{a_0-c})>G(y_{b_0-c-1})}} \left(\frac{F(x_{a_0-c}) - G(y_{b_0-c-1})}{F(x_{a_0-c}) - F(x_{a_0-c-1})} \cdot \right.\right.\\
&\qquad\qquad \left.\left. (F(x_{a_0-c}) - F(x_{a_0-c-1})) \right) \right)\\
&\geq (x_{a_0+1} - y_{b_0}) \cdot \left(F(x_{a_0}) - \frac{F(x_{a_0}) - G(y_{b_0-1})}{F(x_{a_0}) - F(x_{a_0-1})} \cdot \right.\\
&\qquad\qquad \left. \sum_{\substack{c\in\mathbb{N}_0:\\ F(x_{a_0-c})>G(y_{b_0-c-1})}} (F(x_{a_0-c}) - F(x_{a_0-c-1})) \right)\\
&\geq (x_{a_0+1} - y_{b_0}) \cdot \left(F(x_{a_0}) - \frac{F(x_{a_0}) - G(y_{b_0-1})}{F(x_{a_0}) - F(x_{a_0-1})} \cdot F(x_{a_0}) \right)\\
&\geq \frac{x_{a_0+1} - y_{b_0}}{2} \cdot \left(1 - \frac{F(x_{a_0}) - G(y_{b_0-1})}{F(x_{a_0}) - F(x_{a_0-1})} \right).
\end{aligned} \tag{7.51}$$

Now, by combining (7.45), which is also valid in Case 2, with (7.49) and (7.51), it follows that

$$\begin{aligned}
&\mathrm{MDMAD}(G) - \mathrm{MDMAD}(F) \\
\geq\ & \frac{x_{a_0+1} - y_{b_0}}{2} \cdot \left(1 - \frac{G(y_{b_0}) - F(x_{a_0})}{F(x_{a_0+1}) - F(x_{a_0})} - \frac{F(x_{a_0}) - G(y_{b_0-1})}{F(x_{a_0}) - F(x_{a_0-1})}\right) \\
\geq\ & \frac{x_{a_0+1} - y_{b_0}}{2} \cdot \left(1 - \frac{q_{b_0}}{\min\{p_{a_0}, p_{a_0+1}\}}\right) \geq 0,
\end{aligned}$$

where the last inequality holds since $(a_0, b_0), (a_0 + 1, b_0) \in R(\leftrightharpoons)$ is true by assumption.

Case 3: $G(y_{b_0-1}) < F(x_{a_0-1}) < G(y_{b_0}) < F(x_{a_0})$

This case is completely analogous to Case 2 for reasons of symmetry.

7.5. Application to Specific Distributions

In this section, we analyze whether popular families of discrete distributions preserve the discrete dispersive orders. Since all of the distributions considered in the following are lattice distributions with defining distance equal to one, all previously defined discrete dispersive orders are equivalent and it is sufficient to consider the order $\leq_{D-pm}^{disc}$ (see Corollary 7.23a)). In the formulation of the results, the order $\leq_{disp}^{\wedge-disc}$ is used since it is the discrete order that is closest to $\leq_{disp}$. The only considered non-lattice distribution is the discrete uniform distribution on arbitrary finite sets, which is discussed at the end of the following subsection.

7.5.1. The Discrete Uniform Distribution and Empirical Distributions

The discrete uniform distribution is the simplest discrete distribution. In its usual variant, it puts the same amount of probability mass on a finite number of points that are equidistantly spaced with distance 1. Since all of the dispersion orders and measures considered in the previous chapters are location invariant, it is sufficient to consider uniform distributions with supports $\{1, \ldots, n\}, n \in \mathbb{N}_{\geq 2}$. If $\mathbb{P}(X = k) = \frac{1}{n}$ for all $k \in \{1, \ldots, n\}$, we denote this by $X \sim \mathcal{U}[n]$. These distributions are also used in Example 6.7 in order to establish that the original dispersive order is far from sufficient for discrete distributions. However, it is easy to show that any two discrete uniform distributions are ordered with respect to the discrete dispersive orders introduced in this work.

Proposition 7.42. *Let $n, m \in \mathbb{N}_{\geq 2}, n < m$, and let $X \sim \mathcal{U}[n]$ and $Y \sim \mathcal{U}[m]$. Then, $X \leq_{disp}^{\wedge-disc} Y$ holds.*

Proof. Since X and Y both have lattice distributions with the same defining distance, Corollary 7.23a) states that $X \leq_{disp}^{\wedge-disc} Y$ is equivalent to $X \leq_{D-pm}^{disc} Y$. Since the height of all

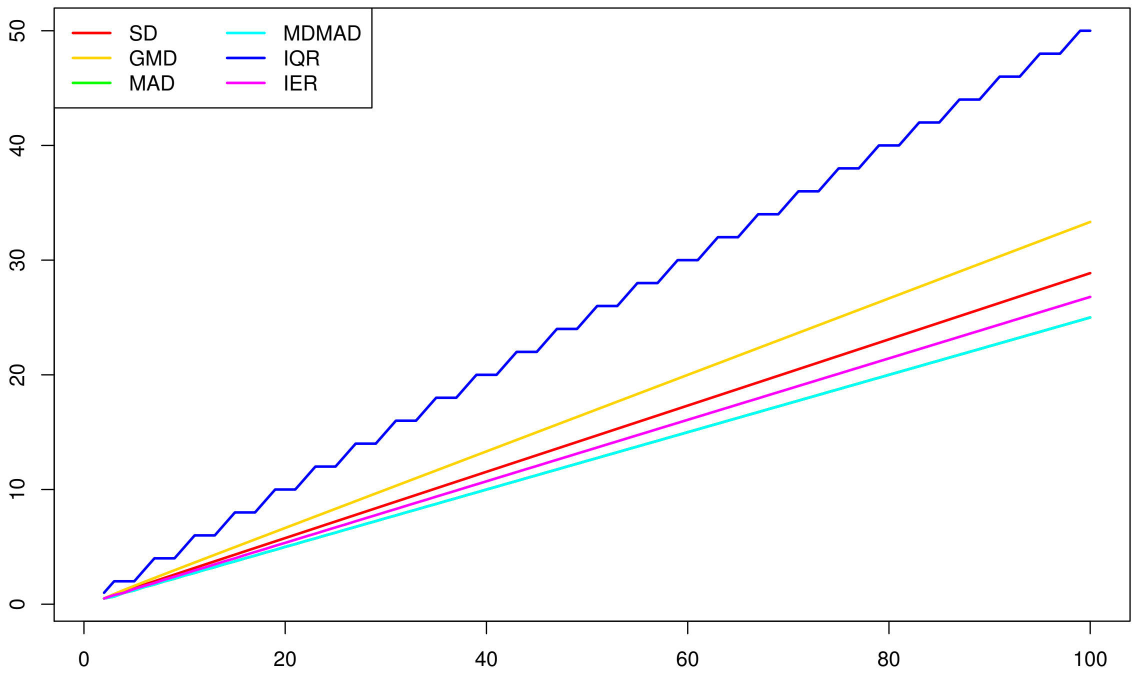

Figure 7.11.: Plot of $\tau(X)$ for six different dispersion measures τ and $X \sim \mathcal{U}[n]$, as a function of $n \in \{2, \ldots, 100\}$. The measures MAD and MDMAD coincide here because of symmetry.

jumps of the cdf of X is equal to $\frac{1}{n}$ and, therefore, larger than $\frac{1}{m}$, which is the height of all jumps of the cdf of Y, $X \leq_{D-pm}^{disc} Y$ holds. ▨

The behaviour of the dispersion measures from Chapter 7.4 for discrete uniform distributions as a function of the parameter n is depicted in Figure 7.11. It shows that five of the six dispersion measures are almost linearly increasing as a function of n, although the slight deviations from linearity can barely be seen in Figure 7.11. The average slopes differ between the measures; only MAD and MDMAD are exactly the same since the distribution is symmetric. The graph of the interquartile range IQR has a different shape since it only takes values in the natural numbers when applied to a lattice distribution with defining distance 1. This lack of granularity is also somewhat indicative of its lack of compatibility with discrete distributions that is formalized in Theorem 7.40. However, a counterexample for the proof of Theorem 7.40 cannot be constructed using this class of discrete uniform distributions.

The concept of discrete uniform distributions can be generalized to arbitrary finite sets. Let $S \subset \mathbb{R}$ with $|S| = n \in \mathbb{N}_{\geq 2}$. If now $\mathbb{P}(X = s) = \frac{1}{n}$ for any $s \in S$, then X is discretely uniformly distributed on S, denoted by $X \sim \mathcal{U}(S)$. Note that the set of all generalized discrete uniform distributions is equal to the set $\mathcal{E}_{nt}$ of all non-tied empirical distributions. Because of the complexity of this family of distributions, we refrain from trying to obtain general results with respect to discrete dispersive orders. Instead, we discuss a number of special cases for the order $\leq_{disp}^{\wedge-disc}$ in the following example.

Example 7.43. Let $S, T \subset \mathbb{R}$ with $2 \leq |S| = m < n = |T|$ and let $X \sim \mathcal{U}(S)$, $Y \sim \mathcal{U}(T)$.

a) Let n be a multiple of m, so there exists a $c \in \mathbb{N}_{\geq 2}$ such that $n = c \cdot m$. Because of

$$\begin{aligned} F(D_F) &= \{\tfrac{k}{m} : k \in \{1, \ldots, m-1\}\} = \{\tfrac{c \cdot k}{n} : k \in \{1, \ldots, m-1\}\} \\ &\subset \{\tfrac{k}{n} : k \in \{1, \ldots, n-1\}\} = G(D_G), \end{aligned}$$

$F \leq_{disp}^{\wedge-disc} G$ is equivalent to $F \leq_{disp} G$ in this case. According to Proposition 7.4, $F \leq_{disp}^{\wedge-disc} G$ holds if $x_{k+1} - x_k \leq y_{c \cdot k+1} - y_{c \cdot k}$ holds for all $k \in \{1, \ldots, m-1\}$. Hence, $m-1$ comparisons are made overall, one per constant interval of F.

b) Let n be a multiple of m plus one, so there exists a $c \in \mathbb{N}$ such that $n = c \cdot m + 1$. It follows that m and n are coprime and $F(D_F) \cap G(D_G) = \emptyset$. Furthermore, for each $k \in \{1, \ldots, m-1\}$, it holds that $\frac{k}{m} = \frac{c \cdot k}{n-1} \in (\frac{c \cdot k}{n}, \frac{c \cdot k+1}{n})$. Hence, $\mathrm{NN}_F^G(k+1) = \{\frac{c \cdot k}{n}, \frac{c \cdot k+1}{n}\}$ for all $k \in \{1, \ldots, m-1\}$. Since $F \leq_{D-pm}^{disc} G$ is obviously satisfied, Proposition 7.19a) states that $F \leq_{disp}^{\wedge-disc} G$ is equivalent to

$$\begin{aligned} x_{k+1} - x_k &\leq y_{c \cdot k+1} - y_{c \cdot k} \quad \text{and} \\ x_{k+1} - x_k &\leq y_{c \cdot k+2} - y_{c \cdot k+1} \end{aligned}$$

for all $k \in \{1, \ldots, m-1\}$. Hence, $2m-2$ comparisons are made overall, two per constant interval of F.

c) Let the greatest common divisor d of m and n satisfy $1 < d < m$, so there exist $c_F, c_G \in \mathbb{N}_{\geq 2}$ such that $m = c_F \cdot d$ and $n = c_G \cdot d$. Then, $F(D_F) \cap G(D_G) = \{\frac{k}{d} : k \in \{1, \ldots, d-1\}\}$ since

$$F(D_F) \ni \frac{k \cdot c_F}{m} = \frac{k \cdot c_F}{c_F \cdot d} = \frac{k}{d} = \frac{k \cdot c_G}{c_G \cdot d} = \frac{k \cdot c_G}{n} \in G(D_G)$$

holds for all $k \in \{1, \ldots, d-1\}$ and because c_F and c_G are coprime by assumption. It follows that $|\,\mathrm{NN}_F^G(\ell+1)| = 1$, if ℓ is a multiple of c_F, and $|\,\mathrm{NN}_F^G(\ell+1)| = 2$ otherwise. In the former case, the corresponding comparisons (by Proposition 7.19a)) have the same structure as in part a), and in the latter case they have the same structure as in part b). Overall, there are $2m - 2 - (d-1) = (2c_F - 1)d - 3$ comparisons to be made. Note that the edge cases $d = m$ and $d = 1$ give the situation in part a) and part b), respectively.

7.5.2. The Geometric Distribution

Except for the discrete uniform distribution, the geometric distribution is the only popular type of discrete distribution with an explicit representation of the cdf. For other popular families of discrete distributions, like the binomial or the Poisson distribution, the cdf is only available as sum of the values of the corresponding pdf.

We use the following version of the geometric distribution: if $X \sim \mathrm{Geom}(\pi)$ with $\pi \in (0,1)$, then $\mathbb{P}(X = k) = \pi \cdot (1-\pi)^{k-1}$ for $k \in \mathbb{N}$. The cdf of X is then given by $F(t) = 1 - (1-\pi)^{\lfloor t \rfloor}$ for $t \geq 0$. Graphically, the dispersion of the distribution seems to decrease as the parameter π increases. Furthermore, for $F = \mathrm{Geom}(\pi_F)$ and $G = \mathrm{Geom}(\pi_G)$ with $0 < \pi_G < \pi_F < 1$, $F \leq_{st} G$ and even $F <_{st} G$ obviously holds, which already implies $G \not\leq_{disp}^{\wedge-disc} F$ according to Proposition 7.28. The following result gives a sufficient condition for the ordering of two geometric distributions with respect to $\leq_{disp}^{\wedge-disc}$.

Theorem 7.44. *Let $X \sim \mathrm{Geom}(\pi_F)$ and $Y \sim \mathrm{Geom}(\pi_G)$ with $0 < \pi_G < \pi_F < 1$ have cdf's F and G. If*

$$(\pi_F, \pi_G) \in \left\{ (1-\lambda^{\varrho}, 1-\lambda) : \frac{1}{2} < \lambda < 1, \varrho \geq \frac{\log(2\lambda - 1)}{\log(\lambda)} - 1 \right\},$$

then $F \leq_{disp}^{\wedge-disc} G$ holds.

Proof. Let $\lambda = 1 - \pi_G \in (0,1)$ and $\varrho = \frac{\log(1-\pi_F)}{\log(1-\pi_G)} > 1$, then $\pi_F = 1 - \lambda^{\varrho}$ and $\pi_G = 1 - \lambda$. Note that $F \leq_{D-supp}^{\wedge-disc} G$ holds because both are lattice distributions with defining distance 1. We start by finding an equivalent condition for $(a,b) \in R(\Leftrightarrows)$, where $a, b \in \mathbb{N} = A = B$. The statement $(a,b) \in R(\Leftrightarrows)$ is equivalent to $F(a-1) < G(b)$ and $G(b-1) < F(a)$ holding simultaneously. For these two inequalities, the following equivalences hold:

$$\begin{aligned}
&F(a-1) < G(b) \Leftrightarrow 1 - \lambda^{\varrho(a-1)} < 1 - \lambda^{b} \Leftrightarrow \lambda^{\varrho(a-1)} > \lambda^{b} \Leftrightarrow \varrho(a-1) < b \Leftrightarrow a < \tfrac{b}{\varrho} + 1, \\
&G(b-1) < F(a) \Leftrightarrow 1 - \lambda^{b-1} < 1 - \lambda^{\varrho a} \Leftrightarrow \lambda^{b-1} > \lambda^{\varrho a} \Leftrightarrow b - 1 < \varrho a \Leftrightarrow a > \tfrac{b-1}{\varrho}.
\end{aligned}$$

Overall, $(a,b) \in R(\Leftrightarrows)$ is equivalent to $a \in (\frac{b-1}{\varrho}, \frac{b}{\varrho} + 1)$. Because a is required to be a natural number and $(\frac{b}{\varrho} + 1) - \frac{b-1}{\varrho} = 1 + \frac{1}{\varrho} \in (1,2)$, there are either one or two possible values for a for a fixed $b \in \mathbb{N}$. Particularly, it follows that

$$a \in \left\{ \left\lfloor \frac{b}{\varrho} \right\rfloor + 1, \left\lceil \frac{b-1}{\varrho} \right\rceil \right\}.$$

The two values are different, if there exists an $n \in \mathbb{N}$ such that $\frac{b-1}{\varrho} \leq n \leq \frac{b}{\varrho}$. We consider the two elements separately. To this end, note that

$$\begin{aligned}
\lambda^{b-1} - \lambda^{b} &= \mathbb{P}(Y = b) \\
&\leq \mathbb{P}(X = a) = \lambda^{\varrho(a-1)} - \lambda^{\varrho a} \quad \forall (a,b) \in \bigcup_{b \in \mathbb{N}} \left\{ \left(\left\lfloor \frac{b}{\varrho} \right\rfloor + 1, b \right), \left(\left\lceil \frac{b-1}{\varrho} \right\rceil, b \right) \right\}
\end{aligned}$$

is equivalent to $F \leq_{D-pm}^{disc} G$, and therefore to $F \leq_{disp}^{\wedge-disc} G$.

Case 1: Let $a = \left\lceil \frac{b-1}{\varrho} \right\rceil$.
It holds that

$$\lambda^{\varrho(a-1)} - \lambda^{\varrho a} = \lambda^{\varrho(a-1)}(1-\lambda^{\varrho}) > \lambda^{\varrho \cdot \frac{b-1}{\varrho}}(1-\lambda^{\varrho}) = \lambda^{b-1} - \lambda^{b-1+\varrho} > \lambda^{b-1} - \lambda^{b},$$

where the last inequality is true because $\varrho > 1$ implies $\lambda^{b-1+\varrho} < \lambda^{b}$.

Case 2: Let $a = \left\lfloor \frac{b}{\varrho} \right\rfloor + 1$.
It holds that

$$\lambda^{\varrho(a-1)} - \lambda^{\varrho a} = \lambda^{\varrho(a-1)}(1-\lambda^{\varrho}) > \lambda^{\varrho \cdot \frac{b}{\varrho}}(1-\lambda^{\varrho}) = \lambda^{b}(1-\lambda^{\varrho}).$$

Thus, it is sufficient to show

$$\lambda^{b-1}(\lambda - \lambda^{\varrho+1}) \geq \lambda^{b-1} - \lambda^{b}$$

or, equivalently,

$$2\lambda \geq \lambda^{\varrho+1} + 1.$$

This, in turn, is equivalent to

$$\begin{aligned} \lambda(2-\lambda^{\varrho}) \geq 1 \Leftrightarrow \lambda^{\varrho} &\leq 2 - \tfrac{1}{\lambda} \\ \Leftrightarrow \varrho &\geq \frac{\log(2-\frac{1}{\lambda})}{\log(\lambda)} = \frac{\log(2\lambda-1)}{\log(\lambda)} - 1, \end{aligned}$$

which is true by assumption.

▨

The set of parameter pairs from Theorem 7.44 is visualized in the left panel of Figure 7.12, where it is the green area on the lower right. The grey area represents those combinations of parameters, for which no theoretical result could be obtained. In order to determine the behaviour in these grey areas, a numerical analysis was conducted. Since the support of the geometric distribution is infinite, the cdf's and pdf's were cut off at 10^6. The results with 0.01 as increment for the parameters π_F and π_G are depicted in the right panel of Figure 7.12. The numerical results look almost identical to the theoretical results with the grey area filled in red. It is not clear whether the few sparse green dots in that area actually represent $F \leq_{disp}^{\wedge-disc} G$ holding or they represent numerical inaccuracies. Either way, the numerical results suggest that the implication in Theorem 7.44 is close to being an equivalence as the number of counterexamples for the reverse implication is very small.

The behaviour of the dispersion measures from Section 7.4 applied to geometric distributions is shown in Figure 7.13. First, it is immediately obvious that the graphs all have similar shapes. While that includes the interquartile range IQR to a certain degree, its graph is the only one

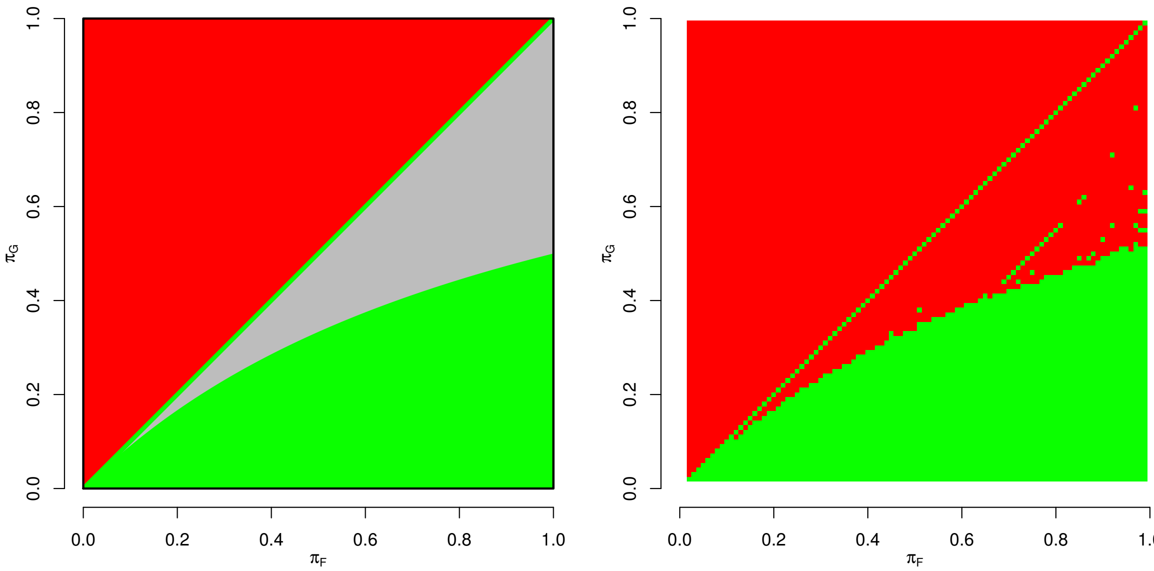

Figure 7.12.: Plot of theoretical (left panel) and numerical (right panel) results concerning $F \leq_{disp}^{\wedge-disc} G$ for all possible parameter choices $\pi_F, \pi_G \in (0,1)$ for $F = \mathrm{Geom}(\pi_F)$ and $G = \mathrm{Geom}(\pi_G)$. In green areas, $F \leq_{disp}^{\wedge-disc} G$ holds; in red areas, $F \not\leq_{disp}^{\wedge-disc} G$ holds; in grey areas, no result could be obtained.

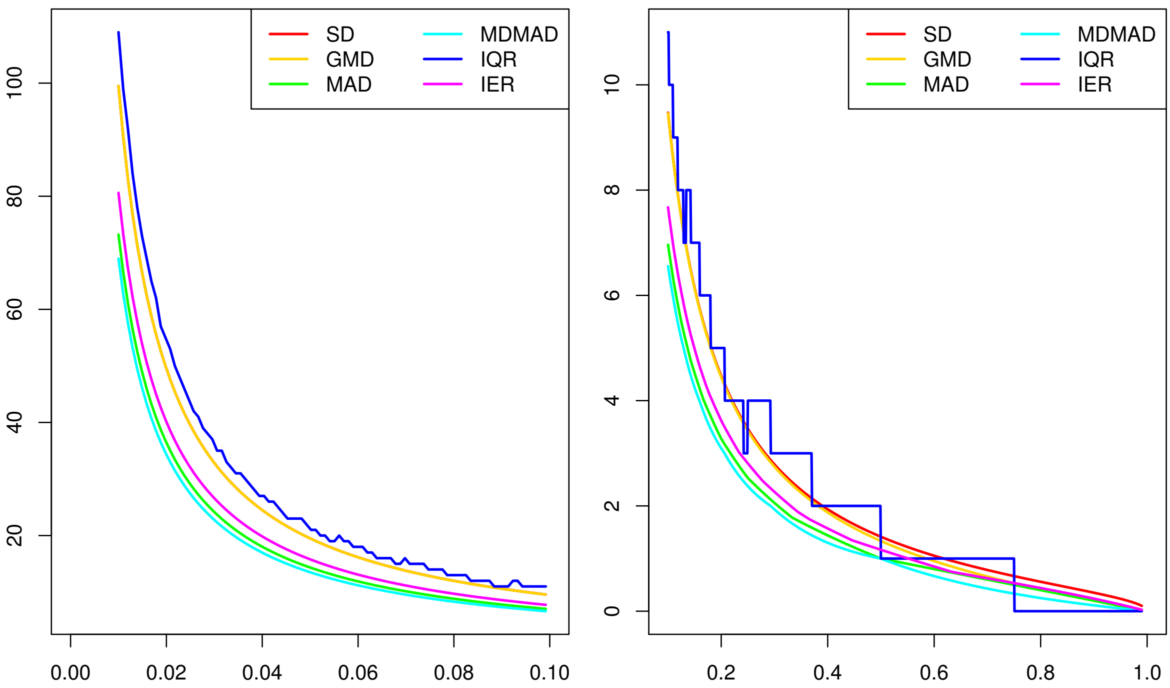

Figure 7.13.: Plot of $\tau(X)$ for six different dispersion measures τ and $X \sim \mathrm{Geom}(\pi)$, as a function of $\pi \in (0,1)$. Note the different scales in the two panels.

that is not decreasing on the entire parameter space. Furthermore, the slopes of all graphs decrease significantly for increasing parameter values. Thus, there is a smaller difference in dispersion between two similarly high values of π than there is between two similarly low values of π. This observation is in agreement with the behaviour of the discrete dispersion order for geometric distributions. Consider the following example: according to Theorem 7.44 and Figure 7.12, $F \leq_{disp}^{\wedge-disc} G$ holds for $\pi_F = 0.15$ and $\pi_G = 0.12$ while $F \not\leq_{disp}^{\wedge-disc} G$ holds for $\pi_F = 0.9$ and $\pi_G = 0.72$, which differ from each other by the same factor.

7.5.3. The Binomial Distribution

The binomial distribution has two parameters to be varied, namely the sample size n and the success probability π. However, if we consider two distributions $F = \mathrm{Bin}(n, \pi_F)$ and $G = \mathrm{Bin}(n, \pi_G)$ with $n \in \mathbb{N}_{\geq 2}$ and $\pi_F, \pi_G \in (0,1), \pi_F \neq \pi_G$, Proposition 7.26 states that neither $F \leq_{disp}^{\wedge-disc} G$ nor $G \leq_{disp}^{\wedge-disc} F$ holds. That is because of $D_F = D_G = [0, n)$, which yields $\lambda^1(D_F) = \lambda^1(D_G) = n < \infty$. Heuristically, the binomial distribution seems to be most dispersed when it is symmetric. Its dispersion declines, if the success probability becomes markedly high or low as then, the probability mass is concentrated heavily on one side. This observation is reflected in the left panel of Figure 7.14, which depicts the behaviour of the dispersion measures from Chapter 7.4 for fixed n and varying π.

Furthermore, the plot shows that the dispersion measures display differing degrees of smoothness as a function of the success probability. While IQR is the only measure that is not continuous, MAD, MDMAD and IER also exhibit some lack of smoothness. Solely the graphs of SD and GMD look like they could stem from an infinitely often differentiable function.

For binomial distributions with fixed success probability π and varying sample size n, we restrict ourselves to the symmetric case $\pi = \frac{1}{2}$. If we consider two distributions $F = \mathrm{Bin}(m, \frac{1}{2})$ and $G = \mathrm{Bin}(n, \frac{1}{2})$ with $m, n \in \mathbb{N}_{\geq 2}$ and $m < n$, we can once again invoke Proposition 7.26 to obtain $G \not\leq_{disp}^{\wedge-disc} F$. The remaining question is: if at all, under which conditions concerning m and n does $F \leq_{disp}^{\wedge-disc} G$ hold? Because of the non-explicit structure of the cdf of the binomial distribution, no theoretical result answering this question could be proved. Instead, we have to rely solely on numerical computations.

The results are depicted in Figure 7.15. They generally support the graphical impression that the (symmetric) binomial distribution becomes more dispersed as its sample size increases. However, the difference between the two sample sizes m and n needs to be quite large for $F \leq_{disp}^{\wedge-disc} G$ to hold. For $n < 5m$, $F \leq_{disp}^{\wedge-disc} G$ only holds very sporadically. For $5m \leq n \leq 10m$, $F \leq_{disp}^{\wedge-disc} G$ holds in some cases, depending on the compatibility of the two distributions. However, as n approaches $10m$, the share of positive results seems to increase. Finally, for $n > 10m$, $F \leq_{disp}^{\wedge-disc} G$ always holds with very few exceptions if $n \approx 10m$. It is notable that the borders between the red and the mixed area as well as between the mixed and the green area both seem to be approximately linear. According to further numerical evaluations for larger sample sizes, the factors of 5 and 10 seem to grow a bit further to

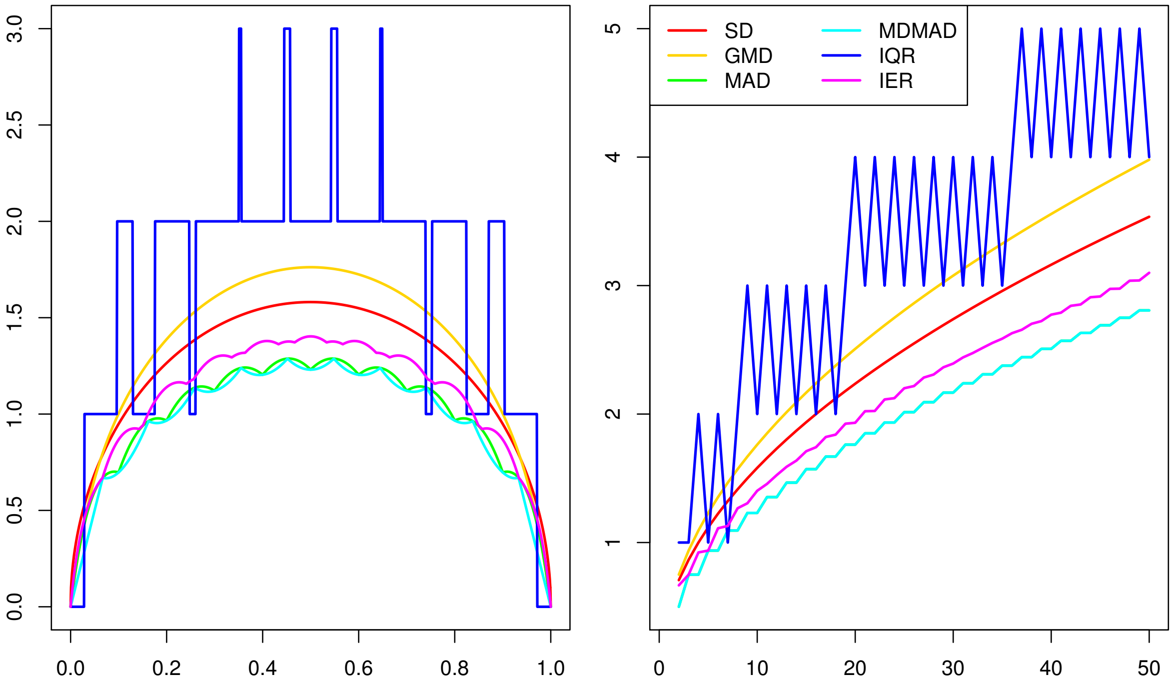

Figure 7.14.: Plot of $\tau(X)$ for six different dispersion measures τ and $X \sim \text{Bin}(n, \pi)$. Left panel: $n = 10$ fixed, $\tau(X)$ as a function of π. Right panel: $\pi = \frac{1}{2}$ fixed, $\tau(X)$ as a function of n.

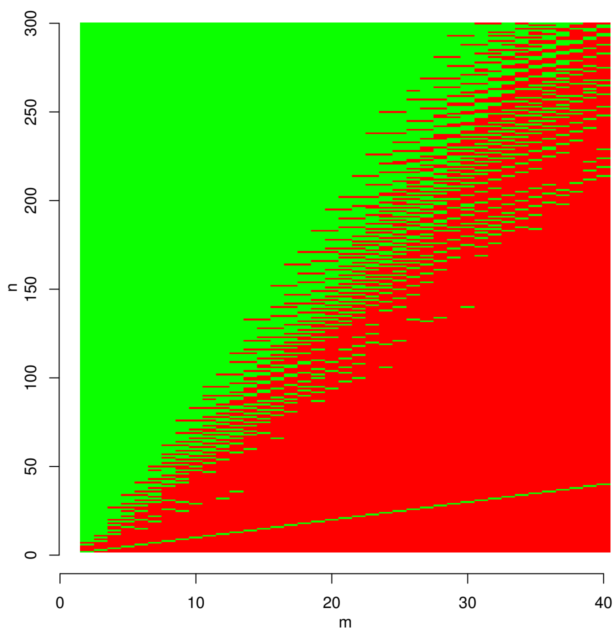

Figure 7.15.: Plot of the numerical results concerning $F \leq_{disp}^{\wedge-disc} G$ for selected parameter values for $F = \text{Bin}(m, \frac{1}{2})$ and $G = \text{Bin}(n, \frac{1}{2})$. In green areas, $F \leq_{disp}^{\wedge-disc} G$ holds; in red areas, $F \not\leq_{disp}^{\wedge-disc} G$ holds. Note the different scales of the two axes.

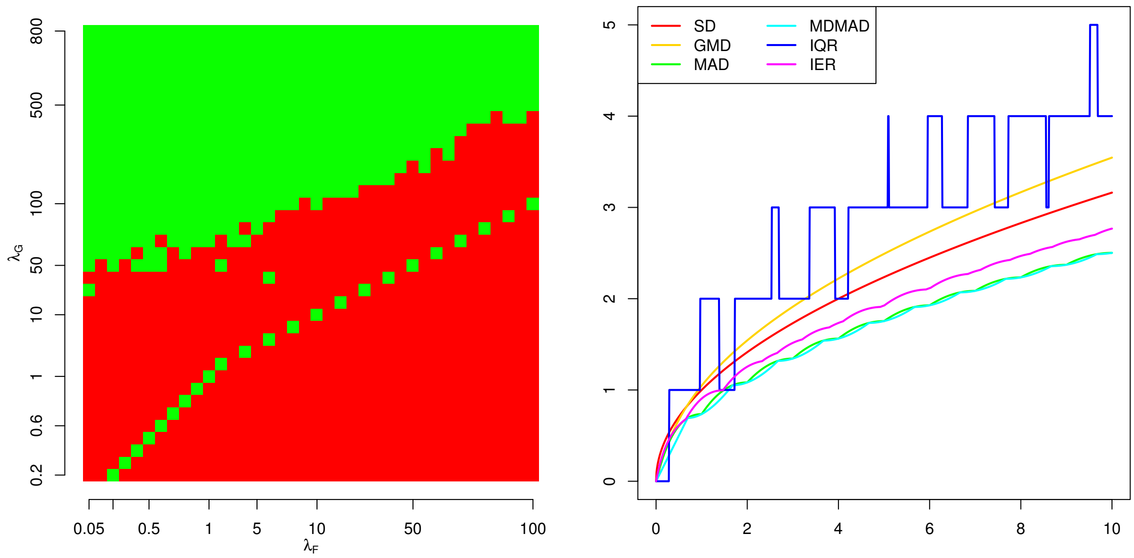

Figure 7.16.: Left panel: Plot of the numerical results concerning $F \leq_{disp}^{\wedge-disc} G$ (green: yes; red: no) for selected parameter values for $F = \text{Pois}(\lambda_F)$ and $G = \text{Pois}(\lambda_G)$. Note that the scale on both axes is not linear and only contains exemplary values. Right panel: Plot of $\tau(X)$ for six different dispersion measures τ and $X \sim \text{Pois}(\lambda)$, as a function of λ.

approximately 8.5 and 12 at $m = 400$.

The behaviour of the dispersion measures when applied to symmetric binomial distributions is similar to our previous observations. All graphs of the corresponding plot in the right panel of Figure 7.14 are increasing, once again with the exception of IQR. Their slopes slightly decrease as n is increasing. Their smoothness properties coincide with our observations from the left panel of Figure 7.14, where π varies instead of n.

7.5.4. The Poisson Distribution

The last exemplary discrete distribution considered in this section is the Poisson distribution. For that, we consider two distributions $F = \text{Pois}(\lambda_F)$ and $G = \text{Pois}(\lambda_G)$ with $\lambda_F, \lambda_G > 0$. Similarly to the geometric distribution, it is easy to show that $F \leq_{st} G$ and even $F <_{st} G$ holds, if $\lambda_F < \lambda_G$. By Proposition 7.28, $G \not\leq_{disp}^{\wedge-disc} F$ follows in that case. Whether $F \leq_{disp}^{\wedge-disc} G$ holds can once again only be analyzed numerically since the cdf of the Poisson distribution also does not have an explicit form.

The results for selected values of λ_F and λ_G are depicted in the left panel of Figure 7.16. As for the binomial distribution, $F \leq_{disp}^{\wedge-disc} G$ only seems to hold, if λ_G is sufficiently large compared to λ_F. However, the differing factor between λ_F and λ_G at the border between the red and the green area decreases for increasing λ_F. For $\lambda_F = 0.05$, that factor is equal to 600, and it is subsequently reduced: to 70 for $\lambda_F = 1$, to 10 for $\lambda_F = 10$, and to 5 for $\lambda_F = 100$. It is unclear whether this reduction is representative of the actual interaction between the Poisson distribution and the order $\leq_{disp}^{\wedge-disc}$ or it is a numerical phenomenon. The latter explanation is supported by the fact that, with increasing parameter λ, the amount of probability mass within jumps too small to register numerically also increases. Therefore, more relevant jumps cannot be compared properly.

The behaviour of the dispersion measures plotted in the right panel of Figure 7.16 is similar to the previous distribution families. The declining slope of the graphs is indicative of smaller differences in dispersion for higher parameter values and therefore suggests that our observations about the left panel are indeed due to numerical inaccuracies.

CHAPTER 8

Conclusion and Outlook

8.1. Alternative Approaches

In Section 7.1, the proposals for discrete dispersive orders are carefully constructed from the information available about the original dispersive order, particularly about its behaviour on discrete distributions. However, one could of course pursue other approaches to define such an order. A number of these approaches are named in this section along with a brief discussion of their advantages and disadvantages.

The first alternative approach arises from the inconsistency of $\leq_{disp}^{\wedge-disc}$ noted at the end of Example 7.16b) and only differs from our definitions in the way that the supports are compared. More specifically, it does not underlie the limitation that only the distances between pairs of *neighbouring* points in the respective supports are compared. Instead, referring back to the general setting in Example 7.16b), one could compare the distances $x_2 - x_1$ and $x_3 - x_2$ with the distances $y_3 - y_1$, $y_6 - y_3$ and $y_8 - y_6$ (see also Figure 7.3). In a general setting, $F \in \mathcal{D}_0$ would be deemed less dispersed than $G \in \mathcal{D}_0$, if $F \leq_{D-pm}^{disc} G$ holds as well as a modification of $F \leq_{D-supp}^{\wedge-disc} G$. The pointwise comparison of the supports indicated by that modification is again dictated by F, but the length $x_a - x_{a-1}$ of each constant interval of F is compared with the cumulated length $\sum_{b \in \underline{B}: p_{a-1} \leq q_{b-1} < p_a} (y_b - y_{b-1})$ of all constant intervals of G lower than the next constant interval of F. $x_a - x_{a-1}$ is also be compared with the cumulated length $\sum_{b \in \underline{B}: p_{a-2} < q_{b-1} \leq p_{a-1}} (y_b - y_{b-1})$ of constant intervals of G below the interval $[x_{a-1}, x_a)$. This weakens the strong order $\leq_{disp}^{\wedge-disc}$, but seemingly to a small enough degree that the new order is still meaningful. We conjecture that this new order, just like $\leq_{disp}^{\vee-disc}$, is transitive, but not equivalent to $\leq_{disp}$ on their joint area of applicability. Thus, it seems like a suitable order to explore in further work. A downside is that it cannot be described with the relation $\leftrightharpoons$ as intuitively as the other discrete orders. Therefore, it might be difficult or even impossible to replicate some of the results in Sections 7.3 and 7.4, while the results in Section 7.5 are not going to be improved since they only concern lattice distributions.

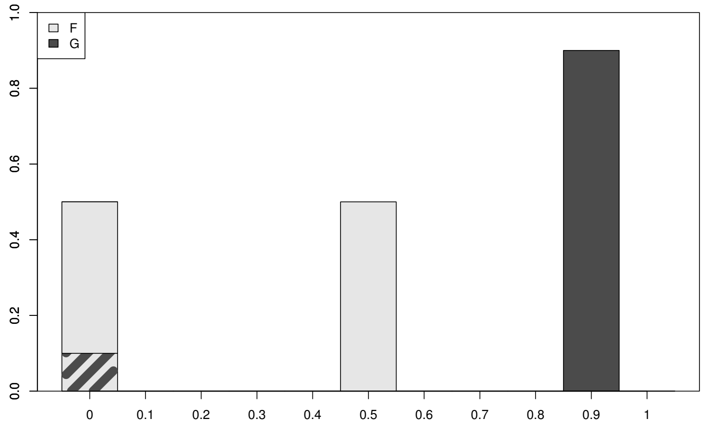

Figure 8.1.: Exemplary pair of discrete distributions, for which the second alternative approach in Section 8.1 behaves counterintuitively.

A second approach arises from using the characterization of $\leq_{disp}$ based on RIDF's, which is often very useful in a continuous setting. However, it is discussed in Chapter 6 that the concept of RIDF's has several shortcomings in a discrete setting. The fact that this is particularly true for the dispersive order is extended upon in Section 6.2 with the use of the specific Example 6.5. There, a dispersion order that is defined by the corresponding RIDF being increasing is shown to behave counterintuitively for discrete distributions. Thus, this approach is not sensible in a discrete setting. An adaptation solution in the form of a generalization of the concept of RIDF's is proposed in Appendix A, but it cannot replicate a number of crucial properties of the original concept.

The third approach is to only have one requirement for the discrete order instead of two as for $\leq_{disp}^{\vee-disc}$ and $\leq_{disp}^{\wedge-disc}$. Note that, under suitable regularity conditions, $F \leq_{disp} G$ is equivalent to $\frac{\mathrm{d}F^{-1}(p)}{\mathrm{d}p} \leq \frac{\mathrm{d}G^{-1}(p)}{\mathrm{d}p}$ for all $p \in (0,1)$. Hence, a canonical combination of the requirements concerning the probability mass and the support for discrete distributions is to consider the slopes of their appropriately interpolated discrete quantile functions. The points to be interpolated are given by $(x, F^{-1}(x))$ for $x \in F(\mathrm{supp}(F))$. However, this approach has several disadvantages. First, it imposes no restriction on the first jumps of the two cdf's to be compared, which is obviously necessary to compare them in terms of dispersion. Even if the first jump of F is additionally required to be higher than the first jump of G, one can still construct examples, for which the resulting discrete dispersion order disagrees with all well-known measures of dispersion. One such example is obtained for $X = \frac{1}{2}\tilde{X}$ with $\tilde{X} \sim \mathrm{Bin}(1, \frac{1}{2})$ and $Y = (1-\varepsilon)\tilde{Y}$ with $\tilde{Y} \sim \mathrm{Bin}(1, 1-\varepsilon)$ and $\varepsilon > 0$ sufficiently small (see Figure 8.1, where $\varepsilon = \frac{1}{10}$). Here, this new order would deem G more dispersed than F although G converges to the cdf of a degenerate distribution for $\varepsilon \searrow 0$. Furthermore, the interpolation order changes significantly, if we use another definition of the quantile function, e.g. if it is defined by $p \mapsto \sup\{t \in \mathbb{R} : F(t) \leq p\}$.

In conclusion, the only alternative approach from this section that seems worth pursuing further is the first one. It is conjectured to be a transitive weakening of $\leq_{disp}^{\wedge-disc}$ that, on the other hand, is more difficult to handle. All other discussed approaches have critical drawbacks.

8.2. A Discrete Skewness Order

The canonical next step after having found discrete versions of the dispersive order in Chapter 7 is to consider skewness orders for discrete distributions. The most popular and also most fundamental skewness order is the convex transformation order introduced by van Zwet (1964, p. 48), see Definition 2.11. Contrary to the stochastic order and the dispersive order, which are usually defined according to Definition 2.8b) in the literature (see, e.g., Oja, 1981, p. 157, Shaked and Shanthikumar, 2006, p. 148 and Müller and Stoyan, 2002, p. 40), the convex transformation order is normally defined using RIDF-based characterization in Proposition 2.9 (see, e.g., van Zwet, 1964, p. 48, Oja, 1981, p. 160 and Groeneveld and Meeden, 1984, p. 392). In the papers that do so, this does not make a difference because they generally at least assume that all involved distributions are absolutely continuous and have interval support, i.e., they lie in $\mathcal{P}_I$. This assumption is sufficient for the two different definitions to be equivalent. However, just like for the stochastic order and the dispersive order, these characterizations differ, if they are applied to discrete distributions. Thus, a choice needs to be made in order to analyze the convex transformation order $\leq_c$ on discrete distributions.

For the sake of consistency, we use the same definition for $\leq_c$, the order of the second convex characteristic, as we used for the orders of zeroth and the first convex characteristic, $\leq_{st}$ and $\leq_{disp}$. Another reason to not use the RIDF-based characterisation of $\leq_c$ as its definition is that RIDF's behave counterintuitively on non-continuous distributions. This is evidenced generally in Appendix A and specifically for the orders $\leq_{st}$ and $\leq_{disp}$ in Chapter 6. The basic problem is that RIDF's generally do not retain their crucial property for discrete distribution, which states that R_{FG} is capable of transforming an F-distributed random variable into a G-distributed random variable (see Proposition A.3).

The general Definition 2.8 of the order of the k-th convex characteristic is examined in the case $k = 2$ in Example 2.10c). More specifically, the definition of $F \leq_c G$ for two cdf's $F, G \in \mathcal{P}$ is given in (2.11) by

$$\frac{F^{-1}(w) - F^{-1}(v)}{F^{-1}(v) - F^{-1}(u)} \leq \frac{G^{-1}(w) - G^{-1}(v)}{G^{-1}(v) - G^{-1}(u)} \tag{8.1}$$

for all $0 < u < v < w < 1$ with $F^{-1}(u) < F^{-1}(w)$ and $G^{-1}(u) < G^{-1}(w)$, where division by zero is allowed with the value ∞ assigned in that case. Note that, according to Lemma 3.9, this is equivalent to the comparison of standardized second order differences of the two quantile functions.

One can easily see that the shortcomings of the dispersive order $\leq_{disp}$ for discrete distributions discussed in Section 6.2 apply to the convex transformation order $\leq_c$ in a similar way.

This means that, just like for $\leq_{disp}$, a discrete version of $\leq_c$ is needed. Eberl and Klar (2019) also demonstrated the need for a fundamental discrete skewness order, although they used a RIDF-based definition of $\leq_c$.

The first step towards this kind of discrete modification is to find out what exactly the order $\leq_c$ means in a discrete context. Analogously to Proposition 7.4, the equivalent characterization can be divided into requirements concerning the jumps heights and requirements concerning the support or the jump points.

Theorem 8.1. *Let $F, G \in \mathcal{D}_0$. Then, $F \leq_c G$ is equivalent to the following three conditions:*

(i) $F(D_F) \subseteq G(D_G) \cup (0, \inf G(D_G)]$,

(ii) $G(D_G) \subseteq F(D_F) \cup [\sup F(D_F), 1)$,

(iii)
$$\frac{x_{a+1} - x_a}{x_a - x_{a-1}} \leq \frac{y_{b+1} - y_b}{y_b - y_{b-1}} \tag{8.2}$$
holds for all $(a, b) \in \overline{\underline{A}} \times \overline{\underline{B}}$ with $F(x_{a-1}) = G(y_{b-1})$ and $F(x_a) = G(y_b)$.

Proof. We start by proving $F \leq_c G \Rightarrow (i)$ by contradiction. For that, assume $F(D_F) \setminus (G(D_G) \cup (0, \inf G(D_G)]) \neq \emptyset$ and let $r \in (0, 1)$ be in that set. Note that, since 0 and 1 are the only possible accumulation points of $G(D_G)$ (because of $G \in \mathcal{D}_0$), either $\inf G(D_G) = 0$ or $\inf G(D_G) = \min G(D_G)$ holds. Hence, the set $G(D_G) \cap (0, r]$ is non-empty and has a maximum. Define r_G as that maximum, so $r_G = \max(G(D_G) \cap (0, r])$. Obviously, $r_G \in G(D_G)$ holds as well as $r_G < r$. Choose $v \in (r_G, r]$ and $u \leq r_G$. Furthermore, choose $w \in (r, G(y_{b_r+1})]$, where if $b_r \in \overline{B}$ is the unique index such that $G(y_{b_r}) = r_G$. Now, $F^{-1}(v) \leq F^{-1}(r)$ and $F^{-1}(w) > F^{-1}(r)$ yield $\frac{F^{-1}(w)-F^{-1}(v)}{F^{-1}(v)-F^{-1}(u)} > 0$, whereas $G^{-1}(v) = y_{b_r+1} = G^{-1}(w)$ yields $\frac{G^{-1}(w)-G^{-1}(v)}{G^{-1}(v)-G^{-1}(u)} = 0$, thereby contradicting (8.1). Thus, $F \leq_c G$ is contradicted, which proves the implication $F \leq_c G \Rightarrow (i)$.

The proof of the implication $F \leq_c G \Rightarrow (ii)$ is very similar. We assume $G(D_G) \setminus (F(D_F) \cup [\sup F(D_F), 1)) \neq \emptyset$ and now let r be in that set. We define $r_F = \min(F(D_F) \cap [r, 1))$, which is again existent and finite. We choose u, v and w such that $F(x_{a_r-1}) < u \leq r < v \leq r_F < w$, where $a_r \in A$ is the unique index satisfying $F(x_{a_r}) = r_F$. It can now be shown that $F^{-1}(v) - F^{-1}(u) = 0 < G^{-1}(v) - G^{-1}(u)$ holds, which contradicts (8.1) and therefore also $F \leq_c G$.

For the implication $F \leq_c G \Rightarrow (iii)$, let $(a, b) \in \overline{\underline{A}} \times \overline{\underline{B}}$ with $F(x_{a-1}) = G(y_{b-1})$ and $F(x_a) = G(y_b)$. Now, (iii) follows directly from the definition (8.1) of $F \leq_c G$ by setting $u = G(y_{b-1})$, $v = G(y_b)$ and $w = G(y_{b+1})$. Note that, according to (i) and (ii), $G(y_{b+1}) \leq F(x_{a+1})$ holds (either they are equal or $G(y_{b+1})) \in [\sup F(D_F), 1)$ and $F(x_{a+1}) = 1$), resulting in $F^{-1}(w) = x_{a+1}$.

It remains to be shown that (i), (ii) and (iii) together imply $F \leq_c G$. For that, let $0 < u < v < w < 1$ with $F^{-1}(u) < F^{-1}(w)$ and $G^{-1}(u) < G^{-1}(w)$. By Definition 7.8, there exist pairs $(a_u, b_u), (a_v, b_v), (a_w, b_w) \in R(\leftrightharpoons)$ such that

$$\begin{aligned} F^{-1}(u) &= x_{a_u}, \quad F^{-1}(v) = x_{a_v}, \quad F^{-1}(w) = x_{a_w}, \\ G^{-1}(u) &= y_{b_u}, \quad G^{-1}(v) = y_{b_v}, \quad G^{-1}(w) = y_{b_w}. \end{aligned}$$

Hence, $F \leq_c G$ is equivalent to

$$\frac{x_{a_w} - x_{a_v}}{x_{a_v} - x_{a_u}} \leq \frac{y_{b_w} - y_{b_v}}{y_{b_v} - y_{b_u}} \tag{8.3}$$

holding for all pairs $(a_u, b_u), (a_v, b_v), (a_w, b_w) \in R(\leftrightharpoons)$, which are ordered from low to high and may be equal, and with $a_u < a_w$ and $b_u < b_w$. We prove this in three cases.

Case 1: $F(x_{a_u}) \geq \inf G(D_G)$ and $G(y_{b_w-1}) \leq \sup F(D_F)$.

With these additional assumptions, (i) and (ii) dictate that $F(x_{a_u}) = G(y_{a_u})$, $F(x_{a_u+1}) = G(y_{b_u+1})$, $\ldots$, $F(x_{a_w-1}) = G(y_{b_w-1})$. If the equivalent conditions $a_u = a_v$ and $b_u = b_v$ are fulfilled, both sides of (8.3) are infinite; if the equivalent conditions $a_v = a_w$ and $b_v = b_w$ are fulfilled, both sides of (8.3) vanish. In both cases, inequality (8.3) is fulfilled, so we can assume without restriction $a_u < a_v < a_w$ and $b_u < b_v < b_w$.

Now, (iii) can be invoked for all pairs $(a_u+1, b_u+1), (a_u+2, b_u+2), \ldots, (a_w-1, b_w-1)$. For all $(\alpha, a, \beta, b) \in A^2 \times B^2$ with $a_u+1 \leq \alpha \leq a \leq a_w-1$ and $a_w - a = b_w - b, \alpha - a_u = \beta - b_u$ (so α and a are at the same points in the chain $a_u + 1, \ldots, a_w - 1$ as β and b are in the chain $b_u + 1, \ldots, b_w - 1$), it follows that

$$\frac{x_{a+1} - x_a}{x_\alpha - x_{\alpha-1}} = \prod_{j=\alpha}^{a} \frac{x_{j+1} - x_j}{x_j - x_{j-1}} \leq \prod_{j=\beta}^{b} \frac{y_{j+1} - y_j}{y_j - y_{j-1}} = \frac{y_{b+1} - y_b}{y_\beta - y_{\beta-1}}.$$

Because of the identity

$$\begin{aligned} \frac{x_{a_w} - x_{a_v}}{x_{a_v} - x_{a_u}} &= \sum_{a=a_v}^{a_w-1} \frac{x_{a+1} - x_a}{x_{a_v} - x_{a_u}} = \sum_{a=a_v}^{a_w-1} \left(\sum_{\alpha=a_u+1}^{a_v} \frac{x_\alpha - x_{\alpha-1}}{x_{a+1} - x_a} \right)^{-1} \\ &= \sum_{a=a_v}^{a_w-1} \left(\sum_{\alpha=a_u+1}^{a_v} \left(\frac{x_{a+1} - x_a}{x_\alpha - x_{\alpha-1}} \right)^{-1} \right)^{-1} \end{aligned}$$

and, analogously,

$$\frac{y_{b_w} - y_{b_v}}{y_{b_v} - y_{b_u}} = \sum_{b=b_v}^{b_w-1} \left(\sum_{\alpha=b_u+1}^{b_v} \left(\frac{y_{b+1} - y_b}{y_\beta - y_{\beta-1}} \right)^{-1} \right)^{-1},$$

which are increasing in every summand, (8.3) follows.

Case 2: $F(x_{a_u}) < \inf G(D_G)$ and $G(y_{b_w-1}) \leq \sup F(D_F)$.

It directly follows that $\inf G(D_G) = \min G(D_G)$ as $\inf G(D_G) > 0$. Furthermore, $b_u = \min B$ follows because the minimum is the only element $b \in B$ that satisfies $a_u \leftrightharpoons b$. Note that $F(x_{a_w}) > \min G(D_G)$ holds since, otherwise, $b_w = b_u$ would follow in a similar way. If $F(x_{a_v}) < \min G(D_G)$, then $b_v = b_u$ follows, yielding that the right hand side of (8.3) is infinite, which proves the inequality to be true. Hence, we now assume $F(x_{a_v}) \geq \min G(D_G)$. Define $\tilde{a}_u = \min\{a \in A : F(x_a) \geq \min G(D_G)\}$, which yields $a_u < \tilde{a}_u \leq a_v$. According to assumptions (i) and (ii), we obtain $F(x_{\tilde{a}_u}) = \min G(D_G) = G(y_{b_u})$. It follows that Case 1 can be invoked for the pairs $(\tilde{a}_u, b_u)$, (a_v, b_v) and (a_w, b_w), yielding

$$\frac{x_{a_w} - x_{a_v}}{x_{a_v} - x_{a_u}} \leq \frac{x_{a_w} - x_{a_v}}{x_{a_v} - x_{\tilde{a}_u}} \leq \frac{y_{b_w} - y_{b_v}}{y_{b_v} - y_{b_u}}.$$

Case 3: $G(y_{b_w-1}) > \sup F(D_F)$.

We proceed mostly analogous to Case 2. First, both $\sup F(D_F) = \max F(D_F)$ and $a_w = \max A$ follow directly. Furthermore, $G(y_{b_u}) \leq \max F(D_F)$ holds since, otherwise, $a_u = a_w$ would follow. If $G(y_{b_v}) > \max F(D_F)$, then $a_v = a_w$ follows, yielding that the left hand side of (8.3) is zero, which proves the inequality to be true. We now assume $G(y_{b_v}) \leq \max F(D_F)$ and subsequently define $\tilde{b}_w \in B$ by $\tilde{b}_w - 1 = \max\{b \in B : G(y_b) \leq \max F(D_F)\} < b_w - 1$. Since $G(y_{\tilde{b}_w - 1}) = \max F(D_F) = F(x_{\tilde{a}_w - 1})$ follows due to (i) and (ii) holding, Cases 1 and 2 (which, combined, cover the case $G(y_{b_w-1}) \leq \sup F(D_F)$) can be invoked for the pairs (a_u, b_u), (a_v, b_v) and $(a_w, \tilde{b}_w)$, yielding

$$\frac{x_{a_w} - x_{a_v}}{x_{a_v} - x_{a_u}} \leq \frac{y_{\tilde{b}_w} - y_{b_v}}{y_{b_v} - y_{b_u}} \leq \frac{y_{b_w} - y_{b_v}}{y_{b_v} - y_{b_u}}.$$

▨

Note that properties (i) and (ii) from Theorem 8.1 are equivalent to the following. If we disjointly divide the set $F(D_F) \cup G(D_G)$ into the three sets $S_{FG} = F(D_F) \cap G(D_G)$, $S_F = F(D_F) \setminus G(D_G)$ and $S_G = G(D_G) \setminus F(D_F)$, then $p < r < q$ holds for all $p \in S_F$, $q \in S_G$ and $r \in S_{FG}$. Simply put, there exists a subset of the unit interval, within which the sets $F(D_F)$ and $G(D_G)$ coincide; values smaller or larger than the elements of that common set can only be elements of $F(D_F)$ or $G(D_G)$, respectively. This equivalent characterization is utilized multiple times in the proof of Theorem 8.1.

Since properties (i) and (ii) are very restrictive, and at the same time are necessary conditions for $F \leq_c G$, Theorem 8.1 implies that $\leq_c$ is not a suitable skewness order for discrete distributions. This observation is in agreement with Eberl and Klar (2019). The only notable subclass of $\mathcal{D}_0$, for which $\leq_c$ is a sensible skewness order, is the class of non-tied empirical distributions with the same sample size. In order to obtain a skewness order suitable for all purposive discrete distributions, the definition of $\leq_c$ needs to be modified, similarly to the definition of discrete dispersive orders as a modification of the usual dispersive order in

Section 7.1. In the remainder of this section, we develop a possible definition of a discrete skewness order based on $\leq_c$ and Theorem 8.1.

Property (iii) from Theorem 8.1 already suggests the form of one criterion for two discrete distributions to be ordered with respect to skewness. The quantities to be compared are the ratios of the lengths of two successive constant intervals; it remains to be established for which pairs of indices this comparison is necessary. These pairs in property (iii) of Theorem 8.1 are all $(a, b) \in \overline{\underline{A}} \times \overline{\underline{B}}$ with $F(x_{a-1}) = G(y_{b-1})$ and $F(x_a) = G(y_b)$, which obviously only makes sense under restrictions (i) and (ii). For a suitable generalization of this requirement, consider the following result.

Proposition 8.2. *Let $F, G \in \mathcal{D}_0$ satisfy properties (i) and (ii) from Theorem 8.1. Then, for $(a, b) \in \overline{\underline{A}} \times \overline{\underline{B}}$, all of the following statements are equivalent:*

(A) $F(x_{a-1}) = G(y_{b-1}) \quad \text{and} \quad F(x_a) = G(y_b)$,

(B) $a \leftrightharpoons b$,

(C) $a \leftrightharpoons_\wedge b \quad \text{or} \quad a + 1 \leftrightharpoons_\wedge b + 1$,

(D) $a \leftrightharpoons_\wedge b \quad \text{and} \quad a + 1 \leftrightharpoons_\wedge b + 1$.

Proof. The implication (A)⇒(D) follows directly from Proposition 7.19, the implication (D)⇒(C) is obviously true and the implication (C)⇒(B) holds by definition of the relation $\leftrightharpoons_\wedge$ (see Definition 7.14a)). For the remaining implication (B)⇒(A), let $(a, b) \in (\overline{\underline{A}} \times \overline{\underline{B}}) \cap R(\leftrightharpoons)$. If $\min G(D_G)$ and therefore $\min B$ does not exist, $\inf G(D_G) = 0$ follows; if $\max F(D_F)$ and therefore $\max A$ does not exist, $\sup F(D_F) = 1$ follows.

Case 1: Both $\min B$ and $\max A$ do not exist.
In this case, properties (i) and (ii) dictate that $F(D_F) = G(D_G)$, which yields that (A) and (B) are equivalent.

Case 2: $\min B$ exists, but $\max A$ does not exist.
Properties (i) and (ii) dictate that $G(D_G) \setminus F(D_F) = \emptyset$. Furthermore, since $b > \min B$, $G(y_{b-1}) \geq \min G(D_G)$ follows. Hence, $G(y_{b-1+k}) \in F(D_F) \cap G(D_G)$ holds for all $k \in \mathbb{N}_0$. Because of $a \leftrightharpoons b$, this particularly implies $F(x_{a-1}) = G(y_{b-1})$ and $F(x_a) = G(y_b)$.

Case 3: $\max A$ exists, but $\min B$ does not exist.
The proof here is basically analogous to Case 2. Properties (i) and (ii) dictate that $F(D_F) \setminus G(D_G) = \emptyset$. Furthermore, since $a < \max A$, $F(x_a) \leq \max F(D_F)$ follows. Hence, $F(x_{a-k}) \in F(D_F) \cap G(D_G)$ holds for all $k \in \mathbb{N}_0$. Because of $a \leftrightharpoons b$, this particularly implies $F(x_{a-1}) = G(y_{b-1})$ and $F(x_a) = G(y_b)$.

Case 4: Both $\min B$ and $\max A$ exist.
Here, $G(y_{b-1}) \geq \min G(D_G)$ and $F(x_a) \leq \max F(D_F)$ holds. Since the assumption

$a \leftrightharpoons b$ yields $G(y_{b-1}) < F(x_a)$, it follows that both values are elements of $F(D_F) \cap G(D_G)$. Hence, $F(x_{a-1}) = G(y_{b-1})$ and $F(x_a) = G(y_b)$ holds.

▨

Proposition 8.2 presents us with three possible alternative methods of determining which pairs of indices have to be compared in the sense of (8.2). However, it is not difficult to find cdf's $F, G \in \mathcal{D}_0$ such that the set of pairs obtained through method (D) is empty. Methods (B) and (C) both seem to be sensible choices. The set of pairs obtained through method (C) is a subset of the set obtained through method (B) as the implication (C)$\Rightarrow$(B) obviously holds for all $F, G \in \mathcal{D}_0$. However, since method (C) seems to be sufficient for a number of exemplary pairs of distributions, we use this method in our definition for a discrete skewness order as it requires a smaller number of comparisons. We refrain from attempting to prove any rigorous results advocating the use of method (C), which could possibly be obtained similarly to Proposition 7.19. An upside of using method (B) instead could be that the set of index-pairs to be compared can be computed more easily. Overall, our proposed requirement concerning the supports for $F \in \mathcal{D}_0$ to be deemed less skewed (to the right) than $G \in \mathcal{D}_0$ is

$$\frac{x_{a+1} - x_a}{x_a - x_{a-1}} \leq \frac{y_{b+1} - y_b}{y_b - y_{b-1}} \quad \forall (a, b) \in R(\leftrightharpoons_3), \tag{8.4}$$

where the relation $\leftrightharpoons_3$ on the set $\overline{\underline{A}} \times \overline{\underline{B}}$ is defined by

$$\begin{aligned} a \leftrightharpoons_3 b &\Leftrightarrow (a \leftrightharpoons_\wedge b) \vee (a + 1 \leftrightharpoons_\wedge b + 1) \\ &\Leftrightarrow \ ((a - 1 \leftrightharpoons b - 1) \wedge (a \leftrightharpoons b)) \vee ((a \leftrightharpoons b) \wedge (a + 1 \leftrightharpoons b + 1)) \\ &\Leftrightarrow (a \leftrightharpoons b) \wedge ((a - 1 \leftrightharpoons b - 1) \vee (a + 1 \leftrightharpoons b + 1)) \end{aligned}$$

for $a \in \overline{\underline{A}}, b \in \overline{\underline{B}}$.

However, this requirement alone is not sufficient to deem one discrete distribution unambiguously less skewed than another. Similarly to our process of finding discrete dispersive orders, a requirement concerning the jump heights seems to be missing. Or, in other words, the requirements (i) and (ii) on the jump heights given in Theorem 8.1 need to be weakened. In order to obtain the corresponding requirement for the discrete dispersive orders, we substituted the Lebesgue densities in an equivalent characterization of $F \leq_{disp} G$ on $\mathcal{P}_I^1$ for the pmf's of the discrete distributions in question. That equivalent characterization is given by $g(G^{-1}(p)) \leq f(F^{-1}(p))$ for all $p \in (0, 1)$. On $\mathcal{P}_I^2$, the statement $F \leq_c G$ can be equivalently characterized in a similar way, namely by

$$\frac{g'(G^{-1}(p))}{(g(G^{-1}(p)))^2} \leq \frac{f'(F^{-1}(p))}{(f(F^{-1}(p)))^2} \quad \forall p \in (0, 1)$$

(see the proof of Theorem 3.12c)). Here, the Lebesgue densities cannot simply be substituted for pmf's. However, the fact that the derivatives of Lebesgue densities appear in the above characterization suggests that some sort of comparison between successive values of the pmf's is necessary. Thus, our proposal is to formulate the requirement similarly to (8.4): this results in requiring

$$\frac{q_b}{q_{b-1}} \leq \frac{p_a}{p_{a-1}} \tag{8.5}$$

for all pairs (a, b) in a suitable subset of $\underline{A} \times \underline{B}$. The direction of the $\leq$-sign in (8.5) is obvious by comparing a discrete uniform distribution to a distribution with the same support, but with decreasing pmf. The latter distribution is obviously skewed to the right and the corresponding ratios considered in (8.5) are smaller than 1, while the uniform distribution is obviously symmetric with the corresponding ratios always being equal to 1.

The only thing missing is a suitable subset of $\underline{A} \times \underline{B}$ that contains all pairs (a, b) to be compared in the sense of (8.5). Two already known subsets of $\underline{A} \times \underline{B}$ suitable for a kind of pointwise comparison between two cdf's are $R(\leftrightharpoons_\wedge)$ and $R(\leftrightharpoons_\vee)$. Since the set in association with (8.5) should be a weakening of properties (i) and (ii) from Theorem 8.1, the following result is helpful in choosing between the two presented alternatives.

Proposition 8.3. *Let $F, G \in \mathcal{D}_0$ satisfy properties (i) and (ii) from Theorem 8.1. Then,*

$$\frac{q_b}{q_{b-1}} \leq \frac{p_a}{p_{a-1}} \quad \forall (a, b) \in R(\leftrightharpoons_\wedge) \tag{8.6}$$

follows, whereas

$$\frac{q_b}{q_{b-1}} \leq \frac{p_a}{p_{a-1}} \quad \forall (a, b) \in R(\leftrightharpoons_\vee) \tag{8.7}$$

does generally not hold.

Proof. It is easy to see that $R(\leftrightharpoons_\wedge) = \{(a, b) \in \underline{A} \times \underline{B} : F(x_{a-1}) = G(y_{b-1})\}$. Consequently, for all pairs $(a, b) \in R(\leftrightharpoons_\wedge)$ except for the lowest and the highest (if those exist), $F(x_{a-j}) = G(y_{b-j}), j = 0, 1, 2$ holds, resulting in

$$\frac{q_b}{q_{b-1}} = \frac{G(y_b) - G(y_{b-1})}{G(y_{b-1}) - G(y_{b-2})} = \frac{F(x_a) - F(x_{a-1})}{F(x_{a-1}) - F(x_{a-2})} = \frac{p_a}{p_{a-1}}.$$

$R(\leftrightharpoons_\wedge)$ has a lowest pair, if and only if $\min G(D_G)$ exists, and it has a highest pair, if and only if $\max F(D_F)$ exists. In the first case, if (a, b) denotes the pair in question, $F(x_{a-1}) = G(y_{b-1})$ holds along with $G(y_{b-2}) = 0 \leq F(x_{a-2})$. If $|R(\leftrightharpoons_\wedge)| > 1$, then $F(x_a) = G(y_b)$ holds; otherwise, $F(x_a) = 1 \geq G(y_b)$ holds. It follows

$$\frac{q_b}{q_{b-1}} = \frac{G(y_b) - G(y_{b-1})}{G(y_{b-1}) - G(y_{b-2})} \leq \frac{F(x_a) - F(x_{a-1})}{F(x_{a-1}) - F(x_{a-2})} = \frac{p_a}{p_{a-1}}.$$

In the case of $\max F(D_F)$ existing, the proof is analogous.

For the counterexample to (8.7), let $F \in \mathcal{D}_0$ be defined by $A = \{1, \ldots, 4\}$ and $(x_j, p_j) =$

$(j, \frac{1}{4})$ for $j \in A$, i.e. $X \sim \mathcal{U}[4]$. Furthermore, let $G \in \mathcal{D}_0$ be defined by $B = \{1, \dots, 5\}$, $y_j = j$ for $j \in B$ and $(q_1, \dots, q_5) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{16}, \frac{3}{16})$. It follows that $F(D_F) = \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$ and $G(D_G) = \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{13}{16}\}$, so both (i) and (ii) are satisfied. However, since $(4, 5) \in R(\leftrightharpoons_\vee)$, $\frac{q_5}{q_4} = 3 > 1 = \frac{p_4}{p_3}$ contradicts (8.7). ▨

With this result, our proposed definition of a discrete skewness order arises from the combination of (8.4) and (8.6).

Definition 8.4. Let $F, G \in \mathcal{D}_0$. G is said to be *at least as discretely skewed (to the right) as* F, denoted by $F \leq_{skew}^{disc} G$, if

$$\begin{aligned} \frac{q_b}{q_{b-1}} &\leq \frac{p_a}{p_{a-1}} && \forall (a, b) \in R(\leftrightharpoons_\wedge) \quad \text{and} \\ \frac{x_{a+1} - x_a}{x_a - x_{a-1}} &\leq \frac{y_{b+1} - y_b}{y_b - y_{b-1}} && \forall (a, b) \in R(\leftrightharpoons_3). \end{aligned}$$

Corollary 8.5. *Let $F, G \in \mathcal{D}_0$ satisfy properties (i) and (ii) from Theorem 8.1. Then, $F \leq_c G$ and $F \leq_{skew}^{disc} G$ are equivalent.*

Proof. The result follows from the combination of Theorem 8.1 and Propositions 8.2 and 8.3. ▨

Similarly to our considerations concerning discrete dispersion orders, this order could be analyzed further. This includes the derivation of desirable properties, its compatibility with popular distribution families and skewness measures as well as the discussion of alternative orders. However, for this outlook, the proposal of a discrete skewness order should suffice.

8.3. Comparing Discrete Distributions

Generally, quantiles are not an optimal tool for describing and comparing discrete distributions. They seem to suffice for the relatively simple concept of location, which can be described by the calculation and comparison of single quantiles (see Section 6.1). However, as soon as the characteristic of interest requires differences of quantiles to be considered, quantile-based tools like the Q-Q-plots are no longer satisfactory. This is shown for the characteristic of dispersion in Section 6.2 and it is indicated in Section 8.2 for skewness.

One alternative way to describe distributions is to consider expectiles instead of quantiles. Because all expectile functions e_F are strictly increasing on $(\inf(\text{supp}(F)), \sup(\text{supp}(F)))$ (see Proposition 2.22c)), the disadvantages of quantiles are not present for expectiles. This is because expectiles are quantiles of the transformed cdf $\breve{F}$ defined in (2.14), which lies in the set $\mathcal{P}_I$ for all underlying distributions. Thus, the problems with discrete distributions are eliminated by transforming them into sufficiently similar continuous distributions. On the other hand, expectiles do not describe distributions in the same fundamental and intuitive sense as quantiles, making them more difficult to interpret.

Another approach at describing and comparing discrete distributions is developed throughout Chapter 7. Crucial points in this development are given by the restriction to purposive discrete distributions in Definition 7.1, their equivalent description by indexing sets and identifying sequences in Proposition 7.3, and by the definition and characterization of the relation $\leftrightharpoons$ in Definition 7.8 and Proposition 7.10. These tools allow us to describe and handle the discrete dispersive orders in a fairly simple way. The relation $\leftrightharpoons$ in particular modifies the meaning of pointwise comparisons to be meaningfully applicable to discrete distributions. It does so in a way that is informed by the Q-Q-plot (see Figure 6.3, Example 7.11a) and the subsequent remarks), and is therefore connected to the methodology used in a continuous setting.

Within this thesis, the theory for the comparison of discrete distributions described above is mainly used for the characteristic of dispersion. This begs the question whether it is generally suitable for (convex) characteristics of discrete distributions. Lemma A.12 states that the stochastic order $\leq_{st}$ can easily be described using the relation $\leftrightharpoons$. Furthermore, Section 8.2 contains an approach to use it for the construction of a discrete version of the skewness order $\leq_c$. Although the resulting Definition 8.4 is more complex than the discrete orders of location and dispersion, the requirements are of a similar form and are based on the relation $\leftrightharpoons$. Since discrete skewness is only discussed briefly, the utility of the discrete skewness order is difficult to assess. However, it is possible that convex characteristics of an even higher order could be described in a similar way. For the example of kurtosis, it would be interesting to see whether the problems that arise from a lack of standardization with respect to skewness in Chapter 4 also play a role in these discrete orders.

The framework for comparing discrete distributions from Chapter 7 also exhibits a few disadvantages. First, the first alternative approach for a discrete dispersive order presented in Section 8.1 looks promising, but it does not seem to fit into this framework. Of course, this could also be seen as a weakness of the proposed order. Second, the discrete orders of convex characteristics cannot be equivalently characterized by simple functions like RIDF's. A generalization of RIDF's developed in Appendix A succeeds in characterizing the stochastic order, but not the discrete dispersive orders. The third disadvantage of the methodology is that it is not applicable to all discrete distributions, but only to the subset $\mathcal{D}_0$. Example 7.2 gives a discrete distribution outside of $\mathcal{D}_0$. Although virtually all practically useful discrete distributions are in $\mathcal{D}_0$, this shows that the methodology is somewhat limited. However, for some $F \in \mathcal{D} \setminus \mathcal{D}_0$, it seems possible to divide F into multiple distributions in $\mathcal{D}_0$, which are then treated separately and can finally be added up to again obtain F.

Finally, it is unclear whether the transition from the discrete dispersive orders to the original dispersive order is smooth in some sense. For example, one could consider a continuous cdf F and approximate it by discrete cdf's F_n such that $F_n \overset{n\to\infty}{\to} F$ in a given mode of convergence. If the same is given for another continuous cdf G, a desirable result would be that $F_n \leq_{disp}^{\wedge-disc} G_n$ for all sufficiently large n implies $F \leq_{disp} G$. This and similar open questions could be the topic of future research.

APPENDIX A

Relative Inverse Distribution Functions of Discrete Distributions

In this appendix, we consider RIDF's and their application to discrete distributions. The goal is to obtain exact limits of their applicability and propose a suitable modification. Throughout the appendix, we assume $F \mathrel{\hat{=}} (A, (x_j, p_j)_{j\in A})$ and $G \mathrel{\hat{=}} (B, (y_j, q_j)_{j\in B})$ whenever $F, G \in \mathcal{D}_0$ (see Proposition 7.3).

For two cdf's F and G, the corresponding RIDF's are given by $R_{GF} = F^{-1} \circ G$ and $R_{FG} = G^{-1} \circ F$. Under certain regularity conditions (e.g. $F, G \in \mathcal{P}_I$), these mappings are strictly increasing functions that transform one random variable into another. Specifically, if $X \sim F$ and $Y \sim G$, then $R_{GF}(Y) =_{st} X$ and $R_{FG}(X) =_{st} Y$. This is explicitly proved in the following (see also van Zwet, 1964, p. 48, and Oja, 1981, p. 156).

Proposition A.1. *Let $F, G \in \mathcal{P}_I$. Then, $\varphi(X) =_{st} Y$ with $\varphi : D_F \to D_G$ strictly increasing is equivalent to $\varphi = R_{FG}$.*

Proof. Preliminarily note that, if φ is bijective, we obtain

$$H_{\varphi(X)}(t) = \mathbb{P}(\varphi(X) \le t) = \mathbb{P}(X \le \varphi^{-1}(t)) = F(\varphi^{-1}(t)) \tag{A.1}$$

for all $t \in D_G$. For the implication from left to right, it follows from φ being strictly increasing that it is also injective. For the surjectivity of φ, note that both $\varphi(D_F)$ and D_G are open intervals. Thus, assuming $\varphi(D_F) \subsetneq D_G$ would imply $\lambda^1(\varphi(D_F)) < \lambda^1(D_G)$, which contradicts $\varphi(X) =_{st} Y$. This means that (A.1) holds and therefore, $\varphi(X) =_{st} Y$ is equivalent to $F(\varphi^{-1}(t)) = G(t)$ for all $t \in D_G$. Since F, G and φ are all strictly increasing, this is equivalent to $\varphi^{-1} = F^{-1} \circ G$ and also to $\varphi = G^{-1} \circ F$. For the direction from right to left, note that φ is strictly increasing as a composition of strictly increasing functions. Since $\varphi(D_F) = D_G$

holds by assumption, φ is bijective. Due to (A.1), we obtain

$$H_{\varphi(X)}(t) = F(\varphi^{-1}(t)) = F(F^{-1}(G(t))) = G(t),$$

which, as noted before, is equivalent to $\varphi(X) =_{st} Y$. ▨

The domain and the codomain of a RIDF are chosen in such a way that they almost surely contain all values that X and Y take, respectively. In the continuous setting in Proposition A.1, the interiors D_F and D_G of the supports $\text{supp}(F)$ and $\text{supp}(G)$ are chosen, because using the supports themselves would lead to problems with the surjectivity of φ. However, if we are instead dealing with discrete distributions, the supports $\text{supp}(F)$ and $\text{supp}(G)$ would be the canonical choices for the domain and the codomain since they are then a union of singleton atoms instead of a closed interval.

The requirement that φ is strictly increasing is made in order to obtain a intuitive transformation, meaning that the probability mass on low values of X is transformed onto low values of G and the mass on high values is transformed onto high values. It prevents unnecessary rearrangements leading to the same result. Consider the following simple example.

Example A.2. Let $X \sim \mathcal{U}(0,2)$ and $Y \sim \mathcal{U}(2,4)$. Because of $F, G \in \mathcal{P}_I$, R_{FG} is the unique strictly increasing transformation that returns a G-distributed random variable, if applied to X. Because of $F : [0,2] \to [0,1], t \mapsto \frac{t}{2}$, and $G : [2,4] \to [0,1], t \mapsto \frac{t}{2} - 1$, we obtain $R_{FG}(t) = G^{-1}(\frac{t}{2}) = 2\left(\frac{t}{2} + 1\right) = t + 2$. However, without the transformation being required to be strictly increasing, one could, e.g., also propose $\varphi(t) = (t+3)\mathbb{1}_{[0,1]}(t) + (t+1)\mathbb{1}_{(1,2]}(t)$. Because of

$$\begin{aligned}\mathbb{P}(\varphi(X) \le t) &= \mathbb{P}(X \le 1, X \le t-3) + \mathbb{P}(X > 1, X \le t-1)\\ &= \frac{t-3}{2}\mathbb{1}_{[3,4]}(t) + \frac{t-2}{2}\mathbb{1}_{[2,3]}(t) + \frac{1}{2}\mathbb{1}_{(3,4]}(t) = \frac{t}{2} - 1 = G(t)\end{aligned}$$

for all $t \in [2,4]$, this transformation also satisfies $\varphi(X) =_{st} Y$. However, since it is not increasing, the nature of the transformation is non-monotone and very unintuitive, as illustrated by Figure A.1.

Since, under some regularity conditions, R_{FG} is the unique strictly increasing function that transforms a F-distributed random variable into a G-distributed random variable, our next step is to see whether this transformation works for all $F, G \in \mathcal{P}$. The following proposition not only states that this is, in general, not true, but also gives a necessary and sufficient condition for its validity. The result and its proof are improved versions of Lemma 1c) from Eberl and Klar (2019, p. 265). In particular, a number of equations in the proof are taken directly from there.

Proposition A.3. *Let $F, G \in \mathcal{P}$. Then, $R_{FG}(X) =_{st} Y$ holds, if and only if $G(D_G) \subseteq \overline{F(D_F)}$.*

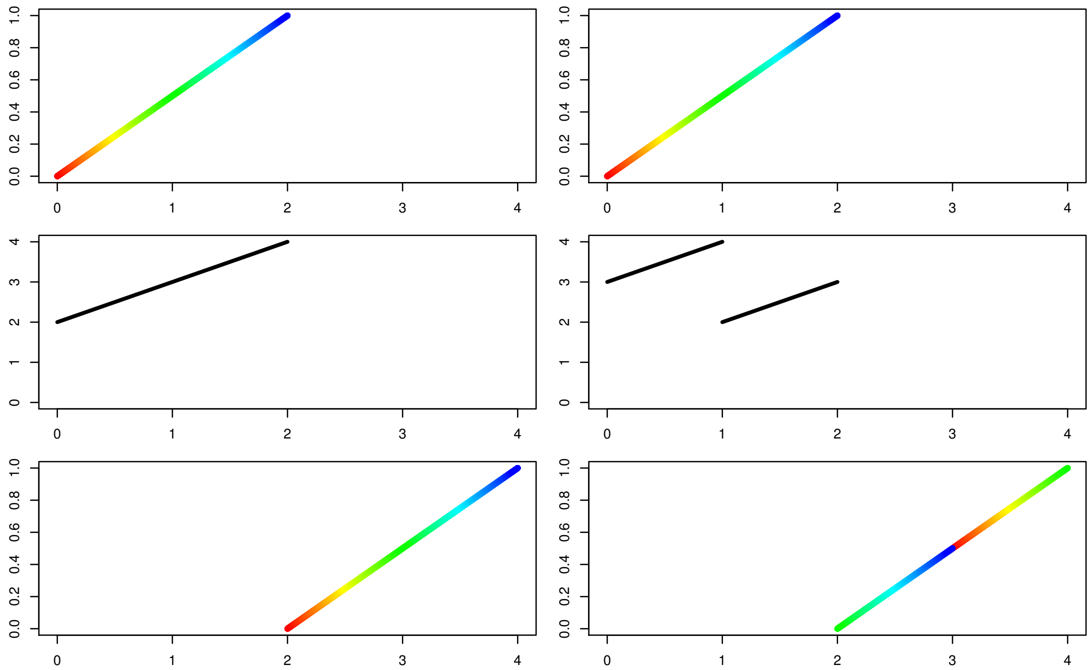

Figure A.1.: Illustration of Example A.2 with the transformation via $G^{-1} \circ F$ on the left and the transformation via φ on the right. Upper panels: Cdf of X with the colour signifying the order of the values (low values red, high values blue). Central panels: Transformation function. Lower panels: Cdf of the transformation of X with the colour signifying the order of the values as before.

Proof. Preliminarily, note that the following three equivalences hold for any cdf $H \in \mathcal{P}$:

$$H^{-1}(H(t)) = t \iff \forall s \in D_H, s < t : H(s) < H(t) \tag{A.2}$$

for all $t \in D_H$,

$$H(H^{-1}(p)) = p \iff p \in H(\mathbb{R}), \tag{A.3}$$

for all $p \in (0, 1)$, and

$$p \leq H(t) \iff H^{-1}(p) \leq t \tag{A.4}$$

for all $p \in (0, 1)$ and $t \in \mathbb{R}$ (see Shorack and Wellner, 1986, pp. 5–6). Considering (A.4), the cdf of $R_{FG}(X)$ is given by

$$H_{R_{FG}(X)}(t) = \mathbb{P}(G^{-1}(F(X)) \leq t) = \mathbb{P}(F(X) \leq G(t)) = H_{F(X)}(G(t)) \tag{A.5}$$

for $t \in \mathbb{R}$. If $t \notin D_G$, it immediately follows that either $G(t) = 0$ and

$$H_{R_{FG}(X)}(t) = H_{F(X)}(0) = \mathbb{P}(F(X) \leq 0) = \mathbb{P}(F(X) = 0) = 0$$

hold (since X does not have any probability mass where F is zero), or $G(t) = 1$ and

$$H_{R_{FG}(X)}(t) = H_{F(X)}(1) = \mathbb{P}(F(X) \leq 1) = 1$$

hold. Hence, $H_{R_{FG}(X)}(t) = G(t)$ for $t \in \mathbb{R} \setminus D_G$.

For the case $t \in D_G$, we first focus on the cdf of $F(X)$ at a point $p \in (0, 1)$. If $p \in F(D_F) = F(\mathbb{R}) \setminus \{0, 1\}$, we obtain

$$\begin{aligned} H_{F(X)}(p) &= \mathbb{P}(F(X) \leq p) = \mathbb{P}(F^{-1}(F(X)) \leq F^{-1}(p)) \\ &= \mathbb{P}(X \leq F^{-1}(p)) = F(F^{-1}(p)) = p. \end{aligned}$$

Here, the second equality follows by combining (A.3) and (A.4), using $p \in F(\mathbb{R})$. The third equality follows from (A.2) by using that X almost surely does not take any realization that lies outside of D_F, or on which F is constant. The last inequality holds due to (A.3) and $p \in F(\mathbb{R})$.

Now consider the case $p \in \overline{F(D_F)} \setminus F(D_F)$. Since any cdf H satisfies $\lim_{t \to -\infty} H(t) = 0$ and $\lim_{t \to \infty} H(t) = 1$ and has at most countably many discontinuities, $H(D_H)$ is a countable union of disjoint intervals. This means that p is an endpoint of one of these intervals (now relating to $F(D_F)$ instead of $H(D_H)$) that is excluded from the interval itself. Since any cdf is right continuous, the left endpoint is always included in the interval. Hence, there exists an interval $I \subseteq F(D_F)$, of which p is the right endpoint. It follows

$$H_{F(X)}(p) = \lim_{\substack{r \in I \\ r \nearrow p}} \mathbb{P}(F(X) \leq r) + \mathbb{P}(F(X) = p) = \lim_{\substack{r \in I \\ r \nearrow p}} r = p,$$

where $\mathbb{P}(F(X) = p) = 0$ is implied by $p \notin F(\mathbb{R})$. Hence, $H_{F(X)}(p) = p$ for $p \in \overline{F(D_F)}$. Now let $p \notin \overline{F(D_F)}$. Since $F(D_F) \cap (\sup(F(D_F) \cap [0, p]), p] = \emptyset$, we obtain

$$H_{F(X)}(p) = \mathbb{P}(F(X) \leq p) = \mathbb{P}(F(X) \leq \underbrace{\sup(F(D_F) \cap [0, p])}_{\in \overline{F(D_F)}}) = \sup(F(D_F) \cap [0, p]) < p.$$

Therefore, $H_{F(X)}(p) = p$ holds, if and only if $p \in \overline{F(D_F)}$ (see Shorack and Wellner, 1986, p. 5, Proposition 2, proof as exercise). Because of (A.5), this means that $H_{R_{FG}(X)}(t) = G(t)$ is equivalent to $G(t) \in \overline{F(D_F)} \cup \{0, 1\}$. Considering that $R_{FG}(X) =_{st} Y$ is equivalent to $H_{R_{FG}(X)}(t) = G(t)$ for all $t \in \mathbb{R}$, this concludes the proof. ▨

Note that the sufficient condition for the existence of an increasing transformation function of one random variable into another, given in Proposition A.3, is highly similar to the necessary condition for being ordered with respect to the dispersive order, given in Proposition 6.6.

Now we turn our attention to purposive discrete distributions. Since the only possible accumulation points of $H(D_H)$ with $H \in \mathcal{D}_0$ are 0 and 1, we obtain $\overline{H(D_H)} \cap (0,1) = H(D_H)$. Hence, for $F, G \in \mathcal{D}_0$, the equivalent characterization in Proposition A.3 is given by $G(D_G) \subseteq F(D_F)$. As discussed after Proposition 6.6, this condition poses a major restriction for pairs of purposive discrete cdf's. This suggests that, for most $F, G \in \mathcal{D}_0$, there exists no transformation function $\varphi : \operatorname{supp}(F) \to \operatorname{supp}(G)$ such that $\varphi(X) =_{st} Y$. The following proposition proves this result in a more rigorous way.

Proposition A.4. *Let $F, G \in \mathcal{D}_0$. Then, there exists an increasing function $\varphi : \operatorname{supp}(F) \to \operatorname{supp}(G)$ such that $\varphi(X) =_{st} Y$, if and only if $G(D_G) \subseteq F(D_F)$.*

Proof. The implication from right to left holds due to Proposition A.3 and since, as stated above, $\overline{F(D_F)} \cap (0,1) = F(D_F)$. For the implication from left to right, assume $G(D_G) \not\subseteq F(D_F)$ or, equivalently, $G(D_G) \setminus F(D_F) \neq \emptyset$. Since the support of X is countable and φ relocates the points in the support in an increasing way, we obtain

$$\begin{aligned} H_{\varphi(X)}(t) &= \mathbb{P}(\varphi(X) \leq t) = \mathbb{P}(X \in \varphi^{-1}((-\infty, t])) = \mathbb{P}(X \leq \sup\{s \in \operatorname{supp}(F) : \varphi(s) \leq t\}) \\ &= F(\sup\{s \in \operatorname{supp}(F) : \varphi(s) \leq t\}) \end{aligned}$$

for any $t \in \mathbb{R}$. Therefore, $H_{\varphi(X)}(\mathbb{R}) \subseteq F(\mathbb{R}) = F(D_F) \cup \{0,1\}$ holds. Now, $G(D_G) \subseteq G(\mathbb{R})$ yields

$$G(\mathbb{R}) \setminus H_{\varphi(X)}(\mathbb{R}) \supseteq G(D_G) \setminus (F(D_F) \cup \{0,1\}) = G(D_G) \setminus F(D_F) \neq \emptyset,$$

which contradicts $H_{\varphi(X)}(t) = G(t)$ for all $t \in \mathbb{R}$ or, equivalently, $\varphi(X) =_{st} Y$. ▨

As shown throughout this thesis, the concept of RIDF's is highly relevant for the comparison of probability distributions. This role is based on the crucial property $R_{FG}(X) =_{st} Y$, which implies that $\Delta_{FG}(X) = R_{FG}(X) - X$ is the difference of a G-distributed and an F-distributed random variable. The function Δ_{FG} can then be utilized to obtain useful and easy to interpret equivalent characterizations of orders of convex characteristics, see Proposition 2.9.

However, this concept only works smoothly, if the RIDF in question actually succeeds in transforming one random variable into the other. In order to meaningfully apply this useful concept to all purposive discrete distributions, we introduce modified RIDF's for this case. Afterwards, we examine whether the new concept has the desired properties and whether it is connected to orders of convex characteristics or their discrete versions. Preliminarily, for two arbitrary sets S and T, denote the the set of all functions from S to T by $\mathrm{M}(S, T)$.

Definition A.5. Let $F, G \in \mathcal{D}_0$. A random element φ on $\mathrm{M}(\mathrm{supp}(F), \mathrm{supp}(G))$ is said to be a *random relative inverse distribution function (RRIDF) from F to G*, if it satisfies

$$\mathbb{P}(\varphi(x_a) = y_b) = \frac{r_{(a,b)}}{p_a}$$

for all $a \in A$ and $b \in B$ and with $r_{(a,b)}$ being defined as in (7.6). We denote the set of all RRIDF's from F to G by Φ_F^G.

Since, for $F, G \in \mathcal{D}_0$, $\mathrm{supp}(F)$ is countable, the set $\mathrm{M}(\mathrm{supp}(F), \mathrm{supp}(G))$ is isomorphic to the set of sequences $(\mathrm{supp}(G))^A = \{(y_a)_{a\in A} : y_a \in \mathrm{supp}(G)\ \forall a \in A\}$ via the isomorphism $\varphi \mapsto (\varphi(x_a))_{a\in A}$. Therefore, any RRIDF from F to G can also be interpreted as a sequence of random variables, indexed by A. Moreover, all of these random variables have themselves purposive discrete distributions since they only take values in $\mathrm{supp}(G)$, which is order-isomorphic to a subset of $\mathbb{Z}$ by assumption.

With that in mind, the distribution of any random element φ in $\mathrm{M}(\mathrm{supp}(F), \mathrm{supp}(G))$ is given by the joint distribution of the sequence $(\varphi(x_a))_{a\in A}$. However, since the defining condition of a RRIDF only specifies the marginal distributions of that sequence, there are generally multiple RRIDF's for one pair of distributions F and G. These different elements of Φ_F^G are set apart by having a variety of dependence structures between the individual components of the sequences $(\varphi(x_a))_{a\in A}$. One of the simplest and easiest to use elements of Φ_F^G is usually obtained, if each component of the sequence has the required marginal distribution and all components are stochastically independent. This observation, along with the facts that $0 \le r_{(a,b)} \le 1$ and $\sum_{b\in B} r_{(a,b)} = \sum_{b\in B: a \rightleftharpoons b} r_{(a,b)} = p_a$ for all $a \in A$ and $b \in B$, already proves that $\Phi_F^G \neq \emptyset$ for all $F, G \in \mathcal{D}_0$.

Next, we show that, for $F, G \in \mathcal{D}_0$, Φ_F^G already contains all random transformations of $X \sim F$ into $Y \sim G$ that are reasonably 'intuitive', which means that they transform the probability mass in an increasing way. However, requiring the transformation function φ to be almost surely increasing in the sense that $\mathbb{P}(\varphi(x_0) \le \varphi(x_1)) = 1$ for all $x_0, x_1 \in \mathrm{supp}(F)$ with $x_0 < x_1$ is not sufficient. It still allows for non-intuitive transformations, if the joint distribution of the components of φ is chosen accordingly, as demonstrated in the following example.

Example A.6. Let $X \sim \mathcal{U}(\{1, 2\})$ and $Y \sim \mathcal{U}(\{1, 2, 3, 4\})$. Define φ as a random element on $\mathrm{M}(\mathrm{supp}(F), \mathrm{supp}(G))$ by

$$\mathbb{P}(\varphi(1) = 1, \varphi(2) = 2) = \mathbb{P}(\varphi(1) = 3, \varphi(2) = 4) = \tfrac{1}{2}.$$

Hence, $\varphi(1) \sim \mathcal{U}(\{1, 3\})$ and $\varphi(2) \sim \mathcal{U}(\{2, 4\})$. Obviously, φ does not adhere to the guideline that low values of X are transformed onto low values of Y and the same with high values. Still, $\varphi(X) =_{st} Y$ is easily seen to be true and the formal requirement of φ being almost surely increasing is satisfied because $\varphi(1) < \varphi(2)$ holds for the two values of $(\varphi(1), \varphi(2))$ with probability mass on them.

This example shows that the property of being almost surely increasing is not sufficient for a random transformation between purposive discrete distributions to behave intuitively. Instead, we use a stronger property that is defined as follows.

Definition A.7. Let S and T be countable sets and let φ be a random element on $\mathrm{M}(S,T)$. Then, φ is said to be *increasing with respect to the marginal distributions (MD-increasing)*, if

$$\mathbb{P}(\varphi(x_\ell) = y_u) = 0 \quad \text{or} \quad \mathbb{P}(\varphi(x_u) = y_\ell) = 0$$

holds for all $x_\ell, x_u \in S$ with $x_\ell < x_u$ and all $y_\ell, y_u \in T$ with $y_\ell < y_u$.

With this definition, we can now show that the set of all 'intuitive' transformation functions of one random variable into another is equal to the corresponding set of RRIDF's.

Theorem A.8. *Let $F, G \in \mathcal{D}_0$ and let φ be a random element on $\mathrm{M}(\mathrm{supp}(F), \mathrm{supp}(G))$ that is independent from X and that is MD-increasing. Then, the following equivalence holds:*

$$\varphi(X) =_{st} Y \Longleftrightarrow \varphi \in \Phi_F^G.$$

Proof. For the implication from right to left, let $t \in \mathbb{R}$. If $\mathrm{supp}(G) \cap (-\infty, t] \neq \emptyset$ and $\sup(\mathrm{supp}(G) \cap (-\infty, t]) < \infty$, let $b_t \in B$ be the unique index that satisfies $y_{b_t} = \sup(\mathrm{supp}(G) \cap (-\infty, t])$. If $\mathrm{supp}(G) \cap (-\infty, t] = \emptyset$, let $b_t = -\infty$; if $\sup(\mathrm{supp}(G) \cap (-\infty, t]) = \infty$, let $b_t = \infty$. It follows that

$$\begin{aligned}
H_{\varphi(X)}(t) &= \mathbb{P}(\varphi(X) \leq t) = \sum_{a \in A} \mathbb{P}(\varphi(X) \leq t | X = x_a) \mathbb{P}(X = x_a) \\
&= \sum_{a \in A} \mathbb{P}(\varphi(x_a) \leq t) \mathbb{P}(X = x_a) = \sum_{a \in A} p_a \sum_{\substack{y \in \mathrm{supp}(G): \\ y \leq t}} \mathbb{P}(\varphi(x_a) = y) \\
&= \sum_{a \in A} p_a \sum_{\substack{j \in B: \\ j \leq b_t}} \frac{r_{(a,j)}}{p_a} = \sum_{\substack{(a,j) \in R(\Leftarrow): \\ j \leq b_t}} r_{(a,j)} \\
&= \sum_{\substack{j \in B: \\ j \leq b_t}} \sum_{\substack{a \in A: \\ a \Leftarrow j}} r_{(a,j)} = \sum_{\substack{j \in B: \\ j \leq b_t}} q_j \\
&= G(y_{b_t}) = G(t).
\end{aligned}$$

This only leaves the implication from left to right to prove; for that, let $b \in B$. Then, there exists an $a \in A$ such that $\mathbb{P}(\varphi(x_a) = y_b) > 0$ since otherwise, $\mathbb{P}(\varphi(X) = y_b) = 0$ would follow, which contradicts $\varphi(X) =_{st} Y$ because of $\mathbb{P}(Y = y_b) = q_b > 0$. Denote the non-empty set of all such a's by

$$A_b = \{a \in A : \mathbb{P}(\varphi(x_a) = y_b) > 0\}.$$

Assume now that A_b does not have a minimum. Because of $A_b \subseteq A \subseteq \mathbb{Z}$, this is equivalent to $\inf A_b = -\infty$. If we additionally assume that there exists a $b_0 \in B, b_0 < b$, then, for all $a_0 \in A_{b_0}$, there exists an $a \in A_b$ with $a < a_0$. However, this contradicts the assumption that φ is MD-non-decreasing, which means that $b = \min B$ follows, if A_b does not have a minimum. Analogously, $b = \max B$ follows if A_b does not have a maximum.

Now we consider the case $b \in \underline{\overline{B}}$, meaning that both $\min A_b$ and $\max A_b$ exist. For all $b_0 \in B, b_0 < b$, the fact that φ is MD-increasing dictates that $a_0 \leq \min A_b$ for all $a_0 \in A_{b_0}$. It follows

$$\begin{aligned}
\mathbb{P}(\varphi(X) < y_b) &= \sum_{\substack{b_0 \in B: \\ b_0 < b}} \mathbb{P}(\varphi(X) = y_{b_0}) \\
&= \sum_{\substack{b_0 \in B: \\ b_0 < b}} \sum_{a \in A} \mathbb{P}(\varphi(X) = y_{b_0} | X = x_a)\mathbb{P}(X = x_a) \\
&= \sum_{\substack{b_0 \in B: \\ b_0 < b}} \sum_{a \in A} \underbrace{\mathbb{P}(\varphi(x_a) = y_{b_0})}_{=0 \text{ if } a \notin A_{b_0}} p_a \\
&= \sum_{a \in \bigcup_{b_0 \in B: b_0 < b} A_{b_0}} p_a \sum_{\substack{b_0 \in B: \\ b_0 < b}} \mathbb{P}(\varphi(x_a) = y_{b_0}) \\
&= \sum_{a \in \left(\bigcup_{b_0 \in B: b_0 < b} A_{b_0}\right) \setminus \{\min A_b\}} p_a \sum_{\substack{b_0 \in B: \\ b_0 < b}} \mathbb{P}(\varphi(x_a) = y_{b_0}) \\
&\quad + \mathbb{1}_{\bigcup_{b_0 \in B: b_0 < b} A_{b_0}}(\min A_b) \cdot p_{\min A_b} \sum_{\substack{b_0 \in B: \\ b_0 < b}} \mathbb{P}(\varphi(x_{\min A_b}) = y_{b_0})
\end{aligned} \tag{A.6}$$

Let $a \in \left(\bigcup_{b_0 \in B: b_0 < b} A_{b_0}\right) \setminus \{\min A_b\}$. Then all $b_0 \in B$ such that $\mathbb{P}(\varphi(x_a) = y_{b_0}) > 0$ are smaller than b, yielding $\sum_{b_0 \in B: b_0 < b} \mathbb{P}(\varphi(x_a) = y_{b_0}) = 1$. If $\min A_b \notin \bigcup_{b_0 \in B: b_0 < b} A_{b_0}$, then there exists no $b_0 \in B, b_0 < b$, such that $\mathbb{P}(\varphi(x_{\min A_b}) = y_{b_0}) > 0$, yielding $\sum_{b_0 \in B: b_0 < b} \mathbb{P}(\varphi(x_{\min A_b}) = y_{b_0}) = 0$. This renders the characteristic function in (A.6) redundant. Furthermore, if $\min A_b \in \bigcup_{b_0 \in B: b_0 < b} A_{b_0}$, we obtain $\sum_{b_0 \in B: b_0 < b} \mathbb{P}(\varphi(x_{\min A_b}) = y_{b_0}) = \mathbb{P}(\varphi(x_{\min A_b}) < y_b)$. Overall, this yields

$$\begin{aligned}
\mathbb{P}(\varphi(X) < y_b) &= \sum_{a \in \left(\bigcup_{b_0 \in B: b_0 < b} A_{b_0}\right) \setminus \{\min A_b\}} p_a + p_{\min A_b} \mathbb{P}(\varphi(x_{\min A_b}) < y_b) \\
&= F(x_{\min A_b - 1}) + p_{\min A_b} \mathbb{P}(\varphi(x_{\min A_b}) < y_b) \\
&= F(x_{\min A_b}) - p_{\min A_b} \mathbb{P}(\varphi(x_{\min A_b}) \geq y_b).
\end{aligned} \tag{A.7}$$

With analogous reasoning, we obtain

$$\begin{aligned}\mathbb{P}(\varphi(X) > y_b) &= \sum_{a\in\left(\bigcup_{b_0\in B:b_0>b} A_{b_0}\right)\setminus\{\max A_b\}} p_a + p_{\max A_b}\mathbb{P}(\varphi(x_{\max A_b}) > y_b) \\ &= 1 - F(x_{\max A_b}) + p_{\max A_b}\mathbb{P}(\varphi(x_{\max A_b}) > y_b) \qquad \text{(A.8)} \\ &= 1 - F(x_{\max A_b - 1}) - p_{\max A_b}\mathbb{P}(\varphi(x_{\max A_b}) \leq y_b).\end{aligned}$$

It follows

$$\begin{aligned}F(x_{\min A_b - 1}) &= \mathbb{P}(\varphi(X) < y_b) - p_{\min A_b}\mathbb{P}(\varphi(x_{\min A_b}) < y_b) \\ &\leq \mathbb{P}(\varphi(X) \leq y_{b-1}) = \mathbb{P}(Y \leq y_{b-1}) = G(y_{b-1}), \qquad \text{(A.9)} \\ F(x_{\min A_b}) &= \mathbb{P}(\varphi(X) < y_b) + p_{\min A_b}\mathbb{P}(\varphi(x_{\min A_b}) \geq y_b) \\ &\geq \mathbb{P}(\varphi(X) \leq y_{b-1}) + p_{\min A_b}\mathbb{P}(\varphi(x_{\min A_b}) = y_b) > \mathbb{P}(\varphi(X) \leq y_{b-1}) = G(y_{b-1}), \\ & \qquad \text{(A.10)} \\ F(x_{\max A_b}) &= 1 - \mathbb{P}(\varphi(X) > y_b) + p_{\max A_b}\mathbb{P}(\varphi(x_{\max A_b}) > y_b) \\ &\geq \mathbb{P}(\varphi(X) \leq y_b) = \mathbb{P}(Y \leq y_b) = G(y_b), \qquad \text{(A.11)} \\ F(x_{\max A_b - 1}) &= 1 - \mathbb{P}(\varphi(X) > y_b) - p_{\max A_b}\mathbb{P}(\varphi(x_{\max A_b}) \leq y_b) \\ &\leq \mathbb{P}(\varphi(X) \leq y_b) - p_{\max A_b}\mathbb{P}(\varphi(x_{\max A_b}) = y_b) < \mathbb{P}(\varphi(X) \leq y_b) = G(y_b). \quad \text{(A.12)}\end{aligned}$$

yielding that, for all $a \in A$, $a \leftrightharpoons b$ is equivalent to $\min A_b \leq a \leq \max A_b$, so to $a \in A_b$. This directly implies that, for all $a \in A \setminus A_b$,

$$\mathbb{P}(\varphi(x_a) = y_b) = 0 = r_{(a,b)} = \frac{r_{(a,b)}}{p_a},$$

leaving only $a \in A_b$ to be considered. We now distinguish between two cases concerning the cardinal number of A_b.

Case 1: $|A_b| = 1$

Let a_b denote the one element of A_b. Then, (A.7) and (A.8) yield

$$\begin{aligned}\mathbb{P}(\varphi(X) = y_b) &= 1 - \mathbb{P}(\varphi(X) > y_b) - \mathbb{P}(\varphi(X) < y_b) \\ &= F(x_{a_b}) - F(x_{a_b - 1}) - p_{a_b}\left[\mathbb{P}(\varphi(x_{a_b}) > y_b) + \mathbb{P}(\varphi(X) < y_b)\right] \\ &= p_{a_b}\left[1 - \mathbb{P}(\varphi(x_{a_b}) < y_b) - \mathbb{P}(\varphi(x_{a_b}) > y_b)\right] \\ &= p_{a_b}\mathbb{P}(\varphi(x_{a_b}) = y_b). \qquad \text{(A.13)}\end{aligned}$$

As noted before the case distinction, a_b is the only element $a \in A$ that satisfies $a \leftrightharpoons b$, which implies $r_{(a_b,b)} = q_b$. By plugging this into (A.13), we obtain

$$\mathbb{P}(\varphi(x_{a_b}) = y_b) = \frac{\mathbb{P}(\varphi(X) = y_b)}{p_{a_b}} = \frac{\mathbb{P}(Y = y_b)}{p_{a_b}} = \frac{q_b}{p_{a_b}} = \frac{r_{(a_b,b)}}{p_{a_b}}.$$

This proves the assertion in this case.

Case 2: $|A_b| \geq 2$

In this case, (A.7), (A.9) and (A.10) yield

$$\begin{aligned} r_{(\min A_b,b)} &= F(x_{\min A_b}) - G(y_{b-1}) = F(x_{\min A_b}) - \mathbb{P}(\varphi(X) < y_b) \\ &= F(x_{\min A_b}) - F(x_{\min A_b - 1}) - p_{\min A_b}\mathbb{P}(\varphi(x_{\min A_b}) < y_b) \\ &= p_{\min A_b}\mathbb{P}(\varphi(x_{\min A_b}) \geq y_b). \end{aligned}$$

Furthermore, because there exists an $x > x_{\min A_b}$ with $\mathbb{P}(\varphi(x) = y_b) > 0$, the fact that φ is MD-increasing dictates that $\mathbb{P}(\varphi(x_{\min A_b}) > y_b) = 0$. Dividing by $p_{\min A_b}$ then yields

$$\mathbb{P}(\varphi(x_{\min A_b}) = y_b) = \frac{r_{(\min A_b,b)}}{p_{\min A_b}},$$

as asserted. Similarly, (A.8), (A.11) and (A.12) yield

$$\begin{aligned} r_{(\max A_b,b)} &= G(y_b) - F(x_{\max A_b - 1}) = 1 - \mathbb{P}(\varphi(X) > y_b) - F(x_{\max A_b - 1}) \\ &= F(x_{\max A_b}) - F(x_{\max A_b - 1}) - p_{\max A_b}\mathbb{P}(\varphi(x_{\max A_b}) > y_b) \\ &= p_{\max A_b}\mathbb{P}(\varphi(x_{\max A_b}) \leq y_b). \end{aligned} \tag{A.14}$$

Since there exists an $x < x_{\max A_b}$ with $\mathbb{P}(\varphi(x) = y_b) > 0$, it follows that $\mathbb{P}(\varphi(x_{\max A_b}) < y_b) = 0$ since φ is MD-increasing. Hence,

$$\mathbb{P}(\varphi(x_{\max A_b}) = y_b) = \frac{r_{(\max A_b,b)}}{p_{\max A_b}},$$

Now let $a \in \underline{\overline{A_b}} = A_b \setminus \{\min A_b, \max A_b\}$. Since there exist both larger (e.g. $x_{\max A_b}$) and smaller (e.g. $x_{\min A_b}$) supporting points of F than x_a that are mapped by φ onto y_b with positive probability, it follows that $\mathbb{P}(\varphi(x_a) = y_b) = 1$ holds because φ is MD-increasing. It follows

$$r_{(a,b)} = F(x_a) - F(x_{a-1}) = p_a = p_a\mathbb{P}(\varphi(x_a) = y_b) \Longleftrightarrow \mathbb{P}(\varphi(x_a) = y_b) = \frac{r_{(a,b)}}{p_a}. \tag{A.15}$$

It remains to consider $b = \min B$ and $b = \max B$, provided that the minimum and maximum of B exists, respectively. Note that both cases cannot occur simultaneously since that would imply $|B| = 1$ and, therefore, that the distribution of Y is degenerate. We first consider $b = \min B$; in that case A_b still has a maximum, but not necessarily a minimum. Hence, (A.8), (A.11) and (A.12) still hold. Additionally, since $\inf A_b = \inf A$ and since φ is MD-increasing, $A_b = \{\max A_b, \max A_b - 1, \ldots\} = A \cap (-\infty, \max A_b]$. It then follows from (A.11) and (A.12) that, for all $a \in A$, $a \leftrightharpoons b$ is equivalent to $a \leq \max A_b$, so to $a \in A_b$. This again implies $\mathbb{P}(\varphi(x_a) = y_b) = 0 = r_{(a,b)} = \frac{r_{(a,b)}}{p_a}$ for all $a \in A \setminus A_b$.

In the case $|A_b| = 1$ with a_b denoting the one element of A_b, it once again directly follows that a_b is the only element $a \in A$ with $a \leftrightharpoons b$, yielding $r_{(a_b,b)} = q_b$. Considering $y_b = \min(\mathrm{supp}(G))$, we obtain

$$\begin{aligned} r_{(a_b,b)} &= q_b = \mathbb{P}(Y = y_b) = \mathbb{P}(\varphi(X) = y_b) \\ &= 1 - \mathbb{P}(\varphi(X) > y_b) \\ &= F(x_{a_b}) - p_{a_b}\mathbb{P}(\varphi(x_{a_b}) > y_b) \\ &= p_{a_b}(1 - \mathbb{P}(\varphi(x_{a_b}) > y_b)) \\ &= p_{a_b}\mathbb{P}(\varphi(x_{a_b}) = y_b). \end{aligned}$$

The assertion in this case follows by rearranging the equation.

In the case $|A_b| \geq 2$, (A.14) still holds. Because y_b is the smallest element of $\mathrm{supp}(G)$, $\mathbb{P}(\varphi(x_{\max A_b}) \leq y_b)$ is equal to $\mathbb{P}(\varphi(x_{\max A_b}) = y_b)$. Rearranging then yields the assertion for $a = \max A_b$. For $a \in \overline{A_b} = A_b \setminus \{\max A_b\}$, there exists a larger supporting point of F that x_a (namely $x_{\max A_b}$) that is mapped by φ onto y_b with positive probability, yielding $\mathbb{P}(\varphi(x_a) > y_b) = 0$ because φ is MD-increasing. Since $\mathbb{P}(\varphi(x_a) < y_b) = 0$ was already noted, we obtain $\mathbb{P}(\varphi(x_a) = y_b) = 1$. Hence, (A.15) holds and the proof is completed for the case $b = \min B$.

The case $b = \max B$ is analogous with the roles of minima and maxima being switched. ▨

The following result reinforces the RRIDF's in their asserted role as a discrete generalization of the RIDF. It states that the RIDF is the only possible choice for a (then non-random) RRIDF, if and only if the RIDF is actually successful in transforming one random variable into another (via the equivalent characterization given in Propositions A.3 and A.4).

Proposition A.9. *Let $F, G \in \mathcal{D}_0$. Then, $\Phi_F^G = \{R_{FG}\}$, if and only if $G(D_G) \subseteq F(D_F)$.*

Proof. First, note that the marginal distributions of all $\varphi \in \Phi_F^G$ are the same. Therefore, if R_{FG} as a deterministic function is an element of Φ_F^G, it is also the only element of that set since there is then no dependency structure to be varied. So it remains to be shown that $R_{FG} \in \Phi_F^G$, if and only if $G(D_G) \subseteq F(D_F)$. For any $a \in A$, let $b_a \in B$ be the unique index that satisfies $y_{b_a} = R_{FG}(x_a)$. By definition of the quantile function, we obtain $y_{b_a} = \inf\{t \in \mathbb{R} : G(t) \geq F(x_a)\}$. This yields $G(y_{b_a}) \geq F(x_a) > G(y_{b_a - 1})$ because of the minimality of b_a and, hence, $a \leftrightharpoons b_a$. By Definition A.5, $G^{-1} \circ F \in \Phi_F^G$ is equivalent to $r_{(a,b_a)} = p_a$ for all $a \in A$, which, in turn, is equivalent to $R(\leftrightharpoons) \cap (\{a\} \times B)$ only containing one element for each $a \in A$. This is equivalent to $G(D_G) \cap (F(x_{a-1}), F(x_a)) = \emptyset$ for all $a \in A$

and therefore to

$$\emptyset = \bigcup_{a \in A} (G(D_G) \cap (F(x_{a-1}), F(x_a))) = G(D_G) \cap ((0,1) \setminus F(D_F))$$

or $G(D_G) \subseteq F(D_F)$ (since $F(D_F), G(D_G) \subseteq (0,1)$). ▨

An alternative proof of Proposition A.9 can be obtained by combining Propositions A.3 and A.4 with Theorem A.8 as follows; let $F, G \in \mathcal{D}_0$. Due to Propositions A.3 and A.4, $G(D_G) \subseteq F(D_F)$ is equivalent to $R_{FG}(X) =_{st} Y$, which itself is equivalent to $R_{FG} \in \Phi_F^G$ because of Theorem A.8 and because R_{FG} is increasing. The elements of Φ_F^G all have the same marginal distributions and can only differ based on the structure of their joint distributions. However, because R_{FG} is deterministic, there is no joint distribution to be varied, and so $R_{FG} \in \Phi_F^G$ is equivalent to R_{FG} being the only element of Φ_F^G, which concludes the proof.

The concept of RRIDF's is closely related to the so-called distributional transform, which has been considered by Ferguson (1967), Rüschendorf (1981, 2009), Shorack and Wellner (1986) and Shorack (2017), among others. It generalizes the result that $F(X) \sim \mathcal{U}([0,1])$ holds, if F is a continuous cdf and $X \sim F$, to arbitrary cdf's. Note that this result also follows from Proposition A.3 as a special case, if G is the cdf of the uniform distribution on the unit interval. Moreover, it states that F being continuous (or, equivalently, $(0,1) \subseteq \overline{F(D_F)}$) is a necessary and sufficient condition for $F(X) \sim \mathcal{U}([0,1])$. There are two well known versions of the distributional transform. The version used by Ferguson (1967, p. 216) and Rüschendorf (1981, pp. 330–331, 2009, p. 3922) is defined as

$$F^{*F}(X) = F^F(X, V) \quad \text{with} \quad F^F(x, \lambda) = \mathbb{P}(X < x) + \lambda \cdot \mathbb{P}(X = x),$$

where $x \in \mathbb{R}$, $\lambda \in [0,1]$ as well as $X \sim F$ and $V \sim \mathcal{U}([0,1])$ independent. This is equivalent to the following pointwise definition on the underlying probability space Ω

$$F^{*F}(X)(\omega) = \lim_{t \nearrow X(\omega)} F(t) + V(\omega) \cdot \left(F(X(\omega)) - \lim_{t \nearrow X(\omega)} F(t) \right), \quad \omega \in \Omega.$$

In the case $F \in \mathcal{D}_0$, it can be reformulated to

$$F^{*F}(X) = F(X) - V \cdot \sum_{a \in A} p_a \mathbb{1}_{\{x_a\}}(X).$$

To this end, note that

$$\begin{aligned} F^{*F}(X)(\omega) &= \lim_{t \nearrow X(\omega)} F(t) + V(\omega) \cdot \left(F(X(\omega)) - \lim_{t \nearrow X(\omega)} F(t) \right) \\ &= F(X(\omega)) - \left(F(X(\omega)) - \lim_{t \nearrow X(\omega)} F(t) \right) + V(\omega) \cdot \left(F(X(\omega)) - \lim_{t \nearrow X(\omega)} F(t) \right) \end{aligned}$$

$$= F(X(\omega)) - (1 - V(\omega)) \cdot \left(F(X(\omega)) - \lim_{t \nearrow X(\omega)} F(t) \right), \quad \omega \in \Omega,$$

where $(1 - V) \sim \mathcal{U}([0, 1])$. Shorack and Wellner (1986, p. 106) and Shorack (2017, p. 113) defined the distributional transform in a slightly different way, namely

$$F^{*S}(X) = F(X) - \sum_{a \in A} U_a p_a \mathbb{1}_{\{x_a\}}(X), \tag{A.16}$$

where $(U_a)_{a \in A}$ is a sequence of independent random variables that are uniformly distributed on the unit interval. Moreover, the sequence $(U_a)_{a \in A}$ is independent of X. The representation (A.16) is also applicable in the case $F \notin \mathcal{D}_0$. Then, A is a countable set that indexes all jump discontinuities with the sequence $(x_a)_{a \in A}$ denoting their locations and $(p_a)_{a \in A}$ denoting their jump heights.

It has been shown that, for any $F \in \mathcal{P}$, both $F^{*F}(X)$ and $F^{*S}(X)$ are uniformly distributed on the unit interval (see Ferguson (1967, p. 216, Lemma 1) and Shorack (2017, p. 113, Proposition 3.2)). However, while Ferguson uses the same uniformly distributed random variable V to spread out the probability mass of all jump discontinuities, Shorack assigns a separate independent uniformly distributed random variable $U_a, a \in A$, to each jump discontinuity. One could define even more versions of the distributional transform by changing the joint distribution of these uniformly distributed random variables (other than 'identical' and 'independent', which are already taken). The way, in which the construction of these distributional transforms is rigid in the marginal distributions but variable in the dependence structure or copula, is reminiscent of the sets of RRIDF's from Definition A.5. In fact, both concepts are closely related. In order to see this, define both F^{*F} and F^{*S} as random elements on $\mathrm{M}(\mathrm{supp}(F), (0, 1))$.

Proposition A.10. *Let $F, G \in \mathcal{D}_0$. Then, $\varphi \in \Phi_F^G$ is equivalent to the existence of a sequence $(U_a)_{a \in A}$ with $U_a \sim \mathcal{U}([0, 1]), a \in A$, satisfying $(G^{-1} \circ F^{*U_a}) =_{st} \varphi$, where $F^{*U_a}(t) = F(t) - \sum_{a \in A} U_a p_a \mathbb{1}_{\{x_a\}}(t)$ for $t \in \mathrm{supp}(F)$.*

Proof. '$\Leftarrow$': Obviously, $(G^{-1} \circ F^{*U_a})$ is a random object on $\mathrm{M}(\mathrm{supp}(F), \mathrm{supp}(G))$. For $a \in A$, we obtain $F^{*U_a}(x_a) = F(x_a) - p_a U_a \sim \mathcal{U}((F(x_{a-1}), F(x_a)))$. For any $b \in B$, the set of values that is mapped onto y_b by G^{-1} is given by $(G(y_{b-1}), G(y_b)]$. It follows

$$\begin{aligned}
\mathbb{P}((G^{-1} \circ F^{*U_a})(x_a) = y_b) &= \mathbb{P}(F^{*U_a}(x_a) \in (G^{-1})^{-1}(\{y_b\})) \\
&= \mathbb{P}(F^{*U_a}(x_a) \in (G(y_{b-1}), G(y_b)]) \\
&= \int_{G(y_{b-1})}^{G(y_b)} \frac{1}{F(x_a) - F(x_{a-1})} \mathbb{1}_{(F(x_{a-1}), F(x_a))}(t) \mathrm{d}t \\
&= \frac{\int_{\mathbb{R}} \mathbb{1}_{(F(x_{a-1}), F(x_a)) \cap (G(y_{b-1}), G(y_b))}(t) \mathrm{d}t}{F(x_a) - F(x_{a-1})}
\end{aligned}$$

$$= \frac{r_{(a,b)}}{p_a}$$

for all $a \in A, b \in B$.

'$\Rightarrow$': Let $\varphi \in \Phi_F^G$. Furthermore, let $(U_a)_{a\in A}$ be a sequence of random variables satisfying $U_a \sim \mathcal{U}([0,1])$ for all $a \in A$ as well as

$$\mathbb{P}\left(\bigcap_{a\in A}\left\{U_a \in \left(\frac{F(x_a)-G(y_b)}{F(x_a)-F(x_{a-1})}, \frac{F(x_a)-G(y_{b-1})}{F(x_a)-F(x_{a-1})}\right)\cap(0,1)\right\}\right)$$
$$= \mathbb{P}\left(\bigcap_{a\in A}\{\varphi(x_a) = y_b\}\right) \tag{A.17}$$

for all $b \in B$. Note that the corresponding marginal distributions are the same: if $a \leftrightharpoons b$, we have

$$\mathbb{P}\left(U_a \in \left(\frac{F(x_a)-G(y_b)}{F(x_a)-F(x_{a-1})}, \frac{F(x_a)-G(y_{b-1})}{F(x_a)-F(x_{a-1})}\right)\cap(0,1)\right) = \frac{r_{(a,b)}}{p_a} = \mathbb{P}(\varphi(x_a) = y_b). \tag{A.18}$$

The left hand probability in (A.18) is zero whenever

$$\left(\frac{F(x_a)-G(y_b)}{F(x_a)-F(x_{a-1})}, \frac{F(x_a)-G(y_{b-1})}{F(x_a)-F(x_{a-1})}\right)\cap(0,1) = \emptyset. \tag{A.19}$$

(A.19) is equivalent to $a \not\leftrightharpoons b$ for all $a \in A, b \in B$, which is, in turn, equivalent to $\mathbb{P}(\varphi(x_a) = y_b) = 0$. The sequence $(U_a)_{a\in A}$ exists because of

$$(0,1)\setminus N \subseteq \bigcup_{b\in B}\left(\frac{F(x_a)-G(y_b)}{F(x_a)-F(x_{a-1})}, \frac{F(x_a)-G(y_{b-1})}{F(x_a)-F(x_{a-1})}\right) \quad \text{and}$$
$$\left(\frac{F(x_a)-G(y_b)}{F(x_a)-F(x_{a-1})}, \frac{F(x_a)-G(y_{b-1})}{F(x_a)-F(x_{a-1})}\right)\cap\left(\frac{F(x_a)-G(y_\beta)}{F(x_a)-F(x_{a-1})}, \frac{F(x_a)-G(y_{\beta-1})}{F(x_a)-F(x_{a-1})}\right) = \emptyset$$

for all $a \in A$, all $b, \beta \in B$ with $b \neq \beta$ and for some null set N with respect to the Lebesgue measure. This means that the discretized values of $U_a, a \in A$, which are given by open sub-intervals of $(0,1)$, are mutually exclusive and make up the entire space of values that the corresponding random variable can take (except of a null set). Therefore, the joint discrete distribution of the sequence $(\varphi(x_a))_{a\in A}$ can be transferred to $(U_a)_{a\in A}$ with the behaviour of the latter sequence within the given intervals not being specified. Now,

$$\mathbb{P}\left(\bigcap_{a\in A}\left\{\left(G^{-1}\circ F^{*U_a}\right)(x_a) = y_b\right\}\right)$$
$$= \mathbb{P}\left(\bigcap_{a\in A}\left\{F^{*U_a}(x_a) \in (G(y_{b-1}), G(y_b)]\right\}\right)$$

$$
\begin{aligned}
&= \mathbb{P}\left(\bigcap_{a\in A}\{F(x_a) - p_a \cdot U_a \in (G(y_{b-1}), G(y_b)]\}\right)\\
&= \mathbb{P}\left(\bigcap_{a\in A}\left\{U_a \in \left[\frac{F(x_a) - G(y_b)}{F(x_a) - F(x_{a-1})}, \frac{F(x_a) - G(y_{b-1})}{F(x_a) - F(x_{a-1})}\right)\right\}\right)\\
&= \mathbb{P}\left(\bigcup_{a\in A}\{\varphi(x_a) = y_b\}\right)
\end{aligned}
$$

holds for all $a \in A, b \in B$, where the last identity is due to (A.17). This proves $(G^{-1} \circ F^{*U_a}) =_{st} \varphi$.

Invoking the implication from right to left in the above proposition for the two specific choices of $(U_a)_{a\in A}$ that were already mentioned yields the following corollary.

Corollary A.11. *If $F, G \in \mathcal{D}_0$, then $(G^{-1} \circ F^{*F}), (G^{-1} \circ F^{*S}) \in \Phi_F^G$.*

Let $F, G \in \mathcal{D}_0$ and let $\mathbb{U} = \{(U_a)_{a\in A} : U_a \sim \mathcal{U}([0,1])\ \forall a \in A\}$ denote the set of all $|A|$-dimensional copulas. Then, an equivalence relation $\bowtie_F^G$ can be defined on $\mathbb{U}$ via $(U_a)_{a\in A} \bowtie_F^G (V_a)_{a\in A}$, if and only if $G^{-1} \circ F^{*U_a} =_{st} G^{-1} \circ F^{*V_a}$. Now Proposition A.10 states that the quotient set of $\mathbb{U}$ by $\bowtie_F^G$ is isomorphic to the set Φ_F^G of RRIDF's from F to G, i.e. $\mathbb{U}/\bowtie_F^G \cong \Phi_F^G$.

We now apply the concept of RRIDF's to orders of convex characteristics, starting with the usual stochastic order $\leq_{st}$. Propositions 6.1 and 6.2 state that $\leq_{st}$ is equivalently characterized by the RIDF R_{GF}, but generally not by R_{FG}. Similarly to the general case, the stochastic order can be characterized via the sets of RRIDF's if we restrict ourselves to (purposive) discrete distributions. In this case, both characterizations are indeed equivalent to $F \leq_{st} G$, as demonstrated in Theorem A.13.

Lemma A.12. *Let $F, G \in \mathcal{D}_0$. Then, $F \leq_{st} G$, if and only if $x_a \leq y_b$ for all $(a, b) \in R(\leftrightharpoons)$.*

Proof. The assertion follows from Definition 7.8, similarly to the proof of Proposition 7.9:

$$
\begin{aligned}
F \leq_{st} G \iff & F^{-1}(r) \leq G^{-1}(r) && \forall r \in (0,1)\\
\iff & x_a \leq y_b && \forall (a,b) \in A \times B \text{ such that}\\
& && \quad F^{-1}(r) = x_a, G^{-1}(r) = y_b \text{ for some } r \in (0,1)\\
\iff & x_a \leq y_b && \forall (a,b) \in R(\leftrightharpoons).
\end{aligned}
$$

Theorem A.13. *Let $F, G \in \mathcal{D}_0$. Then, the following three statements are equivalent:*

(i) $F \leq_{st} G$,

(ii) $\mathbb{P}(\varphi(t) \leq t) = 1$ *for all* $t \in \operatorname{supp}(G)$ *and for all* $\varphi \in \Phi_G^F$,

(iii) $\mathbb{P}(\varphi(t) \geq t) = 1$ *for all* $t \in \operatorname{supp}(F)$ *and for all* $\varphi \in \Phi_F^G$.

Proof. We only prove the equivalence of (i) and (ii) since the proof of the equivalence of (i) and (iii) is entirely analogous.

’(i)⇒(ii)’: Let $\varphi \in \Phi_G^F$ and $t \in \operatorname{supp}(G)$. Let $b \in B$ be the unique element satisfying $t = y_b$. By assumption, $(a, b) \in R(\fallingdotseq)$ implies $x_a \leq y_b$. Therefore, the fact that the random variable $\varphi(y_b)$ almost surely only takes values in $\{x_a \in \operatorname{supp}(F) : a \fallingdotseq b\}$, i.e. $\mathbb{P}(\varphi(y_b) \in \{x_a \in \operatorname{supp}(F) : a \fallingdotseq b\}) = 1$, implies that $\mathbb{P}(\varphi(t) \leq t) = \mathbb{P}(\varphi(y_b) \leq y_b) = 1$. Since t was chosen arbitrarily in $\operatorname{supp}(G)$, this proves the asserted implication.

’(ii)⇒(i)’: Let $b \in B$. We obtain

$$\sum_{a \in A : a \fallingdotseq b} \mathbb{P}(\varphi(y_b) = x_a) = 1 = \mathbb{P}(\varphi(y_b) \leq y_b) = \sum_{a \in A : a \fallingdotseq b} \mathbb{P}(\varphi(y_b) = x_a) \cdot \mathbb{1}\{x_a \leq y_b\}.$$

Since each of the probabilities in the sum on the left hand side is positive, this identity implies that all of the indicator functions in the sum on the right hand side are equal to one. It follows that $x_a \leq y_b$ for all $a \in A$ such that $a \fallingdotseq b$.

Since the class of functions, for which the stochastic order cannot be equivalently described using RIDF’s (see Proposition 6.1), also contains purposive discrete distributions, the characterization via RRIDF’s provides a viable alternative. However, that critical class of distributions from Proposition 6.1 also contains other discrete distributions as well as mixtures of continuous and discrete distributions. For these kind of more complex non-continuous distributions, one cannot define a corresponding set of RRIDF’s by Definition A.5 because the concepts of an indexing set, an identifying sequence and of the relation $\fallingdotseq$ are only defined for purposive discrete distributions.

For sufficiently regular continuous distributions, the order of the second convex characteristic, i.e. the dispersive order $\leq_{disp}$, can also be characterized using RIDF’s. It is demonstrated in Example 6.5 that these characterizations are not sensible and also not equivalent to $\leq_{disp}$ for discrete distributions.

A result for a discrete dispersive order that is similar to what is shown for the stochastic order in Theorem A.13 could not be proved, and we conjecture that it does not hold at all. Instead, only a number of implications like the one shown in Theorem A.15 seem to hold. Overall, while RRIDF’s succeed in transforming purposive discrete random variables into one another, they are seemingly not fit to describe orders of convex characteristics other than the stochastic order.

Lemma A.14. *Let $F, G \in \mathcal{D}_0$. For $a, \alpha \in A$ with $a \neq \alpha$, the sets $\{b \in B : a \leftrightharpoons b\}$ and $\{b \in B : \alpha \leftrightharpoons b\}$ have at most one element in common, i.e.*

$$|\{b \in B : a \leftrightharpoons b \text{ and } \alpha \leftrightharpoons b\}| \leq 1.$$

Proof. Assume the existence of $b, \beta \in \{b \in B : a \leftrightharpoons b \text{ and } \alpha \leftrightharpoons b\}$ with $b \neq \beta$. Furthermore, assume without restriction that $\alpha > a$ and $\beta > b$. It follows

$$G(y_b) \leq G(y_{\beta-1}) < F(x_a) \leq F(x_{\alpha-1}),$$

which contradicts $\alpha \leftrightharpoons b$. ∎

Theorem A.15. *For $F, G \in \mathcal{D}_0$, the following implication holds:*

$$F \leq_{disp}^{\wedge-disc} G \Longrightarrow \exists \varphi \in \Phi_G^F : \mathbb{P}(\varphi(t) - \varphi(s) \leq t - s) = 1 \quad \forall s, t \in \operatorname{supp}(G), s < t.$$

Proof. Let $F \mathrel{\hat{=}} (\tilde{A}, (x_j, p_j)_{j \in \tilde{A}})$ and $G \mathrel{\hat{=}} (B, (y_j, q_j)_{j \in B})$. If $\tilde{A} = -\mathbb{N}$, let $A = \tilde{A} + 3$ so that $\max A = 2$. Otherwise, let $A = \tilde{A}$.

Define the discrete random variable Z_1 by $\mathbb{P}(Z_1 = y_b) = \frac{r_{(1,b)}}{p_1}$ for all $b \in B$. This is possible because of $1 \in A$ and uniquely defines the distribution of Z_1 because of $\sum_{b \in B} \mathbb{P}(Z_1 = y_b) = \sum_{b \in B : 1 \leftrightharpoons b} \mathbb{P}(Z_1 = y_b) = 1$. Note that Z_1 only takes values on $\{y_b \in \operatorname{supp}(G) : 1 \leftrightharpoons b\}$. Now we define another discrete random variable Z_2. If $\{b \in B : 1 \leftrightharpoons b \text{ and } 2 \leftrightharpoons b\} = \emptyset$, we define Z_2 independently of Z_1 by $\mathbb{P}(Z_2 = y_b) = \frac{r_{(2,b)}}{p_2}$ for all $b \in B$. If $|\{b \in B : 1 \leftrightharpoons b \text{ and } 2 \leftrightharpoons b\}| = 1$ (any higher cardinalities are impossible due to Lemma A.14), we denote the one element in that set by $b_{1,2}$. It follows that $p_1 \geq q_{b_{1,2}}$, $p_2 \geq q_{b_{1,2}}$ and $G(y_{b_{1,2}-1}) < F(x_1) < G(y_{b_{1,2}})$. By combining these statements, we obtain

$$\begin{aligned} F(x_0) &= F(x_1) - p_1 < G(y_{b_{1,2}}) - q_{b_{1,2}} = G(y_{b_{1,2}-1}) \quad \text{and} \\ F(x_2) &= F(x_1) + p_2 > G(y_{b_{1,2}-1}) + q_{b_{1,2}} = G(y_{b_{1,2}}). \end{aligned}$$

Hence, $1 \leftrightharpoons b_{1,2} - 1$ and $2 \leftrightharpoons b_{1,2} + 1$, which means that the sets $\{b \in B : 1 \leftrightharpoons b\}$ and $\{b \in B : 2 \leftrightharpoons b\}$ both contain at least one element apart from $b_{1,2}$. We define the distribution of Z_2 by

$$\begin{aligned} &\mathbb{P}(Z_1 = y_{b_1}, Z_2 = y_{b_2}) \\ &= \begin{cases} 0 & \text{, if } b_1 = b_{1,2} = b_2, \\ \frac{r_{(1,b_1)}}{p_1 - r_{(1,b_{1,2})}} \cdot \frac{r_{(2,b_{1,2})}}{p_2} & \text{, if } b_1 \neq b_{1,2} = b_2, \\ \frac{r_{(1,b_{1,2})}}{p_1} \cdot \frac{r_{(2,b_2)}}{p_2 - r_{(2,b_{1,2})}} & \text{, if } b_1 = b_{1,2} \neq b_2, \\ \frac{r_{(1,b_1)}}{p_1 - r_{(1,b_{1,2})}} \cdot \frac{r_{(2,b_2)}}{p_2 - r_{(2,b_{1,2})}} \cdot \left(1 - \frac{r_{(1,b_{1,2})}}{p_1} - \frac{r_{(2,b_{1,2})}}{p_2}\right) & \text{, if } b_1 \neq b_{1,2} \neq b_2, \end{cases} \end{aligned}$$

for all $b_1, b_2 \in B$ such that $1 \leftrightharpoons b_1$ and $2 \leftrightharpoons b_2$. Note that all of these probabilities are well-defined, particularly in the last case because of

$$\frac{r_{(1,b_{1,2})}}{p_1} + \frac{r_{(2,b_{1,2})}}{p_2} = \frac{q_{b_{1,2}} \cdot \frac{p_2}{p_1} - r_{(2,b_{1,2})} \cdot \frac{p_2}{p_1}}{p_2} + \frac{r_{(2,b_{1,2})}}{p_2} \leq 1$$

if $p_2 \leq p_1$ and

$$\frac{r_{(1,b_{1,2})}}{p_1} + \frac{r_{(2,b_{1,2})}}{p_2} = \frac{r_{(1,b_{1,2})}}{p_1} + \frac{q_{b_{1,2}} \cdot \frac{p_1}{p_2} - r_{(1,b_{1,2})} \cdot \frac{p_1}{p_2}}{p_1} \leq 1$$

if $p_1 \leq p_2$. Furthermore, we obtain

$$\mathbb{P}(Z_1 = y_{b_{1,2}}) = \frac{r_{(1,b_{1,2})}}{p_1} \frac{1}{p_2 - r_{(2,b_{1,2})}} \sum_{\substack{b_2 \in B\setminus\{b_{1,2}\}: \\ 2 \leftrightharpoons b_2}} r_{(2,b_2)} = \frac{r_{(1,b_{1,2})}}{p_1} \quad \text{and}$$

$$\begin{aligned} \mathbb{P}(Z_1 = y_{b_1}) = {} & \frac{r_{(1,b_1)}}{p_1 - r_{(1,b_{1,2})}} \cdot \\ & \left[\frac{r_{(2,b_{1,2})}}{p_2} + \left(1 - \frac{r_{(1,b_{1,2})}}{p_1} - \frac{r_{(2,b_{1,2})}}{p_2}\right) \frac{1}{p_2 - r_{(2,b_{1,2})}} \sum_{\substack{b_2 \in B\setminus\{b_{1,2}\}: \\ 2 \leftrightharpoons b_2}} r_{(2,b_2)} \right] \\ = {} & \frac{r_{(1,b_1)}}{p_1 - r_{(1,b_{1,2})}} \cdot \left[1 - \frac{r_{(1,b_{1,2})}}{p_1}\right] = \frac{r_{(1,b_1)}}{p_1} \end{aligned}$$

for all $b_1 \in B \setminus \{b_{1,2}\}$ such that $1 \leftrightharpoons b_1$. For all $b_1 \in B$ with $1 \not\leftrightharpoons b_1$, we have $\mathbb{P}(Z_1 = y_{b_1}) = 0 = \frac{r_{(1,b_1)}}{p_1}$. Hence, the definition of Z_2 is compatible with that of Z_1. Furthermore, for reasons of symmetry, the marginal distribution is given by $\mathbb{P}(Z_2 = y_b) = \frac{r_{(2,b)}}{p_2}$ for all $b \in B$. If $3 \in A$, we can now define a random variable Z_3 analogously to the definition of Z_2 with the pair $(1,2)$ being substituted for the pair $(2,3)$. Specifically, the definition of Z_3 is only dependent upon Z_2 and independent of Z_1. The previous considerations for Z_1 and Z_2 can be replicated analogously to obtain that the marginal distribution of Z_3 is given by $\mathbb{P}(Z_3 = y_b) = \frac{r_{(3,b)}}{p_3}$ for all $b \in B$. We continue to recursively define random variables Z_a for all $a \in A, a \geq 1$. Furthermore, if $0 \in A$, define the random variable Z_0 also analogously to Z_2, now substituting the pair $(1,2)$ for the pair $(1,0)$. By, again, continuing to do this recursively, we obtain the random variables Z_a for all $a \in A, a \leq 0$. Overall, we have now defined a sequence $(Z_a)_{a \in A}$ of random variables, which satisfies $\mathbb{P}(Z_a = y_b) = \frac{r_{(a,b)}}{p_a}$ for all $a \in A, b \in B$. Therefore, by defining $\varphi \in \mathrm{M}(\mathrm{supp}(F), \mathrm{supp}(G))$ through $(\varphi(x_a))_{a \in A} = (Z_a)_{a \in A}$, we obtain $\varphi \in \Phi_F^G$.

Let $a_\ell, a_u \in A$ with $a_\ell < a_u$. It remains to be shown that $\mathbb{P}(\varphi(x_{a_u}) - \varphi(x_{a_\ell}) \geq x_{a_u} - x_{a_\ell}) = 1$. Since $\mathbb{P}(\varphi(x_{a_\ell}) = y_{\beta_\ell}) = 0 = \mathbb{P}(\varphi(x_{a_u}) = y_{\beta_u})$ holds for all $\beta_\ell, \beta_u \in B$ with $a_\ell \not\leftrightharpoons \beta_\ell$ or $a_u \not\leftrightharpoons \beta_u$, there exist $b_\ell, b_u \in B$ with $a_\ell \leftrightharpoons b_\ell$ and $a_u \leftrightharpoons b_u$ such that $\mathbb{P}(\varphi(x_{a_\ell}) = y_{b_\ell}, \mathbb{P}(\varphi(x_{a_u}) = y_{b_u}) > 0$. Assume now $x_{a_u} - x_{a_\ell} > y_{b_u} - y_{b_\ell}$. Accordning to (7.20) in Part 1 of the proof of Theorem 7.30, this implies $a_u - a_\ell = b_u - b_\ell + 1$. From Part 2 of the proof of Theorem 7.30, it follows from $a_u - a_\ell = b_u - b_\ell + 1$ that

$$\begin{aligned} &a_\ell \leftrightharpoons b_\ell, \quad a_\ell + 1 \leftrightharpoons b_\ell + 1, \ldots, \; a_u - 2 \leftrightharpoons b_u - 1, \; a_u - 1 \leftrightharpoons b_u, \\ &a_\ell \leftrightharpoons b_\ell - 1, \; a_\ell + 1 \leftrightharpoons b_\ell, \quad \ldots, \; a_u - 1 \leftrightharpoons b_u - 1, \quad a_u \leftrightharpoons b_u, \end{aligned}$$

see (7.21) and (7.25). The situation is illustrated in Figure 7.5. Since $\mathbb{P}(\varphi(x_{a_\ell}) = y_{b_\ell}, \varphi(x_{a_\ell+1}) = y_{b_\ell}) = 0$ holds by construction of φ, $\varphi(x_{a_\ell+1}) = y_{b_\ell+1}$ follows almost surely. From that, because of $\mathbb{P}(\varphi(x_{a_\ell+1}) = y_{b_\ell+1}, \varphi(x_{a_\ell+2}) = y_{b_\ell+1}) = 0$, $\varphi(x_{a_\ell+2}) = y_{b_\ell+2}$ follows almost surely, and so on. Inductively, we obtain $\mathbb{P}(\varphi(x_{a_u-1}) = b_u) = 1$, which contradicts $\mathbb{P}(\varphi(x_{a_u}) = b_u) > 0$ because of $\mathbb{P}(\varphi(x_{a_u-1}) = y_{b_u}, \varphi(x_{a_u}) = y_{b_u}) = 0$. Hence, $y_{b_u} - y_{b_\ell} \geq x_{a_u} - x_{a_\ell}$. Since $y_{b_\ell}, y_{b_u} \in \operatorname{supp}(G)$ were arbitrarily chosen within the sets of (almost surely) possible values of $\varphi(x_{a_\ell})$ and $\varphi(x_{a_u})$, respectively, it follows that $\mathbb{P}(\varphi(x_{a_u}) - \varphi(x_{a_\ell}) \geq x_{a_u} - x_{a_\ell}) = 1$. ▨

APPENDIX B

Notation: Orders and Measures

Orders

In Table B.1, all sufficiently important stochastic orders introduced throughout this thesis with specific notation are recollected along with the pages on which they are introduced. The orders are grouped by the corresponding characteristic. Furthermore, there are separate groups for families of orders for different characteristics and orders that are defined specifically for discrete distributions.

Note that the notation $\leq_s$ is used as both an order of skewness and an order of kurtosis. The notation is used in the literature for both orders and they are neither generally related nor connected throughout the thesis. Furthermore, the name 'weak dispersive order' is used for $\leq_{w-disp}$ as well as another dispersion order that is introduced without a specific notation in the remarks preceding Theorem 7.30. The name is again used for both orders in the literature although they describe different concepts.

Measures

For all measures throughout the thesis, for which a specific notation is introduced, Table B.2 states the page on which it is introduced or on which the corresponding result is presented. On one hand, the measures can be grouped in terms of the characteristic they measure (i.e., central location, dispersion, skewness or kurtosis): this is denoted by the greek letters ν, τ, γ and κ. On the other hand, they can be grouped in terms of how they measure that characteristic (i.e., moment based, quantile based, etc.): this is denoted by the indices of the greek letters, e.g. 'M' for moment based measures.

The sign '—' in a table cell indicates that no corresponding measure is defined within this thesis. The alternate notations for dispersion measures in a discrete context are given in a separate column.

Families	
$\leq_k$ for $k \in \mathbb{N}_0$	p. 23
$\leq_{k-cx}$ for $k \in \mathbb{N}$	p. 29

Location	
$\leq_{st}$	p. 23 / p. 26 (a.k.a. $\leq_0$, $\leq_{1-cx}$)
$\leq_e$	p. 57

Dispersion	
$\leq_{disp}$	p. 23 / p. 26 (a.k.a. $\leq_1$)
$\leq_{cx}$	p. 29 / p. 30 (a.k.a. $\leq_{2-cx}$)
$\leq_{dil}$	p. 30 (a.k.a. $\leq_{2-dil}$)
$\leq_{w-disp}$	p. 58
$\leq_{e-disp}$	p. 58
$\leq_{we-disp}$	p. 59

Skewness	
$\leq_c$	p. 23 / p. 26 (a.k.a. $\leq_2$)
$\leq_{3-cx}$	p. 29 / p. 30
$\leq_{3-dil}$	p. 31
$\leq_\mu^{MAD}$	p. 66
$\leq_s$	p. 66

Kurtosis	
$\leq_3$	p. 23 / p. 77
$\leq_s$	p. 86
$\leq_a$	p. 86
$\leq_S$	p. 75
$\leq_{gs}$	p. 87 (a.k.a. $\leq_{gs}^0$)
$\leq_{gs}^{t_0}$ for $t_0 \in \mathbb{R}$	p. 91
$<_{gss}$	p. 93

Discrete dispersion	
$\leq_{disp}^{disc}$	p. 153
$\leq_{D-pm}^{disc}$	p. 163
$\leq_{D-supp}^{\wedge-disc}$	p. 163
$\leq_{D-supp}^{\vee-disc}$	p. 163
$\leq_{disp}^{\wedge-disc}$	p. 163
$\leq_{disp}^{\vee-disc}$	p. 163

Discrete skewness	
$\leq_{skew}^{disc}$	p. 230

Table B.1.: Page references for the definitions of stochastic orders throughout the thesis.

Based on	Index	Central location	Dispersion	Skewness	Kurtosis	Discrete dispersion
Moments	M	p. 42	p. 42	p. 43	p. 111	p. 202 as SD
L-Moments	LM	p. 44 $(=\nu_M)$	p. 44	p. 44	p. 112	p. 203 as GMD
Quantiles	Q	p. 45	p. 45	p. 45	p. 114	p. 203 as IQR$(\alpha, 1-\alpha)$
Quantiles (integrated)	IQ	—	p. 48	p. 48	—	p. 203 as MDMAD
Quantiles (alternative)	QA	—	—	—	p. 114	—
Quantiles (functional)	QF	—	—	—	p. 117	—
Densities	D	p. 50 $(=\nu_Q)$	p. 50	p. 50	p. 119	—
Densities (integrated)	ID	p. 53	p. 53 (i.a. $=\tau_Q$)	p. 53	p. 122	—
Densities (alternative)	DA	—	—	—	p. 124	—
Densities (functional)	DF	—	—	—	p. 123	—
Mode	$Mode$	p. 54 as M.	—	p. 56	—	—
Expectiles	E	p. 58 $(=\nu_M)$	p. 61	p. 64	p. 124	p. 203 as IER$(\alpha, 1-\alpha)$
Expectiles (alternative)	EA	—	—	p. 68	—	—
Expectiles (lim. value)	EL	—	p. 61	p. 65	p. 126	p. 203 as MAD
Expectiles (lim. val. med.)	EM	—	—	—	p. 127	—

Table B.2.: Page references for the definitions of / results for measures of different characteristics throughout the thesis.

Bibliography

Abadir, K. M. (2005). “The Mean-Median-Mode Inequality: Counterexamples”. *Econometric Theory* 21, pp. 477–482.

Arab, I., Lando, T., and Oliveira, P. E. (2022). “Comparison of L_p-Quantiles and Related Skewness Measures”. *Statistics and Probability Letters* 183, 109339.

Arachchige, C. N. P. G. and Prendergast, L. A. (2019). *Mean Skewness Measures.* arXiv: `1912.06996v2`. URL: `https://arxiv.org/abs/1912.06996v2`.

Arnold, B. C. and Groeneveld, R. A. (1993). “Skewness and Kurtosis Orderings: An Introduction”. *Lecture Notes – Monograph Series* 22, pp. 17–24.

Arnold, B. C. and Groeneveld, R. A. (1995). “Measuring Skewness With Respect to the Mode”. *The American Statistician* 49, pp. 34–38.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). “Coherent Measures of Risk”. *Mathematical Finance* 9, pp. 203–228.

Azzalini, A. (1985). “A Class of Distributions Which Includes the Normal Ones”. *Scandinavian Journal of Statistics* 12, pp. 171–178.

Azzalini, A. and Capitanio, A. (2003). “Distributions Generated by Perturbation of Symmetry With Emphasis on a Multivariate Skew t-Distribution”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65, pp. 367–389.

Bakker, A. and Gravemeijer, K. (2006). “An Historical Phenomenology of Mean and Median”. *Educational Studies in Mathematics* 62, pp. 149–168.

Balanda, K. P. and MacGillivray, H. L. (1988). “Kurtosis—A Critical Review”. *The American Statistician* 42, pp. 111–119.

Balanda, K. P. and MacGillivray, H. L. (1990). “Kurtosis and Spread”. *The Canadian Journal of Statistics* 18, pp. 17–30.

Bellini, F. (2012). “Isotonicity Results for Generalized Quantiles”. *Statistics & Probability Letters* 82, pp. 2017–2024.

Bellini, F., Klar, B., and Müller, A. (2018a). “Expectiles, Omega Ratios and Stochastic Ordering”. *Methodology and Computing in Applied Probability* 20, pp. 855–873.

Bellini, F., Klar, B., Müller, A., and Gianin, E. R. (2014). "Generalized Quantiles as Risk Measures". *Insurance: Mathematics and Economics* 54, pp. 41–48.

Bellini, F., Mercuri, L., and Rroji, E. (2018b). "Implicit Expectiles and Measures of Implied Volatility". *Quantitative Finance* 18, pp. 1851–1864.

Belzunce, F., Martínez-Riquelme, C., and Mulero, J. (2015). *An Introduction to Stochastic Orders*. Academic Press.

Bickel, P. J. and Lehmann, E. L. (1975a). "Descriptive Statistics for Nonparametric Models. I. Introduction". *The Annals of Statistics* 3, pp. 1038–1044.

Bickel, P. J. and Lehmann, E. L. (1975b). "Descriptive Statistics for Nonparametric Models. II. Location". *The Annals of Statistics* 3, pp. 1045–1069.

Bickel, P. J. and Lehmann, E. L. (1976). "Descriptive Statistics for Nonparametric Models. III. Dispersion". *The Annals of Statistics* 4, pp. 1139–1158.

Bickel, P. J. and Lehmann, E. L. (2012). "Descriptive Statistics for Nonparametric Models. IV. Spread". *Selected Works of E. L. Lehmann*. Selected Works in Probability and Statistics. Springer, pp. 519–526.

Birnbaum, Z. W. (1948). "On Random Variables with Comparable Peakedness". *The Annals of Mathematical Statistics* 19, pp. 76–81.

Blest, D. C. (2003). "A New Measure of Kurtosis Adjusted for Skewness". *Australian & New Zealand Journal of Statistics* 45, pp. 175–179.

Bogachev, V. I. (2007). *Measure Theory*. 1st ed. Vol. 2. Springer.

Bowley, A. L. (1901). *Elements of Statistics*. P. S. King & Son.

Chissom, B. S. (1970). "Interpretation of the Kurtosis Statistic". *The American Statistician* 24, pp. 19–22.

Corrado, C. J. and Su, T. (1996). "Skewness and Kurtosis in S&P 500 Index Returns Implied by Option Prices". *The Journal of Financial Research* 19, pp. 175–192.

Crack, T. F. (2019). "Kurtosis as Peakedness: The Phoenix from the Ashes". *Foundations for Scientific Investment*. 8th ed. Timothy F. Crack.

Cristelli, M., Zaccaria, A., and Pietronero, L. (2012). "Universal Relation Between Skewness and Kurtosis in Complex Dynamics". *Physical Review E* 85, 066108.

Critchley, F. and Jones, M. C. (2008). "Asymmetry and Gradient Asymmetry Functions: Density-Based Skewness and Kurtosis". *Scandinavian Journal of Statistics* 35, pp. 415–437.

D'Agostino, R. B., Belanger, A., and D'Agostino, Jr., R. B. (1990). "A Suggestion for Using Powerful and Informative Tests of Normality". *The American Statistician* 44, pp. 316–321.

Darlington, R. B. (1970). "Is Kurtosis Really "Peakedness?"" *The American Statistician* 24, pp. 19–22.

David, F. N. and Johnson, N. L. (1956). “Some Tests of Significance with Ordered Variables”. *Journal of the Royal Statistical Society. Series B (Methodological)* 18, pp. 1–31.

DeCarlo, L. T. (1997). “On the Meaning and Use of Kurtosis”. *Psychological Methods* 2, pp. 292–307.

Doksum, K. A. (1975). “Measures of Location and Asymmetry”. *Scandinavian Journal of Statistics* 2, pp. 11–22.

Eberl, A. and Klar, B. (2019). “On the Skewness Order of van Zwet and Oja”. *Mathematical Methods of Statistics* 28, pp. 262–278.

Eberl, A. and Klar, B. (2020). “Asymptotic Distributions and Performance of Empirical Skewness Measures”. *Computational Statistics and Data Analysis* 146, 106939.

Eberl, A. and Klar, B. (2021). “A Note on a Measure of Asymmetry”. *Statistical Papers* 62, pp. 1483–1497.

Eberl, A. and Klar, B. (2022a). “Expectile Based Measures of Skewness”. *Scandinavian Journal of Statistics* 49, pp. 373–399.

Eberl, A. and Klar, B. (2022b). “Stochastic Orders and Measures of Skewness and Dispersion Based on Expectiles”. *Statistical Papers*. DOI: 10.1007/s00362-022-01331-x. Online First Article.

Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Probability and Mathematical Statistics, Vol. 1. New York: Academic Press.

Fiori, A. M. and Zenga, M. (2009). “Karl Pearson and the Origin of Kurtosis”. *International Statistical Review* 77, pp. 40–50.

Fishburn, P. C. (1976). “Continua of Stochastic Dominance Relations for Bounded Probability Distributions”. *Journal of Mathematical Economics* 3, pp. 295–311.

Fishburn, P. C. (1980). “Continua of Stochastic Dominance Relations for Unbounded Probability Distributions”. *Journal of Mathematical Economics* 7, pp. 271–285.

Gastwirth, J. L. (1971). “On the Sign Test for Symmetry”. *Journal of the American Statistical Association* 66, pp. 821–823.

Georgii, H.-O. (2013). *Stochastics*. 2nd ed. De Gruyter.

Giovagnoli, A. and Wynn, H. P. (1995). “Multivariate Dispersion Orderings”. *Statistics & Probability Letters* 22, pp. 325–332.

Groeneveld, R. A. and Meeden, G. (1984). “Measuring Skewness and Kurtosis”. *Journal of the Royal Statistical Society. Series D (The Statistician)* 33, pp. 391–399.

Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. Wiley Series in Probability and Statistics. Wiley.

Handcock, M. S. and Morris, M. (1998). “Relative Distribution Methods”. *Sociological Methodology* 28, pp. 53–97.

Handcock, M. S. and Morris, M. (1999). *Relative Distribution Methods in the Social Sciences*. Statistics for Social Science and Public Policy. Springer.

Hanook, S., Shahbaz, M. Q., Mohsin, M., and Kibria, B. M. G. (2013). "A Note On Beta Inverse-Weibull Distribution". *Communications in Statistics — Theory and Methods* 42, pp. 320–335.

Hewitt, E. and Stromberg, K. (1975). *Real and Abstract Analysis*. 3rd ed. Springer.

Hildebrand, D. K. (1971). "Kurtosis Measures Bimodality?" *The American Statistician* 25, pp. 42–43.

Hoaglin, D. C. (1985). "Summarizing Shape Numerically: The g-and-h Distributions". *Exploring Data Tables, Trends, and Shapes*. Wiley Series in Probability and Statistics. Wiley.

Hopkins, K. D. and Weeks, D. L. (1990). "Tests for Normality and Measures of Skewness and Kurtosis: Their Place in Research Reporting". *Educational and Psychological Measurement* 50, pp. 717–729.

Horn, R. A. and Johnson, C. R. (2013). *Matrix Analysis*. 2nd ed. Cambridge University Press.

Hosking, J. R. M. (1989). "Some Theoretical Results Concerning L-Moments". *Research Report RC14492. IBM Research*.

Hosking, J. R. M. (1990). "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics". *Journal of the Royal Statistical Society. Series B (Methodological)* 52, pp. 105–124.

Hürlimann, W. (2002). "On Risk And Price: Stochastic Orderings And Measures". *Transactions of the 27th International Congress of Actuaries*.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*. 2nd ed. Vol. 1. Wiley.

Johnson, N. L., Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J., and Mykytka, E. F. (1980). "Letters to the Editor". *Technometrics* 22, p. 135.

Jones, M. C. (1994). "Expectiles and M-Quantiles are Quantiles". *Statistics & Probability Letters* 20, pp. 149–153.

Jones, M. C. and Pewsey, A. (2009). "Sinh-Arcsinh Distributions". *Biometrika* 96, pp. 761–780.

Jones, M. C., Rosco, J. F., and Pewsey, A. (2011). "Skewness-Invariant Measures of Kurtosis". *The American Statistician* 65, pp. 89–95.

Kallenberg, O. (2021). *Foundations of Modern Probability*. 3rd ed. Probability Theory and Stochastic Modelling. Springer.

Karlin, S. (1968). *Total Positivity*. Stanford University Press.

Karlin, S. and Novikoff, A. (1963). "Generalized Convex Inequalities". *Pacific Journal of Mathematics* 13, pp. 1251–1279.

Keating, C. and Shadwick, W. F. (2002). "A Universal Performance Measure". *Journal of Performance Measurement* 6, pp. 59–84.

Klar, B. (2002). “A Note on the $\mathcal{L}$-Class of Life Distributions”. *Journal of Applied Probability* 39, pp. 11–19.

Klar, B. (2015). “A Note on Gamma Difference Distributions”. *Journal of Statistical Computation and Simulation* 85, pp. 3708–3715.

Klar, B., Petney, T. N., and Taraschewski, H. (2010). “Quantifying Differences in Parasite Numbers Between Samples of Hosts”. *Journal of Parasitology* 96, pp. 856–861.

Klenke, A. and Mattner, L. (2010). “Stochastic Ordering of Classical Discrete Distributions”. *Advances in Applied Probability* 42, pp. 392–410.

Koenker, R. (2005). *Quantile Regression.* 1st ed. Econometric Society Monographs. Cambridge University Press.

Kourkoulos, M. and Tzanakis, C. (2010). “History, and Students’ Understanding of Variance In Statistics”. *BSHM Bulletin* 25, pp. 168–178.

Lahcene, B. (2013). “On Pearson Families of Distributions and Its Applications”. *African Journal of Mathematics and Computer Science Research* 6, pp. 108–117.

Lewis, T. and Thompson, J. W. (1981). “Dispersive Distributions, and the Connection between Dispersivity and Strong Unimodality”. *Journal of Applied Probability* 18, pp. 76–90.

MacGillivray, H. L. (1985). “A Crossing Theorem for Distribution Functions and Their Moments”. *Bulletin of the Australian Mathematical Society* 31, pp. 413–419.

MacGillivray, H. L. (1986). “Skewness and Asymmetry: Measures and Orderings”. *The Annals of Statistics* 14, pp. 994–1011.

MacGillivray, H. L. and Balanda, K. P. (1988). “The Relationship between Skewness and Kurtosis”. *Australian Journal of Statistics* 30, pp. 319–337.

MacGillivray, H. L. and Cannon, W. H. (1997). “Generalisations of the g-and-h Distributions and Their Uses”. Only available from the first author.

Mann, H. B. and Whitney, D. R. (1947). “On a Test of Whether One of Two Random Variables Is Stochastically Larger Than the Other”. *The Annals of Mathematical Statistics* 18, pp. 50–60.

Martins, L. R. (1965). “Significance of Skewness and Kurtosis in Environmental Interpretation”. *Journal of Sedimentary Research* 35, pp. 768–770.

McAlevey, L. G. and Stent, A. F. (2018). “Kurtosis: A Forgotten Moment”. *International Journal of Mathematical Education in Science and Technology* 49, pp. 120–130.

Mitrinović, D. S. (1970). *Analytic Inequalities.* 1st ed. Vol. 165. Die Grundlehren der Mathematischen Wissenschaften. Springer.

Moors, J. J. A. (1988). “A Quantile Alternative for Kurtosis”. *Journal of the Royal Statistical Society, Series D (The Statistician)* 37, pp. 25–32.

Mühlbach, G. (1973). "A Recurrence Formula for Generalized Divided Differences and Some Applications". *Journal of Approximation Theory* 9, pp. 165–172.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks.* Wiley Series in Probability and Statistics. Chichester: Wiley.

Newey, W. K. and Powell, J. L. (1987). "Asymmetric Least Squares Estimation and Testing". *Econometrica* 55, pp. 819–847.

Nørlund, N. E. (1926). *Leçons sur les Séries d'Interpolation.* Collection de Monographies sur la Theéorie des Fonctions. Gauthier-Villars.

Oja, H. (1981). "On Location, Scale, Skewness and Kurtosis of Univariate Distributions". *Scandinavian Journal of Statistics* 8, pp. 154–168.

Oja, H. (1983). "Descriptive Statistics for Multivariate Distributions". *Statistics & Probability Letters* 1, pp. 327–332.

Patil, P. N., Patil, P. P., and Bagkavos, D. (2012). "A Measure of Asymmetry". *Statistical Papers* 53, pp. 971–985.

Pearson, K. (1895). "Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material". *Philosophical Transactions of the Royal Society of London. A* 186, pp. 343–414.

Pearson, K. (1901). "Mathematical Contributions to the Theory of Evolution. X. Supplement to a Memoir on Skew Variation". *Philosophical Transactions of the Royal Society of London. A* 197, pp. 443–459.

Pearson, K. (1902). "On the Mathematical Theory of Errors of Judgement, with Special Reference to the Personal Equation". *Philosophical Transactions of the Royal Society of London. A* 198, pp. 235–299.

Pearson, K. (1916). "Mathematical Contributions to the Theory of Evolution. — XIX. Second Supplement to a Memoir on Skew Variation". *Philosophical Transactions of the Royal Society of London. A* 216, pp. 429–457.

Popoviciu, T. (1933). "Sur Quelques Propriétés des Fonctions d'Une ou de Deux Variables Réelles". Doctoral Thesis. La Faculté des Sciences de Paris.

Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R., and Mykytka, E. F. (1979). "A Probability Distribution and its Uses in Fitting Data". *Technometrics* 21, pp. 201–214.

Ramos, H. M. and Sordo, M. A. (2003). "Dispersion Measures and Dispersive Orderings". *Statistics & Probability Letters* 861, pp. 123–131.

Rémillard, B. (2013). *Statistical Methods for Financial Engineering.* CRC Press.

Rhind, A. (1909). "Tables to Facilitate the Computation of the Probable Errors of the Chief Constants of Skew Frequency Distributions". *Biometrika* 7, pp. 127–147.

Rolski, T. (1976). "Order Relations in the Set of Probability Distributions and Their Applications in Queuing Theory". *Dissertationes Mathematicae* 132.

Rolski, T. and Stoyan, D. (1974). "Two Classes of Semi-Orderings and their Application in the Queuing Theory". *Journal of Applied Mathematics and Mechanics* 54, pp. 127–128.

Rosco, J. F., Jones, M. C., and Pewsey, A. (2011). "Skew t Distributions via the Sinh-Arcsinh Transformation". *TEST* 20, pp. 630–652.

Runnenburg, J. T. (1978). "Mean, Median, Mode". *Statistica Neerlandica* 32, pp. 73–79.

Ruppert, D. (1987). "What is Kurtosis?: An Influence Function Approach". *The American Statistician* 41, pp. 1–5.

Rüschendorf, L. (1981). "Stochastically Ordered Distributions and Monotonicity of the OC-Function of Sequential Probability Ratio Tests". *Statistics: A Journal of Theoretical and Applied Statistics* 12, pp. 327–338.

Rüschendorf, L. (2009). "On the Distributional Transform, Sklar's Theorem, and the Empirical Copula Process". *Journal of Statistical Planning and Inference* 139, pp. 3921–3927.

Saunders, I. W. (1978). "Locating Bright Spots in a Point Process". *Advances in Applied Probability* 10, pp. 587–612.

Saunders, I. W. and Moran, P. A. P. (1978). "On the Quantiles of the Gamma and F Distributions". *Journal of Applied Probability* 15, pp. 426–432.

Shaked, M. (1982). "Dispersive Ordering of Distributions". *Journal of Applied Probability* 19, pp. 310–320.

Shaked, M. and Shanthikumar, J. G. (2006). *Stochastic Orders.* Springer Series in Statistics. Springer.

Shorack, G. R. (2017). *Probability for Statisticians.* 2nd ed. Springer Texts in Statistics. Springer.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley Series in Probability and Mathematical Statistics. Wiley.

Sillitto, G. P. (1951). "Interrelations Between Certain Linear Systematic Statistics of Samples from Any Continuous Population". *Biometrika* 38, pp. 377–382.

Sordo, M. A., de Souza, M. C., and Suárez-Llorens, A. (2016). "Testing Variability Orderings by Using Gini's Mean Differences". *Statistical Methodology* 32, pp. 63–76.

Tajuddin, I. H. (1999). "A Comparison Between Two Simple Measures of Skewness". *Journal of Applied Statistics* 26, pp. 767–774.

Tracy, R. L. and Doane, D. P. (2005). "Using the Studentized Range to Assess Kurtosis". *Journal of Applied Statistics* 32, pp. 271–280.

Tukey, J. W. (1977). "Modern Techniques in Data Analysis". SF-Sponsored Regional Research Conference at Southern Massachusetts University. North Dartmouth.

Von Hippel, P. T. (2005). "Mean, Median, and Skew: Correcting a Textbook Rule". *Journal of Statistics Education* 13, pp. 965–971.

Westfall, P. H. (2014). "Kurtosis as Peakedness, 1905–2014. R.I.P." *The American Statistician* 68, pp. 191–195.

Wheeler, D. J. (1975). "An Approximation for Simulation of Gamma Distributions". *Journal of Statistical Computation and Simulation* 3, pp. 225–232.

Whitt, W. (1980). "The Effect of Variability in the GI/G/s Queue". *Journal of Applied Probability* 17, pp. 1062–1071.

Yule, G. U. (1911). *An Introduction to the Theory of Statistics*. 1st ed. Charles Griffin & Co., Ltd.

Van Zwet, W. R. (1964). *Convex Transformations of Random Variables*. Mathematical Centre Tracts, Amsterdam.

Van Zwet, W. R. (1979). "Mean, Median, Mode II". *Statistica Neerlandica* 33, pp. 1–5.