Quantifying differences in parasite numbers between samples of hosts

Bernhard Klar[a]*, Trevor N. Petney[b], Horst Taraschewski[b]

[a]*Institute for Stochastic, Karlsruhe Institute of Technology (KIT), Kaiserstrasse 89, 76133 Karlsruhe, Germany*

[b]*Department of Ecology and Parasitology, Karlsruhe Institute of Technology (KIT), Kornblumenstrasse 13, 76128 Karlsruhe, Germany*

* Corresponding author. Tel.: +49 721 6082047; fax: +49 721 6086066

   E-mail address: Bernhard.Klar@kit.edu (B. Klar)

Abstract

An important question in many parasitological studies is the assessment of differences in parasite numbers between samples of hosts. This is not always easy: while almost everybody will agree that the main task consists in deciding whether the values in one sample tend to be higher than the values of the other sample, there is considerable disagreement about what higher (or lower) should mean. In common use as dissimilarity measures are differences between mean values, medians, geometric means, prevalence rates, relative effects, and more. In general, different measures can lead to different conclusions. However, a debate as to which measure is superior is fruitless; it depends on goals and circumstances of the respective study. In our opinion, it is more important to identify situations in which most of the above mentioned measures coincide, and hence, one can confidently claim that the values in one sample are higher than in the other. This is the case when one sample is stochastically larger than the second. It is the aim of this paper to review this concept using distributional and data examples, and of proposing graphical tools for detecting stochastic dominance.

1. Introduction

In a good part of parasitological research, the quantification and comparison of parasite numbers in different samples, recorded at different times and/or in different habitats, is an important task. The most commonly used measure for this purpose is probably the difference between mean values, i.e. mean abundance or mean intensity. Rózsa et al. (2000) argue that terminological recommendations require the use of these measures. Since small sample sizes can lead to difficulties when using the traditional t-test or Welch's modification, they propose a bootstrap test based on Welch's statistic. They criticize the use of nonparametric tests such as the Mann-Whitney U-test because these tests compare other characteristics instead of means. Indeed, assuming we have samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ of size m and n, the U-test corrected for ties is based on the statistic

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[1(\textit{if } X_i > Y_j) + \frac{1}{2}(\textit{if } X_i = Y_j) \right]$$

(see, e.g., Hollander and Wolfe 1999, p.118). $U/(m \cdot n)$ is an estimator of the relative effect $p = \Pr(X > Y) + (1/2)\Pr(X = Y)$. If $p > 1/2$, observations tend to be larger in group 1 in comparison to group 2 in a certain sense, but the mean of group 1 is not necessarily larger than the mean of group 2. The converse is true for $p < 1/2$. If both groups have the same distribution, then $p = 1/2$. However, the opposite does not hold: if $p = 1/2$, the underlying distributions need not be the same.

In two recent articles, however, Neuhäuser and Poulin (2004) and Neuhäuser and Ruxton (2009) advocate the use of relative effect, since "means are not very useful descriptors for skewed distributions", and they propose the use of the Brunner und Munzel (2000) test which is also based on this measure. They conclude that the relative effect would be the natural measure for a difference between two samples since "the main question is whether the values in one sample tend to be larger (or smaller) than the values of the other sample",

implying that only the relative effect provides the answer to this question. A bootstrap test for the hypothesis $p = 1/2$ based on the rank Welch test statistic is proposed by Reiczigel et al. (2005).

There are other measures in common use, like the comparison of medians or prevalence rates, and clearly, in a specific situation, the use of different measures can lead to different conclusions. However, in our opinion, a dispute about the 'best' measure is fruitless since this depends on goals and circumstances of the respective study. For example, if an infection with a certain parasite is invariably lethal, prevalence rates are adequate measures. In other cases, infection with a moderate number of parasites has no demonstrable negative effects on the host, but a very large number of parasites causes significant morbidity; in such cases, means which give more weight on high intensities are suitable measures. Situations which lie in between call for intermediate measures such as medians or relative effects.

Unfortunately, since the impact of parasites on hosts is seldom known precisely, it is difficult to justify the preference for a particular measure even in specific situations. For this reason, it is important to identify situations in which the above mentioned measures coincide, and one can therefore confidently claim that the values in one sample are higher than in the other. To this aim, it is not sufficient to compare a special characteristic of two distributions, one has to compare the whole distributions. A suitable tool is to determine whether the two distributions are stochastically ordered. We review this concept using distributional and data examples, and propose graphical tools for detecting stochastic dominance.

2. Stochastic dominance between discrete distributions

The (usual) stochastic order is an established concept in probability theory and is in frequent use in reliability theory and econometrics. Since we are interested in comparing parasite numbers, we give the definition for discrete distributions: A discrete random variable

X with cumulative distribution function (cdf) $F_X(k) = \Pr(X \le k)$ is said to be larger than another discrete random variable Y with cdf $F_Y(k) = \Pr(Y \le k)$ with respect to *stochastic order* (written $X \ge_{st} Y$) if

$$1 - F_X(k) \ge 1 - F_Y(k) \quad \text{for} \quad k = 0,1,2,\ldots,$$

and it is called *strictly larger* if $\ge$ can be replaced by $>$ for at least one k (otherwise, the two distributions can be equal). We also say that X is stochastically larger than Y, that X stochastically dominates Y, or that Y is stochastically smaller than X. Obviously, an equivalent condition for $X \ge_{st} Y$ is given by $F_Y(k) - F_X(k) \ge 0$ for all k.

For our purpose, the most important properties of this order between random variables are the following:

- If $X \ge_{st} Y$ then the (theoretical) prevalence $P(X \ge 1)$ of X is larger than the prevalence of Y. This follows directly from the definition putting $k = 0$:

  $$P(X \ge 1) = 1 - F_X(0) \ge 1 - F_Y(0) = P(Y \ge 1).$$

- If $X \ge_{st} Y$ then $EX \ge EY$. This follows from the definition of stochastic order together with the representation of the population mean $EX = \sum_{k \ge 0}\left(1 - F_X(k)\right)$.

- If $X \ge_{st} Y$, then all quantiles of the distribution of X are larger than the corresponding quantiles of Y. In particular, the median of X is larger than the median of Y.

- If $X \ge_{st} Y$ then $p \ge 1/2$ holds for the relative effect (for a proof, see Appendix I).

Hence, if $X \ge_{st} Y$, the typical measures for assessing differences between samples all point in the same direction. Examples for discrete distributions which are stochastically ordered will be dealt with in section 5.

Considering the implications of stochastic order for empirical data: if we look at data sets, probabilities $P(X = k)$ are replaced by relative frequencies $r(k)$, $k = 0,1,\ldots$, and the cdf is replaced by the cumulative relative frequencies $R(k) = r(0) + \ldots + r(k)$ which are sometimes called the empirical distribution function (edf). Since the edf has all the properties of a cdf, the four properties mentioned above carry over immediately to the corresponding empirical measures. For example, if $R_Y(k) - R_X(k) \geq 0$ for all k, (sample) prevalence and sample mean of the x-sample are larger than the corresponding measures of the y-sample.

2.1. Example 1: nematode larvae in Japanese eels

The first example compares counts of larvae of swimbladder-nematodes in two populations of the Japanese eel (*Anguilla japonica*) from southwest Taiwan. Münderle et al. (2006) compared wild eels from the Kao-Ping river (sample 2, $n = 168$) to cultured eels from an adjacent aquaculture farm (sample 1, $n = 71$). All recorded nematodes were of species *Anguillicoloides crassus* (previously *Anguillicola crassus*).

Absolute frequencies and the differences in (cumulative) relative frequencies are given in Table 1. Since all entries in the last row are positive, sample 1 is stochastically larger than sample 2. From the above discussion, we can conclude that the difference between prevalences, means and medians are positive and that the relative effect is larger than 1/2. The exact values are given in the second column of Table 2. In this example, it is fully justified to state that the eels from the aquaculture farm are more heavily infected than the wild eels.

Table 1

Differences in (cumulative) relative frequencies for example 1

k	0	1	2	3	4	5	6	7	8	>8
Abs. freq. sample 1	32	15	8	4	1	1	3	2	1	4
Abs. freq. sample 2	134	19	9	0	1	2	0	1	1	1
$r_Y(k) - r_X(k)$	0.35	-0.10	-0.06	-0.06	-0.01	0.00	-0.04	-0.02	-0.01	-0.05
$R_Y(k) - R_X(k)$	0.35	0.25	0.19	0.13	0.12	0.12	0.08	0.06	0.05	0.00

Table 2

Empirical measures for examples 1-4

	Example 1	Example 2	Example 3	Example 4
Sample sizes	71 and 168	196 and 100	40 and 20	10 and 10
Diff. between prevalences	0.35	0.08	-0.3	-
Diff. between means	2.85	-0.05	5.85	3
Diff. between medians	1	0	-5	-1
Relative effect	0.69	0.53	0.3	0.22

3. Graphical tools for detecting stochastic order

From the definition, two simple graphical procedures for checking stochastic order between two given discrete distributions are evident. The first is a bar plot of the difference $F_Y(k) - F_X(k)$ with positive bars if X stochastically dominates Y, and negative values if Y stochastically dominates X. This plot is called a probability difference plot (P-D plot) in what follows. Alternatively, a scatter plot of $F_Y(k)$ against $F_X(k)$ can be drawn. If one plots the values of $F_X(k)$ on the x-axis, the points are above the diagonal if X stochastically dominates Y. This plot is known as probability-probability plot (P-P plot).

A further possible diagnostic plot is a quantile-quantile plot, which is also well known in the graphical analysis of statistical data. However, for discrete distributions, it is less suitable, particularly due to the need for choosing suitable plotting positions.

In all plots, $F_Y(k)$ and $F_X(k)$ have to be replaced with $R_Y(k)$ and $R_X(k)$ when dealing with observed data instead of probability distributions.

The P-D plot and the P-P plot for example 1 in section 2.1 are shown in Figures 1 and 2, respectively. As expected from the discussion in section 2.1, all values in Figure 1 are positive, and all points in Figure 2 are well above the diagonal. Hence, each plot shows at a glance the stochastic ordering between the two samples.
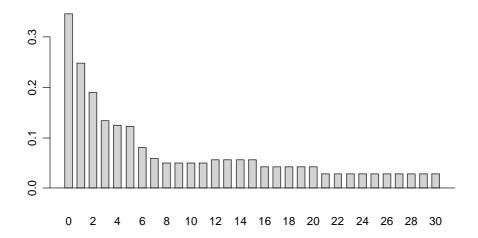
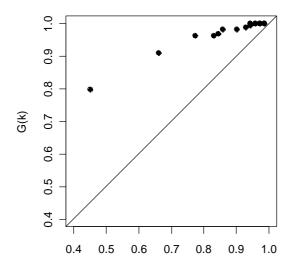Figure 1: Probability-difference plot for the two samples in example 1

Figure 2: Probability-probability plot for the two samples in example 1

4. Further data examples

In this section, we discuss situations in which the two samples are not stochastically ordered. In examples 2 and 3, swimbladder and intestinal nematodes in European eels (*Anguilla anguilla*) are considered. The last example deals with two small samples of infected hosts.

4.1. Example 2: nematode larvae in Rhine eels

As second example, we consider counts of larvae of swimbladder-nematodes in European eels from two different locations: sample 1 ($n = 196$) from the River Rhine near Karlsruhe, and sample 2 ($n = 100$) from the River Rhine near Sulzbach (Münderle 2005). Again, all recorded nematodes were *A. crassus*. Measures such as the difference between means can be found in the third column of Table 2.

The P-D plot (Fig. 3) and the P-P plot (Fig. 4) show that the two samples are not stochastically ordered. Since the bars in Figure 3 are rather short (i.e. the differences $R_Y(k) - R_X(k)$ are small), we may assume that both samples come from the same distribution. This can be confirmed by a test of the hypotheses $F_X = F_Y$ against the alternative $F_X \neq F_Y$. A suitable test for this purpose is a (two-tailed) two-sample Kolmogorov-Smirnov test yielding a non-significant p-value of 0.8.
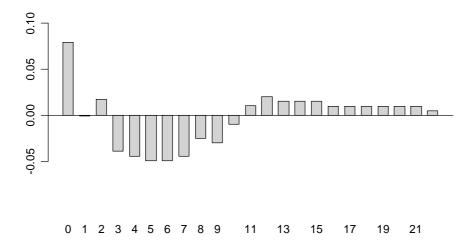
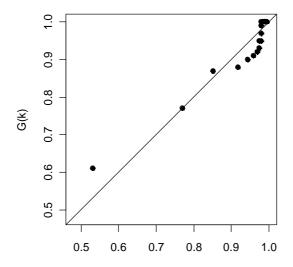Figure 3: P–D plot for the two samples in example 2

Figure 4: P-P plot for the two samples in example 2

4.2. Example 3: intestinal parasites in Rhine eels

The third example compares two samples of counts of intestinal parasites in European eels at the same location (River Rhine near Karlsruhe), but from different years: sample 1 ($n = 40$) was recorded in summer 1999, sample 2 ($n = 20$) in summer 2005 (Thielen 2006).

As in example 2, the P-D (Fig. 5) and P-P plots (Fig. 6) show that the two samples are not stochastically ordered. However, looking at the plots, it is quite obvious that the two samples do not stem from the same distributions. The prevalence in the second sample is 30% higher than in the second, and also the cumulative relative frequencies $R(k)$ are much higher in the second sample for values of k up to 10. The pattern changes for large values of k : there are noticeably more heavily infected eels in sample 1 than in sample 2. These findings are reflected in the numerical values in the fourth column of Table 2.
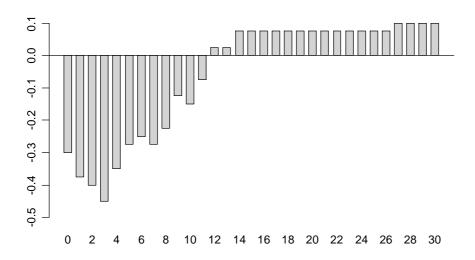

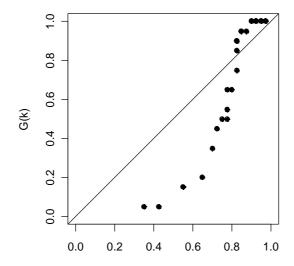Figure 5: P-D plot for the two samples in example 3

Figure 6: P-P plot for the two samples in example 3

The reason behind this change in the distribution of intestinal parasites between 1999 and 2005 is explained in Thielen (2006): the parasite *Paratenuisentis ambiguous*, which was predominant in 1999, vanished due to the extinction of an alternate host and was replaced by several other species (*Pomphorhynchus laevis, Raphidascaris acus, Paraquimperia tenerrima, Proteocephalus macrocephalus, Bothriocephalus claviceps*).

4.3. Example 4: artificial dataset of Rózsa et al. (2000)

As last data example, we consider the dataset discussed in Rózsa et al. (2000) and Neuhäuser and Poulin (2004). Here, 2 samples of infected hosts are considered:

Sample A: 1, 1, 1, 1, 1, 1, 1, 1, 2, 50;

Sample B: 1, 1, 2, 2, 2, 2, 3, 3, 4, 10.

Even though this example deals with intensity rather than abundance, the previous discussion remains valid taking into account that prevalence here has no meaning, and that frequencies are now defined for $k = 1, 2, \ldots$.

Difference between means is positive; however, the difference between medians is negative, and the relative effect is well below 1/2 (Table 2, last column). The P-D plot (Fig. 7) clearly shows that neither sample A stochastically dominates sample B nor B stochastically dominates A. Hence, one cannot generally speak of one sample tending to larger values than the other; one has to be more specific: "Sample A has a larger mean intensity than sample B" or "median intensity is smaller in sample A".
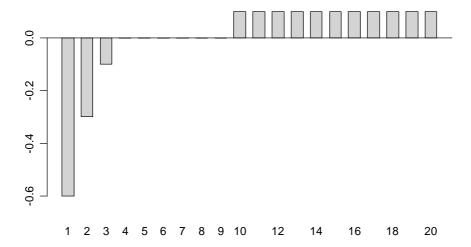
Figure 7: Probability-difference plot for the two samples in example 4

5. Examples of stochastically ordered discrete probability distributions

The opinion is widespread that stochastic ordering between probability distributions is uncommon. Presumably, this is due to the example of the predominant normal distribution: two normal distributions can only be stochastically ordered if they have equal variances (and then, the one with larger mean stochastically dominates the other). However, since equal variance is certainly the exception in practice, the same holds for stochastic ordering between two normal distributions. This finding is connected with the fact that the support of a normal distribution is unbounded in both directions.

This changes entirely if we consider distributions defined only for positive values such as exponential, gamma, Weibull or count distributions. We consider the three most important count distributions, the Poisson, the binomial, and the negative binomial distributions. Particularly the last distribution is often used to model parasite numbers since data show a high degree of overdispersion (Anderson and Gordon 1985). As a final example, we consider the logarithmic distribution, which is long and widely used for modelling intensities (see, e.g., Williams 1964).

5.1. Poisson distributions

Let X be distributed according to a Poisson distribution with probability density function (pdf) $\lambda^k \exp(-\lambda)/k!, k = 0,1,\ldots$, and let Y be Poisson-distributed with pdf $\mu^k \exp(-\mu)/k!, k = 0,1,2,\ldots$. Then, $X \geq_{st} Y$ if and only if $\lambda \geq \mu$, i.e. if the mean of X is larger or equal to the mean of Y (Appendix II).

5.2. Binomial distributions

Let X have a binomial distribution with pdf $\binom{m}{k} p^k (1-p)^{m-k}, k = 0,\ldots,m$, and let Y be binomial distributed with pdf $\binom{n}{k} q^k (1-q)^{n-k}, k = 0,\ldots,n$. Then, $X \geq_{st} Y$ if $p \geq q$ and $m \geq n$, and in this case, $EX \geq EY$ and $Var(X) \geq Var(Y)$ (Appendix II).

5.3. Negative binomial distributions

Let X and Y have negative binomial distributions with pdf $\binom{k+r-1}{r-1} p^r (1-p)^k, k = 0,1,2,\ldots$, and pdf $\binom{k+s-1}{s-1} q^s (1-q)^k, k = 0,1,2,\ldots$, respectively. Then, $X \geq_{st} Y$ if $p \leq q$ and $r \geq s$. In this case, $EX \geq EY$ and $Var(X) \geq Var(Y)$ holds again (Appendix II).

5.4. Logarithmic distributions

Let X have a logarithmic distribution with pdf $-[\ln(1-p)]^{-1} p^k / k, k = 1,2,3,\ldots$, and let Y be logarithmic distributed with pdf $-[\ln(1-q)]^{-1} q^k / k, k = 1,2,\ldots$. Then, $X \geq_{st} Y$ if and only if $p \geq q$ (Appendix II).

Of course, if we have stochastic ordering in one of the four examples above, not only are the means and variances ordered, but also medians, etc. as stated in section 2.

If, on the other hand, the mean of X is smaller than the mean of Y but the variances are ordered conversely, both samples will not be stochastically ordered in general.

As examples, we fitted negative binomial distributions to the datasets in example 1 and 3. This distribution provides a very good model for the data under consideration (see Münderle et al. (2006) for the first example).

In example 1, we obtain $r = 0.28, p = 0.08$ for the first sample and $s = 0.17, q = 0.27$ for the second sample. Since $r > s$ and $p < q$, we can conclude from the result of section 5.3 that the fitted negative binomial distribution for the first sample stochastically dominates the second, in agreement with the previous findings.

In example 3, we have $r = 0.23, p = 0.01$ and $s = 1.60, q = 0.15$. Here, $r < s$ and $p < q$, and a P-D plot (Fig. 8) shows that the two fitted distributions are not stochastically ordered, which is again consistent with the previous results for this dataset.
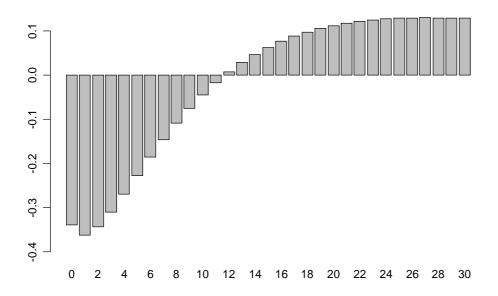

Figure 8: P-D plot for the fitted negative binomial distribution for example 3

6. Conclusion

An important task in many parasitological studies consists in deciding whether the values in one sample tend to be higher than the values of the other sample. However, different dissimilarity measures can lead to different conclusions. A debate as to which measure is superior is unproductive. A similar view with regard to formal tests for the relative effect is expressed in Neuhäuser and Ruxton (2009): "Testing the null hypothesis p=0.5 is not necessarily a better or worse approach than comparing central tendencies such as means or medians. However, it offers a different point-of-view that at least sometimes is more meaningful."

Using the concept of stochastic dominance we proposed tools which help to identify situations in which the most common measures all lead to the same conclusion. Only then it is fully justified to make a general statement that the number of parasites in one sample is larger than in the other sample. Several examples corroborate the applicability of the concept to empirical data as well as to theoretical count distributions.

References

Anderson, R.M., Gordon, D.M., 1982. Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite induced host mortalities. Parasitology 85, 373-398.

Brunner, E., Munzel, U., 2000. The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation. Biometrical Journal 42, 17-25.

Hollander, M., Wolfe, D.A., 1999. Nonparametric statistical methods. John Wiley and Sons, Chichester.

Müller, A., Stoyan, D., 2002. Comparison methods for stochastic models and risks. John Wiley and Sons, Chichester.

Münderle M., 2005. Ökologische, morphometrische und genetische Untersuchungen an Populationen des invasiven Schwimmblasennematoden *Anguillicola crassus* aus Europa und Taiwan. PhD Dissertation, University of Karlsruhe, Karlsruhe, Germany.

Münderle, M., Taraschewski, H., Klar, B., Chang, C.W., Shiao, J.C., Shen, K.N., He, J.T., Lin, S.H., Tzeng, W.N., 2006. Occurrence of *Anguillicola crassus* (Nematoda: Dracunculoidea) in Japanese eels *Anguilla japonica* from a river and an aquaculture unit in southwest Taiwan. Dis. Aquat. Organ. 71, 101-108.

Neuhäuser, M., Poulin, R., 2004. Comparing parasite numbers between samples of hosts. J. Parasitol. 90, 689-691.

Neuhäuser, M., Ruxton, G.D., 2009. Distribution-free two-sample comparisons in the case of heterogeneous variances. Behav. Ecol. Sociobiol. 63, 617-623.

Rózsa, L., Reiczigel, J., Majoros G., 2000. Quantifying parasites in samples of hosts. J Parasitol 86, 228-232.

Reiczigel, J., Zakariás, I., Rózsa, L., 2005. A bootstrap test of stochastic equality of two populations. Am. Stat. 59, 156-161.

Thielen, F., 2006. Der Einfluss einwandernder Amphipodenarten auf die Parasitozönose des Europäischen Aals (*Anguilla anguilla*). PhD Dissertation, University of Karlsruhe, Karlsruhe, Germany.

Williams, C.B., 1964. Patterns in the balance of nature. Academic Press, London.

Appendix I

For the relative effect p, the following chain of equations hold:

$$p = P(X > y) + (1/2)P(X = Y)$$

$$= \sum_{k \geq 0} \left[P(X > k \mid Y = k) + (1/2)P(X = k \mid Y = k) \right] \cdot P(Y = k)$$

$$= \sum_{k \geq 0} \left[P(X > k) + (1/2)P(X = k) \right] \cdot P(Y = k)$$

$$= \sum_{k \geq 0} \left[(1/2)P(X > k) + (1/2)P(X \geq k) \right] \cdot P(Y = k)$$

$$= (1/2) \sum_{k \geq 0} \left[(1 - F_X(k)) + (1 - F_X(k-1)) \right] \cdot P(Y = k).$$

Now assume that X is stochastically larger than Y, i.e. $1 - F_X(k) \geq 1 - F_Y(k)$ holds for all

k. Then

$$p \geq (1/2) \sum_{k \geq 0} \left[(1 - F_Y(k)) + (1 - F_Y(k-1)) \right] \cdot P(Y = k)$$

$$= \sum_{k \geq 0} \left[P(Y > k \mid Y = k) + (1/2)P(Y = k \mid Y = k) \right] \cdot P(Y = k)$$

$$= 0 + (1/2) \sum_{k \geq 0} 1 \cdot P(Y = k)$$

$$= 1/2.$$

Hence, if $X \geq_{st} Y$, then the relative effect is larger or equal to 1/2.

Appendix II

A discrete random variable X (with pdf f_X) is said to be larger than another discrete random variable Y (with pdf f_Y) *in likelihood ratio order* (written $X \geq_{lr} Y$) if

$$f_X(k) / f_Y(k) \leq f_X(k+1) / f_Y(k+1) \qquad \text{for all} \quad k.$$

To show that X is stochastically larger than Y, it is sufficient to show $X \geq_{lr} Y$ since the latter implies the first (see, e.g., Müller and Stoyan 2002, p. 12).

For the Poisson distribution, it is easy to see that $X \geq_{lr} Y$ if $\lambda \geq \mu$. Thus, we also have $X \geq_{st} Y$ in this case. On the other hand, if $X \geq_{st} Y$, then $EX \geq EY$, i.e. $\lambda \geq \mu$. This shows the statement in section 5.1. The assertions in the other examples can be obtained by similar reasoning.