Karlsruher Institut für Technologie

Masterarbeit

Nonparametric Distribution Function Estimation

Nichtparametrische Schätzung von Verteilungsfunktionen

Ariane Hanebeck

05.04.2020

Betreuung: PD Dr. Bernhard Klar

Fakultät für Mathematik

Institut für Stochastik

Karlsruher Institut für Technologie

Contents

1 Introduction

This thesis deals with distribution function estimation, which is a very important topic in statistics. In the following, distribution function estimation is defined and the importance of the topic is explained. After that, a short overview of the thesis is given.

What is distribution function estimation? The setting is that we have an infinite sequence of independent and identically distributed (i.i.d.) random variables $X_1, X_2, ...$ that have an underlying unknown distribution function F. Now, the task is to estimate F, given a finite random sample $X_1, ..., X_n, n \in \mathbb{N}$.

In the case of *parametric* distribution function estimation, the model structure is already defined before knowing the data. It is for example known that the distribution will be of the form $\mathcal{N}(\mu, \sigma^2)$. The only goal is to estimate the parameters, here μ and σ^2.

Compared to this, in the *nonparametric* setting, the model structure is not specified a priori but is determined only by the sample. In this thesis, all the considered estimators are of nonparametric type.

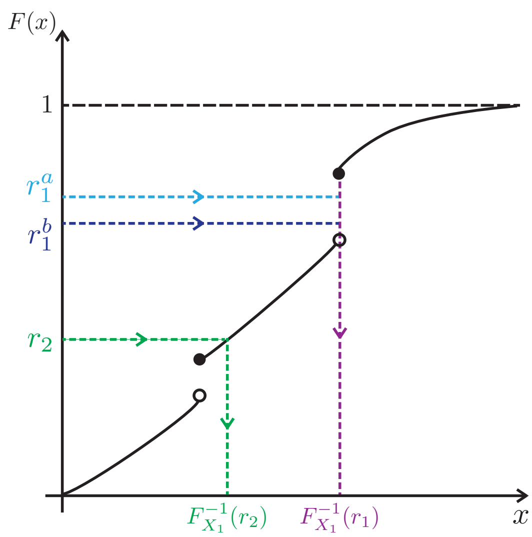

Figure 1: Illustration of the Inverse Transform Sampling.

Why is distribution function estimation important? The goal is to investigate properties of a random sample and its underlying distribution. As the random variables are i.i.d., we here consider the properties of X_1 without loss of generality. One property of the distribution is the probability $\mathbb{P}(a \leq X_1 \leq b) = F(b) - F(a)$, which can directly be estimated without the need to integrate as in the density estimation setting. By taking the inverse of F, it is also possible to calculate quantiles

$$x_p = \inf\{x \in \mathbb{R} : p \leq F(x)\} = F^{-1}(p).$$

Another application of the inverse of F is the so-called Inverse Transform Sampling (ITS). It can be used to generate more samples than already given. The idea is to use the expression

$$Y \sim U[0,1] \Rightarrow F_{X_1}^{-1}(Y) \sim X_1.$$

The intuition of ITS is shown in Figure 1: Given a random number $r \in [0,1]$ that corresponds to Y, the number $F_{X_1}^{-1}(r)$, corresponding to $F_{X_1}^{-1}(Y)$, is the new sample.

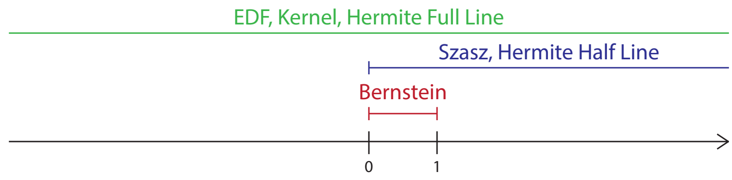

Figure 2: The different intervals of the estimators on the real line.

About this thesis Two very famous distribution function estimators are the empirical density function (EDF) and the kernel estimator. However, there are many other estimators that perform better than these two. Some of them are discussed in this thesis. The goal is to derive properties of the considered estimators and compare them theoretically and in simulation.

One contribution of this thesis is the Szasz estimator that, to the best of my knowledge, is a new way to estimate distribution functions on $[0, \infty)$. Another contribution is that several properties of the estimators in this thesis were not proven before, such as the asymptotic normality of the Hermite estimators.

In Section 2 and Section 3, the EDF and the kernel estimator are introduced. A short summary of the most important definitions and properties is given. In Section 4, the background on function estimation with Bernstein polynomials is explained and in Section 5, the Bernstein distribution function estimator is defined. The Szasz estimator that is derived in Section 6 uses the ideas of the Bernstein estimator but can estimate functions on the real half line. Section 7 on the Hermite estimator is split up into two parts - one for the real half line and the other one for the real line. In Section 8, a theoretical comparison of the estimators is given and in Section 9, the estimators are compared in simulation. In Section 10, the most important findings of the thesis are summarized.

The proofs for each estimator can always be found at the end of the respective section, whereas the proofs for the EDF and the kernel estimator are omitted as the results are well known.

We now give a quick overview of the estimators compared in this thesis.

1.1 The Different Estimators

The first and most obvious difference between the estimators is that all of them are defined on different domains. This is illustrated in Figure 2. The Bernstein estimator for example can only estimate distributions that are supported on $[0, 1]$, while the Szasz estimator and the Hermite estimator on the real half line can estimate distributions supported on $[0, \infty)$. These domains are of course normalized. A distribution on $[a, b], a < b$, can easily be transformed to the unit interval so that the Bernstein estimator can be applied. Furthermore, a distribution on $[-a, \infty), a > 0$, can be shifted to the positive line. However, it is not possible to transform a distribution on $[0, \infty)$ to the unit interval without losing some important properties. This is explained in the beginning of Section 5.

The Bernstein, Szasz, and both Hermite estimators are specifically designed for their respective intervals. The EDF and the kernel estimator serve as a comparison.

It holds that

$$\{\text{Distributions on } [0, 1]\} \subset \{\text{Distributions on } [0, \infty)\} \subset \{\text{Distributions on } (-\infty, \infty)\},$$

which means that the distributions where the Bernstein estimator can be applied are a subset of those where the Szasz estimator can be applied and so on. More explanations about the resulting estimates and their properties can be found in Section 8.

All the estimators use different approaches to estimate the distribution function.

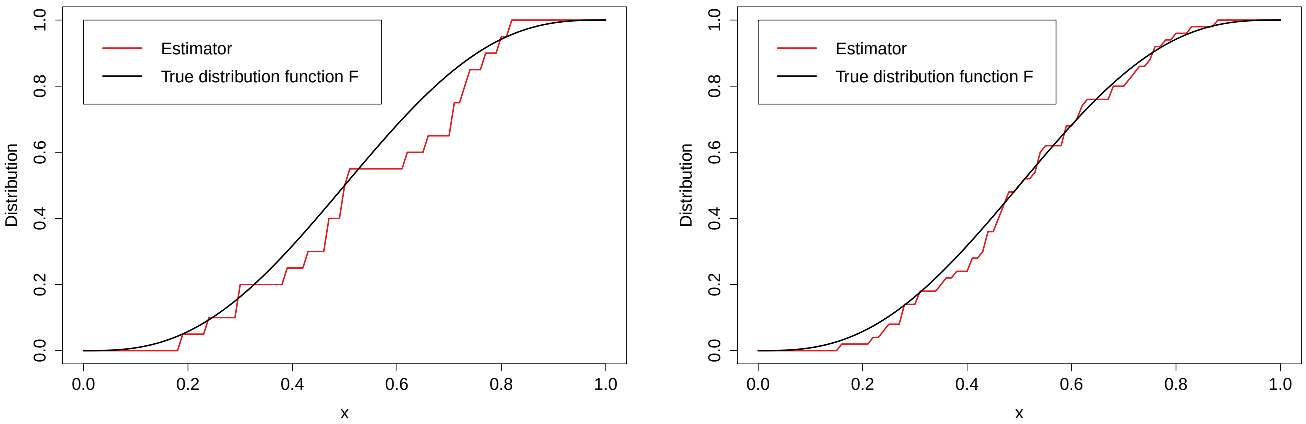

Figure 3: Illustration of the EDF with $n = 20$ and $n = 50$.

2 Empirical Distribution Function

The empirical distribution function (EDF) is the simplest way to estimate the underlying true distribution function, given a finite random sample $X_1, ..., X_n, n \in \mathbb{N}$. The idea is to use the strong law of large numbers. Then, the estimator is defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X_i \leq x),$$

where $\mathbb{I}$ is the indicator function. The estimator is illustrated in Figure 3 for $n = 20$ and $n = 50$. The Glivenko-Cantelli theorem assures the uniform, almost sure convergence of this estimator. In the sequel, some important properties of the EDF are stated. The following theorem follows directly from the central limit theorem.

Theorem 2.1. *For the empirical distribution function it holds for x with $0 < F(x) < 1$ that*

$$n^{1/2}(F_n(x) - \mathbb{E}[F_n(x)]) = n^{1/2}(F_n(x) - F(x)) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right)$$

for $n \to \infty$, where $\sigma^2(x) = F(x)(1 - F(x))$. End Theorem

The next theorem gives the mean squared error (MSE) and the mean integrated squared error (MISE) of the empirical distribution function, which are defined by

$$\mathrm{MSE}[F_n(x)] = \mathbb{E}[(F_n(x) - F(x))^2]$$

and

$$\mathrm{MISE}[F_n] = \mathbb{E}\left[\int_{\mathcal{D}} (F_n(x) - F(x))^2 \,\mathrm{d}x\right],$$

where we integrate over the considered domain $\mathcal{D}$.

Theorem 2.2. *In [1], it can be seen that the MSE of the empirical distribution function is given by*

$$\mathrm{MSE}[F_n(x)] = \mathrm{Var}[F_n(x)] = n^{-1}\sigma^2(x). \tag{2.1}$$

The MISE of the empirical distribution function is given by

$$\mathrm{MISE}[F_n(x)] = n^{-1} \int_{\mathcal{D}} \sigma^2(x) \,\mathrm{d}x,$$

where $\sigma^2(x)$ is defined as in Theorem 2.1. End Theorem

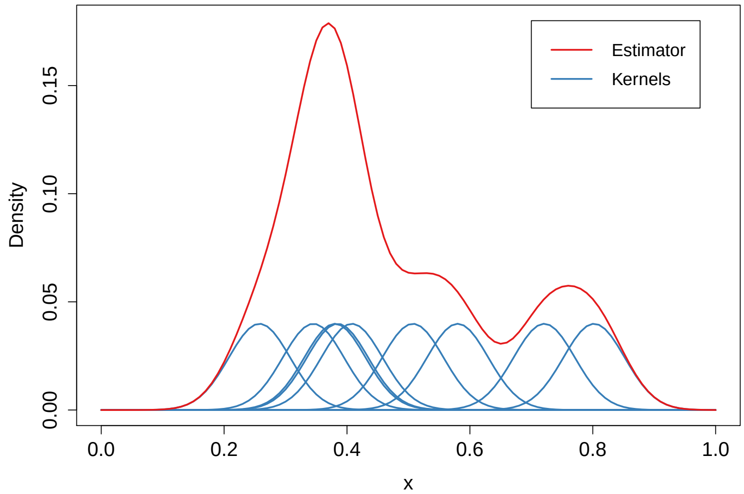

Figure 4: Illustration of the kernel density estimator.

3 Kernel Estimation

One of the most popular density estimators is the kernel density estimator. This estimator is also called the Parzen-Rosenblatt estimator after the two inventors Emanuel Parzen and Murray Rosenblatt who independently came up with the idea, see [2, 3]. It uses a kernel $K : \mathbb{R} \to \mathbb{R}$ that has to fulfill the properties

- $\int K(x)\,\mathrm{d}x = 1$,
- $K(x) \geq 0$ for all x,
- $K(x) = K(-x)$ for all x,
- $\int xK(x)\,\mathrm{d}x = 0$, and
- $\int x^2K(x)\,\mathrm{d}x < \infty$,

following [4]. This means that the kernel is a symmetric density function with zero mean and bounded variance. Some popular kernels are

- the Normal/Gaussian kernel: $K(x) = (2\pi)^{-1/2}\exp\left(-\frac{1}{2}x^2\right)$,
- and the Epanechnikov kernel: $K(x) = \frac{3}{4}\left(1 - x^2\right)\mathbb{I}\left(|x| \leq 1\right)$.

Let $X_1, X_2, ...$ be i.i.d. random variables that have an underlying unknown distribution function F and unknown density function f. Given a finite random sample $X_1, ..., X_n, n \in \mathbb{N}$, the univariate kernel density estimator is defined by

$$f_{h,n}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), x \in \mathbb{R},$$

where the parameter $h \in \mathbb{R}_{>0}$ is called the bandwidth. In Figure 4, the estimator is illustrated. The idea is that the number of kernels is higher in regions with many samples, which leads to a higher density. The width and height of each kernel is determined by the bandwidth h. In this case, the bandwidth is the same for all kernels.

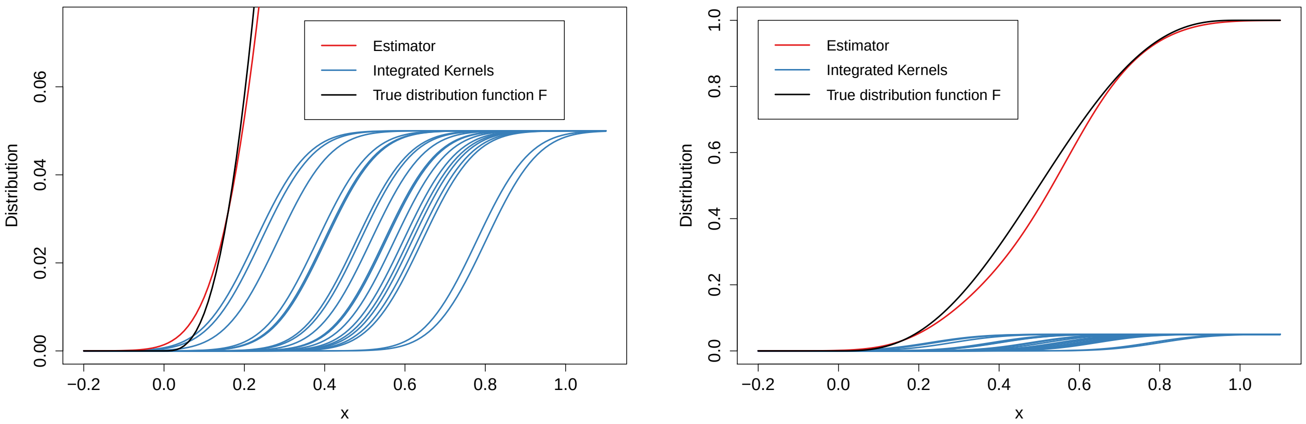

Figure 5: Illustration of the kernel distribution estimator for $n = 20$.

To estimate the distribution function, the kernel density estimator is integrated. This means that if the distribution function is of the form

$$F(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u$$

for a density function f, the kernel distribution estimator is of the form

$$F_{h,n}(x) = \int_{-\infty}^{x} f_{h,n}(u)\,\mathrm{d}u = \int_{-\infty}^{x} \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{u-X_i}{h}\right)\mathrm{d}u = \frac{1}{n}\sum_{i=1}^{n}\mathbb{K}\left(\frac{x-X_i}{h}\right), \tag{3.1}$$

where

$$\mathbb{K}(t) = \int_{-\infty}^{t} K(u)\,\mathrm{d}u$$

is a cumulative kernel function. This estimator was first introduced in [5]. In Figure 5, the integrated kernels $\frac{1}{n}\mathbb{K}\left(\frac{x-X_i}{h}\right)$ that sum up to the estimator (red) are illustrated in blue.

For both the kernel distribution estimator and the kernel density estimator, it is stated in [6, p. 592] and [7, Eq. (27), Eq. (33)] respectively that the mean and the bias are $o(h^2)$. The variance of the density estimator is $O(1/(nh))$ and of the distribution estimator $O(n^{-1}) + O(h/n) = O(n^{-1})$.

The kernel distribution estimator works well when the density is supported on $(-\infty, \infty)$. When the support is finite, problems at the boundaries can arise. In [8], it is shown that the bias at the boundary is $O(h)$, which is worse than $O(h^2)$. This is the reason why other approaches such as the Bernstein estimator are used to estimate distribution functions on bounded intervals.

We now state the most important properties of the kernel estimator. The first property is the asymptotic behavior.

3.1 Asymptotic Behavior

The next result follows from [9, Theorem 6]. It actually even holds for a more general case than kernel estimators.

Theorem 3.1. *It holds for x with $0 < F(x) < 1$ that*

$$n^{1/2}(F_{h,n}(x) - \mathbb{E}[F_{h,n}(x)]) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right)$$

for $n \to \infty$, where $\sigma^2(x)$ is defined as in Theorem 2.1. End Theorem

It is also possible to consider the asymptotic behavior of $n^{1/2}(F_{h,n}(x) - F(x))$. As shown in [8], it holds that

$$\begin{aligned} n^{1/2}(F_{h,n}(x) - F(x)) &= n^{1/2}(F_{h,n}(x) - \mathbb{E}[F_{h,n}(x)]) + n^{1/2}(\mathbb{E}[F_{h,n}(x)] - F(x)) \\ &= n^{1/2}(F_{h,n}(x) - \mathbb{E}[F_{h,n}(x)]) + n^{1/2}\left(\frac{1}{2}h^2 f'(x)\mu_2 + o(h^2)\right) \\ &= n^{1/2}(F_{h,n}(x) - \mathbb{E}[F_{h,n}(x)]) + \frac{1}{2}h^2 n^{1/2} f'(x)\mu_2 + o(h^2 n^{1/2}), \end{aligned}$$

where

$$\mu_2 = \int_{\mathcal{D}} t^2 K(t)\,\mathrm{d}t$$

and the necessary properties of K are defined as in the beginning of this section. Again, we integrate over the considered domain $\mathcal{D}$. Using this equation, the next corollary follows directly.

Corollary 3.1. *Let $n \to \infty$ and $h \to 0$. Then, for x with $0 < F(x) < 1$ and support $[-1, 1]$ of K it holds that*

(a) if $h^{-2}n^{-1/2} \to \infty$, then

$$n^{1/2}(F_{h,n}(x) - F(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(x)\right),$$

(b) if $h^{-2}n^{-1/2} \to c$, where c is a positive constant, then

$$n^{1/2}(F_{h,n}(x) - F(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\mu_2}{2c} f'(x), \sigma^2(x)\right),$$

where $\sigma^2(x)$ is defined as in Theorem 2.1. End Corollary

For the asymptotic normality of the difference to the mean, no restrictions on the bandwidth h are required, while for the behavior with respect to F there are restrictions on h.

Next, the asymptotically optimal h with respect to the MSE is calculated.

3.2 Asymptotically Optimal h with Respect to MSE

The next result follows from [8] and gives the MSE of the kernel distribution estimator.

Theorem 3.2. *The MSE of the kernel distribution estimator is of the form*

$$\mathrm{MSE}[F_{h,n}(x)] = n^{-1}\sigma^2(x) - hn^{-1}\eta f(x) + h^4 v(x) + o\left(h^4\right) + O\left(\frac{h}{n}\right),$$

where

$$\eta = 2\int_{\mathcal{D}} xK(x)\mathbb{K}(x)\,\mathrm{d}x, v(x) = \left[\frac{1}{2}f'(x)\int_{\mathcal{D}} t^2 K(t)\,\mathrm{d}t\right]^2,$$

K and $\mathbb{K}$ are defined as in Section 3. End Theorem

In order to minimize the MSE, we take the derivative with respect to h and asymptotically get

$$\frac{\partial}{\partial h}\mathrm{MSE}[F_{h,n}(x)] = 4v(x)h^3 - n^{-1}\eta f(x).$$

Setting this to zero, we obtain

$$\begin{aligned}
&4v(x)h^3 - n^{-1}\eta f(x) = 0 \\
\Leftrightarrow\ &4v(x)h^3 = n^{-1}\eta f(x) \\
\Leftrightarrow\ &h = n^{-1/3}\left[\frac{\eta f(x)}{4v(x)}\right]^{1/3}.
\end{aligned}$$

This leads to the following corollary.

Corollary 3.2. *Assuming that $f(x) \neq 0$ and $f'(x) \neq 0$, the asymptotically optimal h for estimating $F(x)$ with respect to MSE is*

$$h_{opt} = n^{-1/3}\left[\frac{\eta f(x)}{4v(x)}\right]^{1/3}. \tag{3.2}$$

This gives

$$\text{MSE}\left[\hat{F}_{h_{opt},n}(x)\right] = n^{-1}\sigma^2(x) - \frac{3}{4}n^{-4/3}\left[\frac{(\eta f(x))^4}{4v(x)}\right]^{1/3} + O(n^{-4/3}), \tag{3.3}$$

the optimal MSE. End Corollary

The same is done in the next section for the MISE instead of the MSE.

3.3 Asymptotically Optimal h with Respect to MISE

For the kernel distribution estimator $F_{h,n}$, the MISE is defined as

$$\text{MISE}[F_{h,n}] = \mathbb{E}\left[\int_{\mathcal{D}} (F_{h,n}(x) - F(x))^2 \,\mathrm{d}x\right],$$

where we integrate over the considered domain $\mathcal{D}$. The following result about the MISE of a kernel estimator can be found in [8, 10, 11].

Theorem 3.3. *It holds that*

$$\text{MISE}[F_{h,n}] = n^{-1}\int_{\mathcal{D}} \sigma^2(x)\,\mathrm{d}x - n^{-1}h\eta + \frac{1}{4}h^4\mu_2^2 r_F + o\left(h^4\right) + O\left(\frac{h}{n}\right),$$

where

$$\eta = 2\int_{\mathcal{D}} xK(x)\mathbb{K}(x)\,\mathrm{d}x, \quad r_F = \int_{\mathcal{D}} (f'(x))^2\,\mathrm{d}x, \quad \mu_2 = \int_{\mathcal{D}} t^2K(t)\,\mathrm{d}t,$$

and $\sigma^2(x)$ is defined as in Theorem 2.1. End Theorem

As before, in order to minimize the MISE, we take the derivative with respect to h and asymptotically get

$$\frac{\partial}{\partial h}\text{MISE}[F_{h,n}] = h^3\mu_2^2 r_F - n^{-1}\eta.$$

Setting this to zero leads to

$$\begin{aligned}
&h^3\mu_2^2 r_F - n^{-1}\eta = 0 \\
\Leftrightarrow\ &h^3\mu_2^2 r_F = n^{-1}\eta \\
\Leftrightarrow\ &h = n^{-1/3}\left[\frac{\eta}{r_F\mu_2^2}\right]^{1/3}.
\end{aligned}$$

With this result, the following corollary is trivial.

Corollary 3.3. *It follows that the asymptotically optimal h for estimating F with respect to MISE is*

$$h_{opt} = n^{-1/3} \left[\frac{\eta}{r_F \mu_2^2} \right]^{1/3},$$

which leads to the asymptotic expression

$$\text{MISE}\left[\hat{F}_{h_{opt},n}\right] = n^{-1} \int\limits_{\mathcal{D}} \sigma^2(x)\,\mathrm{d}x - \frac{3}{4} n^{-4/3} \left(\frac{\alpha^4}{r_F} \right)^{1/3} + O(n^{-4/3}), \tag{3.4}$$

where $\alpha = \frac{\eta}{\mu_2^{1/2}}$ *and* $\sigma^2(x)$ *is defined as in Theorem 2.1.* End Corollary

4 Function Estimation With Bernstein Polynomials

In 1912 (see [12]), Sergei Natanowitsch Bernstein introduced the Bernstein polynomial of order m of u

$$B_m(x) = B_m(u;x) = \sum_{k=0}^{m} u\left(\frac{k}{m}\right) P_{k,m}(x) \tag{4.1}$$

for a continuous function u on $[0,1]$, where $P_{k,m}$ are the Bernstein basis polynomials

$$P_{k,m}(x) = \binom{m}{k} x^k (1-x)^{m-k}$$

(see Section 12.1 for more information). The following theorem that can be found in [13, Theorem 1] shows that this function uniformly converges to u.

Theorem 4.1. *If u is a continuous function on $[0,1]$, then as $m \to \infty$,*

$$B_m(u;x) = \sum_{k=0}^{m} u\left(\frac{k}{m}\right) P_{k,m}(x) \to u(x)$$

uniformly for $x \in [0,1]$. End Theorem

At a Congress in Khar'kov in 1930 (see [14]), the three mathematicians Kantorovich, Vronskaya, and Khlodovskii introduced a new way to approximate functions $v \in L[0,1]$ with

$$K_m(x) = K_m(v,x) = (m+1) \sum_{k=0}^{m} P_{k,m}(x) \int_{\frac{k}{m+1}}^{\frac{k+1}{m+1}} v(t)\,\mathrm{d}t,$$

the so-called Kantorovich polynomials. The connection between the Kantorovich polynomials and the Bernstein polynomials is

$$B'_{m+1}(u;x) = (m+1) \sum_{k=0}^{m} \left(u\left(\frac{k+1}{m+1}\right) - u\left(\frac{k}{m+1}\right)\right) P_{k,m}(x) = K_m(v;x)$$

for functions u and v with

$$u(x) = \int_{-\infty}^{x} v(t)\,\mathrm{d}t.$$

Of course, this relation calls to mind the density and distribution functions. The aforementioned approximations seem to be a good way to estimate these functions. This was done in 1975 when Vitale was the first to introduce an estimation of density functions with Bernstein polynomials based on the Kantorovich polynomials, see [15]. For a finite random sample $X_1, ..., X_n, n \in \mathbb{N}$, with unknown density function f supported on $[0,1]$, define

$$A_{k,m}^{(n)} = \text{Number of } X_i \text{ in } \left(\frac{k}{m+1}, \frac{k+1}{m+1}\right]$$

for $k \in \{0, ..., m\}$[1].

[1] In the original work of Vitale, $A_{k,m}^{(n)}$ is defined as the number of X_i in $\left[\frac{k}{m+1}, \frac{k+1}{m+1}\right]$. This definition counts many points twice, which is not very appealing even though the probability that X_i falls into one of these points is zero because of the continuity of our distribution. In the definition given in this work, only the point $x = 0$ is not defined. This could be avoided by defining a special interval $\left[0, \frac{1}{m+1}\right]$.

Using this result, Vitale approximates f with

$$f^V_{m,n}(x) = \frac{m+1}{n} \sum_{k=0}^{m} A^{(n)}_{k,m} P_{k,m}(x).$$

Vitale replaces the unknown distribution function F by the EDF F_n to estimate the density. This can easily be shown by

$$\begin{aligned}
K_m(f;x) &= (m+1) \sum_{k=0}^{m} P_{k,m}(x) \int\limits_{\frac{k}{m+1}}^{\frac{k+1}{m+1}} f(t)\,\mathrm{d}t \\
&= (m+1) \sum_{k=0}^{m} \left(F\left(\frac{k+1}{m+1}\right) - F\left(\frac{k}{m+1}\right)\right) P_{k,m} \\
&\approx (m+1) \sum_{k=0}^{m} \left(F_n\left(\frac{k+1}{m+1}\right) - F_n\left(\frac{k}{m+1}\right)\right) P_{k,m} \\
&= \frac{m+1}{n} \sum_{k=0}^{m} \sum_{i=1}^{n} \mathbb{I}\left(\frac{k}{m+1} < X_i \leq \frac{k+1}{m+1}\right) P_{k,m} \\
&= \frac{m+1}{n} \sum_{k=0}^{m} A^{(n)}_{k,m} P_{k,m} = f^V_{m,n}(x).
\end{aligned}$$

The next section deals with the estimation of distribution functions based on the Bernstein polynomials.

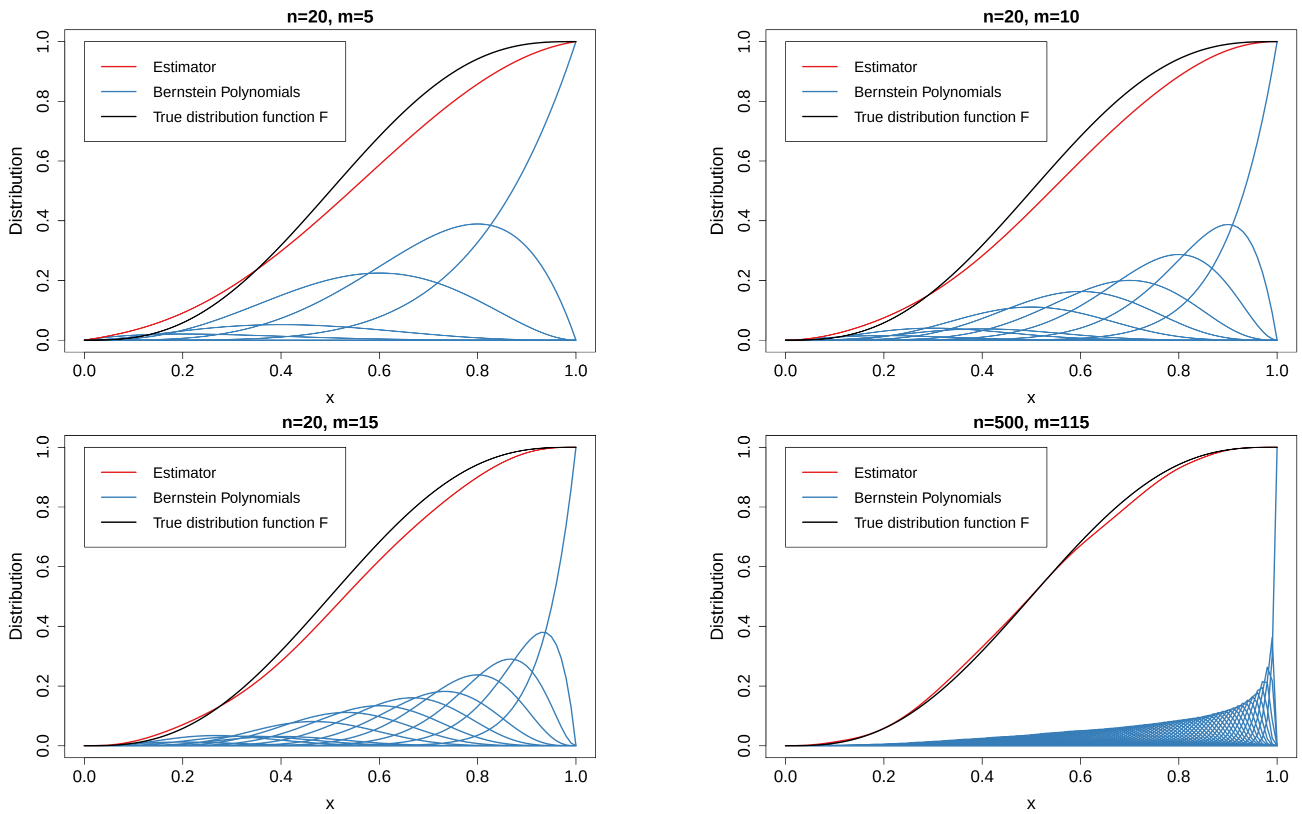

Figure 6: Illustration of the Bernstein estimator for different parameters.

5 Bernstein Distribution Function Estimation on $[0, 1]$

In this section, let $X_1, X_2, ...$ be a sequence of i.i.d. random variables with unknown distribution function F and unknown density function f supported on the unit interval $[0, 1]$.

We now talk about when it is possible to transform distributions to the unit interval without any disadvantages.

In the case where a random variable Y is supported on the compact interval $[a, b], a < b$, it can easily be restricted to $[0, 1]$ by transforming Y to $(Y - a)/(b - a)$. The back-transformation can be done without worrying about optimality or convergence rates.

If a random variable Z occupies the real line $(-\infty, \infty)$, one transformation that changes the support to $(0, 1)$ is $1/2 + (1/\pi) \tan^{-1} Z$. In the case that Z is supported on $[0, \infty)$, a possible approach is $Z/(1 + Z)$, which leads to concentration on $(0, 1)$. Although the resulting random variable is supported on $(0, 1)$ in the last two cases, it is not clear what happens to optimality conditions and convergence rates after the back-transformation.

Another argument against nonlinear transformations is the loss of interpretability. Imagine having two random variables Z_1 and Z_2 on $[0, \infty)$ and transforming them to $Z_1/(1 + Z_1)$ and $Z_2/(1 + Z_2)$. If $Z_1/(1 + Z_1)$ is statistically less than $Z_2/(1 + Z_2)$, it is not directly apparent if this also holds for Z_1 and Z_2. Hence, these transformations have to be treated with care.

The case of the real half line $[0, \infty)$ without the need to transform the random variable is considered in Section 6 and Section 7 deals with both the real half line $[0, \infty)$ and the real line $(-\infty, \infty)$.

Throughout the entire Section 5, we assume the following.

Assumption 5.1. *The distribution function F is continuous and has two continuous and bounded derivatives on* $[0, 1]$.

End Assumption

The continuity is to be understood as right and left continuity at the boundaries, as appropriate. The same holds for the differentiability.

Note that the restrictions at the boundary are only necessary to better derive the rates of convergence. For the convergence itself, the behavior at the boundary is not of importance as the estimator has zero bias and variance there. This insight was gained in [16].

The goal of this section is the estimation of a distribution function F with density f supported on $[0, 1]$, given a finite random sample $X_1, ..., X_n, n \in \mathbb{N}$. As F is continuous in our case, the use of a continuous estimator often makes much more sense than using the empirical distributon function (EDF) F_n. With Theorem 4.1 we know that F can be represented by the expression

$$B_m(F; x) = \sum_{k=0}^{m} F\left(\frac{k}{m}\right) P_{k,m}(x),$$

which converges to F uniformly for $x \in [0, 1]$. As the distribution function F is unknown, the idea now is to replace F with the EDF F_n. Following [17], this leads to the distribution function estimator

$$\hat{F}_{m,n}(x) = \sum_{k=0}^{m} F_n\left(\frac{k}{m}\right) P_{k,m}(x),$$

where $P_{k,m} = \binom{m}{k} x^k (1-x)^{m-k}$ are the Bernstein basis polynomials. We always assume that $m = m_n$ depends on n.

Figure 6 shows how the weighted Bernstein basis polynomials (blue) add up to estimate the distribution function.

We now state and prove some important properties of the Bernstein distribution estimator.

5.1 General Properties

One big advantage of the Bernstein distribution estimator is that it yields very smooth estimates with good behavior at the boundaries. Indeed, the estimator is unbiased with zero variance at the boundary because

$$\hat{F}_{m,n}(0) = 0 = F(0) = B_m(F; 0) \text{ and } \hat{F}_{m,n}(1) = 1 = F(1) = B_m(F; 1) \tag{5.1}$$

with probability one.

The expected value can easily be calculated as

$$\mathbb{E}\left[\hat{F}_{m,n}(x)\right] = B_m(F; x) = \sum_{k=0}^{m} F\left(\frac{k}{m}\right) P_{k,m}(x) \tag{5.2}$$

for $x \in [0, 1], n \geq 1$, which is the Bernstein polynomial of order m of F introduced in Eq. (4.1).

Following [13], we now show that $\hat{F}_{m,n}(x)$ yields a proper continuous distribution function almost surely for all values of m. It is easy to see that $\hat{F}_{m,n}(x)$ is continuous. Furthermore, it holds that $0 \leq \hat{F}_{m,n}(x) \leq 1$ for $x \in [0, 1]$ because of Eq. (5.1) and the fact that $\hat{F}_{m,n}(x)$ is increasing, which is shown now. The following statement can for example be found in [13].

Theorem 5.1. *The function $\hat{F}_{m,n}(x)$ is increasing in x on $[0, 1]$.* End Theorem

Proof. It holds that

$$\hat{F}_{m,n}(x) = \sum_{k=0}^{m} g_n\left(\frac{k}{m}\right) U_k(m, x), \tag{5.3}$$

where

$$g_n(0) = 0 \text{ and } g_n\left(\frac{k}{m}\right) = F_n\left(\frac{k}{m}\right) - F_n\left(\frac{k-1}{m}\right), k = 1, ..., m,$$

and

$$U_k(m,x) = \sum_{j=k}^{m} P_{j,m}(x) = m\binom{m-1}{k-1}\int_0^x t^{k-1}(1-t)^{m-k}\,\mathrm{d}t.$$

The last equation follows from the fact that

$$\begin{aligned}
\sum_{j=k}^{m} P_{j,m}(x) &= \sum_{j=k}^{m}\binom{m}{j}x^j(1-x)^{m-j} \\
&= 1-\sum_{j=0}^{k-1}\binom{m}{j}x^j(1-x)^{m-j} \\
&= 1-m\binom{m-1}{k-1}\int_x^1 t^{k-1}(1-t)^{m-k}\,\mathrm{d}t \\
&= m\binom{m-1}{k-1}\left[\int_0^1 t^{k-1}(1-t)^{m-k}\,\mathrm{d}t-\int_x^1 t^{k-1}(1-t)^{m-k}\,\mathrm{d}t\right],
\end{aligned}$$

where the connection between the binomial distribution and the beta function was used.

Eq. (5.3) holds because of

$$\begin{aligned}
\sum_{k=0}^{m} g_n\left(\frac{k}{m}\right)U_k(m,x) &= \sum_{k=1}^{m}\left[F_n\left(\frac{k}{m}\right)-F_n\left(\frac{k-1}{m}\right)\right]\sum_{j=k}^{m}P_{j,m}(x) \\
&= \sum_{k=1}^{m}\sum_{j=k}^{m}F_n\left(\frac{k}{m}\right)P_{j,m}(x)-\sum_{k=0}^{m-1}\sum_{j=k}^{m}F_n\left(\frac{k}{m}\right)P_{j,m}(x)+\sum_{k=0}^{m-1}F_n\left(\frac{k}{m}\right)P_{k,m}(x) \\
&= F_n\left(\frac{m}{m}\right)P_{k,m}(x)+\sum_{k=0}^{m-1}F_n\left(\frac{k}{m}\right)P_{k,m}(x) \\
&= \hat{F}_{m,n}(x).
\end{aligned}$$

Now, $\hat{F}_{m,n}(x)$ is increasing since $g_n\left(\frac{k}{m}\right)$ is non-negative for at least one k and $U_k(m,x)$ is increasing in x. End Proof

As mentioned in [13], another positive property of the Bernstein estimator is that it takes the knowledge of the support of the distribution into account, which is $[0,1]$ in this section. Other estimators do not have this property and assign positive probability to a region that is actually known to have zero probability. It is also worth noticing that $\hat{F}_{m,n}(x)$ is a polynomial in x and therefore, all derivatives exist.

The next theorem follows from [13, Theorem 2.1] and shows that $\hat{F}_{m,n}(x)$ is uniformly strongly consistent.

Theorem 5.2. *If F is a continuous probability distribution function on $[0,1]$, then*

$$\left\|\hat{F}_{m,n}-F\right\| \to 0$$

almost surely for $m,n\to\infty$. We use the notation $\|G\| = \sup_{x\in[0,1]}|G(x)|$ for a bounded function G on $[0,1]$. End Theorem

Proof. It holds that

$$\left\|\hat{F}_{m,n}-F\right\| \le \left\|\hat{F}_{m,n}-B_m\right\| + \|B_m-F\|$$

and

$$\begin{aligned}\left\|\hat{F}_{m,n}-B_m\right\| &= \left\|\sum_{k=0}^{m}\left[F_n(k/m)-F(k/m)\right]P_{k,m}\right\| \\ &\leq \max_{0\leq k\leq m}\left|F_n(k/m)-F(k/m)\right| \leq \|F_n-F\|.\end{aligned}$$

We know with the Glivenko-Cantelli theorem that $\|F_n-F\|\to 0$ a.s. for $n\to\infty$. The claim follows from Theorem 4.1. End Proof

In [18], it is shown that under certain conditions, the Chung-Smirnov property

$$\limsup_{n\to\infty}\left(\frac{2n}{\log\log n}\right)^{1/2}\sup_{x\in[0,1]}\left|\hat{F}_{m,n}(x)-F(x)\right|\leq 1,\ \text{ a.s.,} \tag{5.4}$$

holds. This result is used in Section 5.7.

In the following, we turn our attention to the bias and the variance of the estimator.

5.1.1 Bias and Variance

We already know that the estimator is unbiased with zero variance at the boundaries. We now take a closer look at the inner interval $(0,1)$. The following result can be found in [19, Section 1.6.1] and leads to the bias of $\hat{F}_{m,n}$.

Lemma 5.1. *It holds for $x\in(0,1)$ that*

$$B_m(F;x)=B_m(x)=F(x)+m^{-1}b(x)+o(m^{-1}),$$

where $b(x)=\frac{x(1-x)f'(x)}{2}$. For the uniform case, this equation simplifies to $B_m(x)=F(x)=x$ for all $m\geq 1, x\in[0,1]$. End Lemma

Proof. Following the proof in [19, Section 1.6.1], it holds that

$$F\left(\frac{k}{m}\right)=F(x)+\left(\frac{k}{m}-x\right)f(x)+\frac{1}{2}\left(\frac{k}{m}-x\right)^2 f'(x)+o\left(\left(\frac{k}{m}-x\right)^2\right).$$

Using this result, it follows that

$$\begin{aligned}B_m(x) &= \sum_{k=0}^{m}F\left(\frac{k}{m}\right)P_{k,m}(x) \\ &= F(x)+\underbrace{f(x)\sum_{k=0}^{m}\left(\frac{k}{m}-x\right)P_{k,m}(x)}_{S_2}+\underbrace{\frac{1}{2}f'(x)\sum_{k=0}^{m}\left(\frac{k}{m}-x\right)^2P_{k,m}(x)}_{S_3} \\ &+\underbrace{\sum_{k=0}^{m}o\left(\left(\frac{k}{m}-x\right)^2\right)P_{k,m}(x)}_{S_4}.\end{aligned}$$

The second summand S_2 simplifies to zero as

$$\sum_{k=0}^{m}\left(\frac{k}{m}-x\right)\binom{m}{k}x^k(1-x)^{m-k}=\frac{1}{m}\mathbb{E}[Y]-x=x-x=0 \tag{5.5}$$

for $x \in [0,1]$, where $Y \sim \text{Bin}(m,x)$.

The third part S_3 can be calculated by using the variance

$$\sum_{k=0}^{m} \left(\frac{k}{m} - x\right)^2 P_{k,m}(x) = \frac{1}{m^2} \sum_{k=0}^{m} (k - mx)^2 P_{k,m}(x) = \frac{1}{m^2} \text{Var}[Y] = \frac{x(1-x)}{m}, \tag{5.6}$$

where Y is again a random variable with distribution $\text{Bin}(m,x)$.

Because of Eq. (5.6), for the last summand S_4, we obtain

$$\begin{aligned} S_4 &= \sum_{k=0}^{m} o\left(\left(\frac{k}{m} - x\right)^2\right) P_{k,m}(x) \\ &= o\left(\sum_{k=0}^{m} \left(\frac{k}{m} - x\right)^2 P_{k,m}(x)\right) \\ &= o\left(\frac{x(1-x)}{m}\right) = o(m^{-1}). \end{aligned}$$

For the uniform case, it is of course well known that

$$f(x) = \begin{cases} 1, & x \in [0,1], \\ 0, & \text{else}, \end{cases} \quad \text{and} \quad F(x) = \begin{cases} 0, & x < 0, \\ x, & x \leq 0 < 1, \\ 1, & x \geq 1, \end{cases}$$

so that

$$b(x) = \frac{x(1-x)f'(x)}{2} = 0 \text{ on } (0,1). \tag{5.7}$$

From Eq. (5.5) we know that

$$\begin{aligned} B_m(x) &= \sum_{k=0}^{m} F\left(\frac{k}{m}\right) P_{k,m}(x) \\ &= \sum_{k=0}^{m} \frac{k}{m} P_{k,m}(x) = x = F(x), \end{aligned}$$

which proves the claim. End Proof

We now turn our attention to the asymptotic expressions for bias and variance of the estimator $\hat{F}_{m,n}$ as $m,n \to \infty$. The theorem can be found in [17, Theorem 1].

Theorem 5.3. *For $x \in (0,1)$, we have that*

$$\text{Bias}\left[\hat{F}_{m,n}(x)\right] = \mathbb{E}\left[\hat{F}_{m,n}\right] - F(x) = m^{-1}b(x) + o(m^{-1}),$$

where $b(x)$ is defined as in Lemma 5.1. For the variance it holds that

$$\text{Var}\left[\hat{F}_{m,n}(x)\right] = n^{-1}\sigma^2(x) - m^{-1/2}n^{-1}V(x) + o_x(m^{-1/2}n^{-1}),$$

where

$$V(x) = f(x)\left[\frac{2x(1-x)}{\pi}\right]^{1/2}$$

and $\sigma^2(x)$ is defined as in Theorem 2.1. End Theorem

For the proof, see Proofs Bernstein.

Note that the bias of the Bernstein estimator $\hat{F}_{m,n}$ is $O(m^{-1}) = O(h)$ if setting the "bandwidth" as $h = 1/m$, which seems a natural step. This order is worse than that of the bias of a kernel distribution estimator that typically is $O(h^2)$ (see Section 3). The variance here is $O(n^{-1})$, which is the same as the variance of the kernel distribution estimator (see Section 3).

Remembering the properties of the kernel distribution estimator, we know that the order of the bias is worse at the boundary. For the Bernstein estimator, the opposite holds true if f'' is also assumed to be bounded and continuous on $[0, 1]$. In [20, p. 2770], it is shown that the bias in the boundary area is of order m^{-2}. Under the extra assumption that $f'(0) = f'(1) = 0$ (the so-called shoulder condition), the order even decreases to m^{-3}. In [20, p. 2770, p. 2771], it can be seen that the variance reduces from order n^{-1} to order $m^{-1}n^{-1}$ at the boundary.

In the following, we turn our attention to the asymptotic behavior of the Bernstein estimator.

5.2 Asymptotic Behavior

We now show that the asymptotic distribution of the difference between the Bernstein estimator and its expected value is the same as for the EDF and the kernel estimator. The next theorem can be found in [17, Theorem 2].

Theorem 5.4. *Let $x \in (0, 1)$, such that $0 < F(x) < 1$. Then, for $m, n \to \infty$, it holds that*

$$n^{1/2}\left(\hat{F}_{m,n}(x) - \mathbb{E}[\hat{F}_{m,n}(x)]\right) = n^{1/2}\left(\hat{F}_{m,n}(x) - B_m(F; x)\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right),$$

where $\sigma^2(x)$ is defined as in Theorem 2.1. End Theorem

The idea for the proof is to use the central limit theorem for double arrays, see Proofs Bernstein for more details.

Note that the asymptotic result holds for all m with $m \to \infty$ in contrast to the density estimation. In the density setting, there are restrictions on the choice of m as can be seen in [13, Proposition 1].

The next corollary deals with the asymptotic behavior of $\hat{F}_{m,n}(x) - F(x)$. This is of more interest than the one of $\hat{F}_{m,n}(x) - \mathbb{E}[\hat{F}_{m,n}(x)]$ as we want to know more about the behavior of the estimator with respect to the true function. With Lemma 5.1, it is easy to see that

$$n^{1/2}\left(\hat{F}_{m,n}(x) - F(x)\right) = n^{1/2}\left(\hat{F}_{m,n}(x) - B_m(F; x)\right) + m^{-1}n^{1/2}b(x) + o(m^{-1}n^{1/2}). \tag{5.8}$$

This directly leads to the following corollary from [17, Corollary 2].

Corollary 5.1. *Let $m, n \to \infty$. Then, for $x \in (0, 1)$ with $0 < F(x) < 1$, it holds that*

(a) if $mn^{-1/2} \to \infty$, then

$$n^{1/2}\left(\hat{F}_{m,n}(x) - F(x)\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right),$$

(b) if $mn^{-1/2} \to c$, where c is a positive constant, then

$$n^{1/2}\left(\hat{F}_{m,n}(x) - F(x)\right) \xrightarrow{D} \mathcal{N}\left(c^{-1}b(x), \sigma^2(x)\right),$$

where $\sigma^2(x)$ and $b(x)$ are defined as in Theorem 5.3. End Corollary

Part (b) can be seen as a border case. Note that for these properties to hold, there are restrictions on the asymptotic behavior of m, as in the kernel setting.

We now calculate the asymptotically optimal m with respect to MSE.

5.3 Asymptotically Optimal m with Respect to MSE

Here, we take a closer look at the mean squared error (MSE), defined by

$$\text{MSE}\left[\hat{F}_{m,n}(x)\right] = \mathbb{E}\left[\left(\hat{F}_{m,n}(x) - F(x)\right)^2\right].$$

In the current setting, for $x \in \{0, 1\}$, the Bernstein estimator and the EDF both have an MSE of zero.

We show that the optimal MSE of the Bernstein estimator and the kernel estimator are of a very similar form. They both asymptotically outperform the EDF.

The next theorem can be found in [17] and follows directly from Theorem 5.3.

Theorem 5.5. *It holds that*

$$\begin{aligned}\text{MSE}\left[\hat{F}_{m,n}(x)\right] &= \text{Var}\left[\hat{F}_{m,n}(x)\right] + \text{Bias}\left[\hat{F}_{m,n}(x)\right]^2 \\ &= n^{-1}\sigma^2(x) - m^{-1/2}n^{-1}V(x) + m^{-2}b^2(x) + o(m^{-2}) + o_x(m^{-1/2}n^{-1})\end{aligned} \tag{5.9}$$

for $x \in (0, 1)$. End Theorem

Note that we use the notation o_x here, which means that the bound is pointwise. For a uniform bound, the notation o is used.

Taking the derivative of Eq. (5.9) with respect to m, we asymptotically obtain

$$\begin{aligned}\frac{\partial}{\partial m}\text{MSE}\left[\hat{F}_{m,n}(x)\right] &= \frac{1}{2}m^{-3/2}n^{-1}V(x) - 2m^{-3}b^2(x) = 0 \\ &\Leftrightarrow m^{-3/2}n^{-1}V(x)) = 4m^{-3}b^2(x) \\ &\Leftrightarrow m^{3/2} = n\frac{4b^2(x)}{V(x)} \\ &\Leftrightarrow m = n^{2/3}\left[\frac{4b^2(x)}{V(x)}\right]^{2/3}.\end{aligned} \tag{5.10}$$

This leads to the following corollary, which can also be found in [17, Corollary 1].

Corollary 5.2. *Assuming that $f(x) \neq 0$ and $f'(x) \neq 0$, the asymptotically optimal choice of m for estimating $F(x)$ with respect to MSE is*

$$m_{opt} = n^{2/3}\left[\frac{4b^2(x)}{V(x)}\right]^{2/3}. \tag{5.11}$$

Using this result, we get for $x \in (0, 1)$ that

$$\text{MSE}\left[\hat{F}_{m_{opt},n}(x)\right] = n^{-1}\sigma^2(x) - \underbrace{\frac{3}{4}n^{-4/3}\left[\frac{V^4(x)}{4b^2(x)}\right]^{1/3}}_{S_2} + o_x(n^{-4/3}), \tag{5.12}$$

where $\sigma^2(x), b(x)$, and $V(x)$ are defined as in Theorem 5.3. End Corollary

By comparing this expression with Eq. (2.1), one can see that the Bernstein estimator asymptotically outperforms the EDF with respect to MSE as the second summand S_2 in Eq. (5.12) is always positive. More details can be found in Section 5.6. The same is true for the kernel estimator, see Eq. (3.3).

We now take a look at the optimal m with respect to MISE.

5.4 Asymptotically Optimal m with Respect to MISE

Similar to the section before we now focus on the mean integrated squared error (MISE), which is defined as

$$\text{MISE}\left[\hat{F}_{m,n}\right] = \mathbb{E}\left[\int_0^1 \left(\hat{F}_{m,n} - F(x)\right)^2 \mathrm{d}x\right]$$

for the Bernstein estimator.

Note that the MISE of the Bernstein estimator cannot be calculated by integrating the expression of MSE $\left[\hat{F}_{m,n}\right]$ given in Eq. (5.9) as the asymptotic expression $o_x(m^{-1/2}n^{-1})$ for the variance of the Bernstein estimator depends on x.

We show that similar to the section before, the Bernstein and the kernel estimator asymptotically outperform the EDF with respect to MISE.

The next theorem can be found in [17, Theorem 3] and gives the MISE of the Bernstein estimator.

Theorem 5.6. *It holds that*

$$\text{MISE}\left[\hat{F}_{m,n}\right] = n^{-1}C_1 - m^{-1/2}n^{-1}C_2 + m^{-2}C_3 + o(m^{-1/2}n^{-1}) + o(m^{-2})$$

with

$$C_1 = \int_0^1 \sigma^2(x)\,\mathrm{d}x \;\;,\;\; C_2 = \int_0^1 V(x)\,\mathrm{d}x \;\; \textit{and} \;\; C_3 = \int_0^1 b^2(x)\,\mathrm{d}x.$$

The definitions of $\sigma^2(x), b(x)$, and $V(x)$ can be found in Theorem 5.3. End Theorem

For the proof, see Proofs Bernstein.

The three integrals C_1, C_2, and C_3 are positive, except in the special case of the uniform distribution, where $C_3 = 0$. This follows from Eq. (5.7).

As in the section for the MSE, it is also possible to compute the optimal m with respect to MISE. The next corollary follows from [17, Corollary 4] and gives this optimal m.

Corollary 5.3. *In the case $C_3 > 0$, it follows that the asymptotically optimal m for estimating F with respect to MISE is*

$$m_{opt} = n^{2/3}\left[\frac{4C_3}{C_2}\right]^{2/3},$$

which leads to

$$\text{MISE}\left[\hat{F}_{m_{opt},n}\right] = n^{-1}C_1 - \underbrace{\frac{3}{4}n^{-4/3}\left[\frac{C_2^4}{4C_3}\right]^{1/3}}_{S_2} + o(n^{-4/3}), \tag{5.13}$$

the optimal MISE. End Corollary

The calculation of the optimal m with respect to MISE is very similar to Eq. (5.10) so that we omit the proof here.

Similar to the MSE case one can compare the Bernstein estimator with the EDF. The fact that the second summand S_2 in Eq. (5.13) is always positive leads to the conclusion that the Bernstein estimator asymptotically outperforms the EDF. For more information on this, see Section 5.6. The same is again true for the kernel estimator, see Eq. (3.4).

We now talk about a property that holds for both the Bernstein estimator and the kernel estimator.

5.5 Difference of the Distribution Estimator to the Density Estimator

The phenomenon that we will talk about here was first observed by Hjort and Walker in [21] for the kernel estimator but can also be applied to the Bernstein estimator.

We first describe this phenomenon in the case of the kernel estimator.

Kernel Distribution Estimator

The recommended bandwidth for estimating the density with a kernel estimator is $O(n^{-1/5})$, see [17]. When using this bandwidth for the estimation of the distribution function, the following problem was observed by Hjort and Walker: the estimated function does not lie in any reasonable confidence band of F based on F_n. This problem is solved by using the optimal bandwidth $o(n^{-1/4})$ for estimating the distribution function, see [17].

Bernstein Estimator

The same phenomenon as for the kernel distribution estimator is observed in [17] for Bernstein estimators. In density estimation, the optimal number m of Bernstein polynomials with respect to MISE is $O(n^{2/5})$, see [22]. For the estimation of the distribution, the optimal choice is $o(n^{2/3})$ (see Corollary 5.3) and when using $O(n^{2/5})$, the estimation lies outside of any confidence band of F. This holds because of the fact that from $mn^{-2/5} \to c$ it follows that $mn^{-1/2} \to 0$ and together with $f'(x) \neq 0$ and Eq. (5.8), it holds that

$$\mathbb{P}\left(n^{1/2}\left|\hat{F}_{m,n}(x) - F(x)\right| > \epsilon\right) \to 1$$

for all $\epsilon > 0$. This shows that for m chosen as $O(n^{2/5})$, $\hat{F}_{m,n}(x)$ does not converge to a limiting distribution centred at $F(x)$ with proper rescaling. It follows that $\hat{F}_{m,n}$ lies outside of confidence bands based on F_n with probability going to one.

5.6 Deficiency

It was already made clear in the Sections 5.3 and 5.4 that it is interesting to compare the performances of the Bernstein estimator and the EDF. Here, the comparison is done with the help of the so-called deficiency.

In [23], the defiency is defined as follows. Consider a statistical procedure A, which is based on n observations and another procedure B that is less effective than the first one. This means that the procedure B needs a larger number $i(n)$ of observations to perform at least as well as procedure A. One way to compare the two procedures is to take a look at the ratio $\frac{i(n)}{n}$. Another way is to inspect the number of required additional observations $i(n) - n$, which is called deficiency. Taking the limit leads to the asymptotic deficiency.

In our case we measure the performance using the MSE and the MISE. The procedure A is of course the Bernstein estimator $\hat{F}_{m,n}$ while the EDF F_n is the procedure B. We define the local and the global number of observations that F_n needs to perform at least as well as $\hat{F}_{m,n}$ as

$$\begin{aligned} i_L(n, x) &= \min\left\{k \in \mathbb{N} : \mathrm{MSE}[F_k(x)] \leq \mathrm{MSE}\left[\hat{F}_{m,n}(x)\right]\right\}, \text{ and} \\ i_G(n) &= \min\left\{k \in \mathbb{N} : \mathrm{MISE}[F_k] \leq \mathrm{MISE}\left[\hat{F}_{m,n}\right]\right\}. \end{aligned}$$

Following from [17, Theorem 4], the next result gives conditions under which F_n is asymptotically efficient. It also gives the asymptotic defiency in two cases.

Theorem 5.7. *Let $x \in (0,1)$ and $m, n \to \infty$. If $mn^{-1/2} \to \infty$, it holds that*

$$i_L(n,x) = n[1 + o_x(1)] \text{ and } i_G(n) = n[1 + o(1)].$$

In addition, the following is true.

(a) If $mn^{-2/3} \to \infty$ and $mn^{-2} \to 0$, then

$$i_L(n,x) - n = m^{-1/2}n[\theta(x) + o_x(1)], \text{ and}$$
$$i_G(n) - n = m^{-1/2}n[C_2/C_1 + o(1)].$$

(b) If $mn^{-2/3} \to c$, where c is a positive constant, then

$$i_L(n,x) - n = n^{2/3}[c^{-1/2}\theta(x) - c^{-2}\gamma(x) + o_x(1)], \text{ and}$$
$$i_G(n) - n = n^{2/3}[c^{-1/2}C_2/C_1 - c^{-2}C_3/C_1 + o(1)],$$

where

$$\theta(x) = \frac{V(x)}{\sigma^2(x)} \text{ and } \gamma(x) = \frac{b^2(x)}{\sigma^2(x)}.$$

$V(x), \sigma^2(x)$ and $b(x)$ are defined as in Theorem 5.3 and C_1, C_2 and C_3 are defined as in Theorem 5.6.

In the case where $x \in \{0,1\}$, it is easy to see that

$$i_L(n,0) = i_L(n,1) = n$$

for any choice of $m > 0$. End Theorem

For the proof, see Proofs Bernstein.

The theorem shows in which cases the Bernstein estimator outperforms the EDF. In the setups (a) and (b), the asymptotic deficiency goes to infinity as n grows. This means that for increasing n, more and more extra observations are needed for the EDF to outperform the Bernstein estimator.

As mentioned before, it can be shown that the EDF is also outperformed by the kernel estimator with respect to MSE and MISE, see [24].

It seems natural that one can also base the selection of an optimal m on the deficiency. Indeed, maximizing the deficiency seems a good way to make sure that the Bernstein estimator outperforms the EDF as much as possible.

With arguments from [17] the following statement holds.

Lemma 5.2. *The optimal m with respect to the global deficiency in the case $mn^{-2/3} \to c$ is of the same order as in Corollary 5.3.* End Lemma

Proof. Note that when $mn^{-2/3} \to c$, $i_G(n) - n$ is asymptotically positive only when

$$c > \left[\frac{C_3}{C_2}\right]^{2/3} = c^*.$$

Now the goal is to choose m so that $mn^{-2/3} \to c$ and $c > c^*$ is chosen to maximize

$$g(c) = c^{-1/2}C_2/C_1 - c^{-2}C_3/C_1.$$

This is achieved by taking the derivative of g with respect to c and setting it to zero

$$\begin{aligned} g'(c) &= 2c^{-3}\frac{C_3}{C_1} - \frac{1}{2}c^{-3/2}\frac{C_2}{C_1} = 0 \\ \Leftrightarrow 2c^{-3}\frac{C_3}{C_1} &= \frac{1}{2}c^{-3/2}\frac{C_2}{C_1} \\ \Leftrightarrow c^{-3/2} &= \frac{1}{4}\frac{C_2}{C_3}. \end{aligned}$$

This leads to

$$c_{opt} = \left[\frac{4C_3}{C_2}\right]^{2/3} = 2^{4/3}c^*.$$

Hence, the optimal order of the Bernstein estimator with respect to the deficiency satisfies

$$m_{opt}n^{-2/3} \to c_{opt} \Leftrightarrow m_{opt} = n^{2/3}[c_{opt} + o(1)],$$

which leads to the assertion. End Proof

The next section deals with a way to choose the optimal m that is different to the approaches before.

5.7 Another Way to Choose the Optimal m

In Sections 5.3 and 5.4, the asymptotically optimal m is obtained by minimizing the MSE and the MISE. The resulting m_{opt} depends on the unknown density. In the sequel, another approach to finding an optimal m is discussed, following from [25]. We will distinguish between a local and a global optimum.

As already mentioned before, $1/m$ plays the same role in Bernstein estimation as the bandwidth does for kernel estimation. This means that a data-driven choice of m seems to be of interest. This approach is used by Dutta in [25]. Such an m depends on the random variables $X_1, ..., X_n, n \in \mathbb{N}$, and hence, is random.

In the entire Section 5.7, we assume that for the kernel density estimator the following assumption holds.

Assumption 5.2. *K is a density, which is differentiable in the interior of the support, where K' satisfies $\int K'(x)\,\mathrm{d}x = 0$ and $\int xK'(x)\,\mathrm{d}x = -1$.* End Assumption

This assumption is satisfied by many well-known kernels such as the Gaussian and the Epanechnikov kernel that were defined in Section 3.

We now calculate the locally optimal m.

5.7.1 Local Choice of m

The aim is to choose m such that $F(x)$ is optimally estimated. In Corollary 5.2, the asymptotically optimal choice of m for estimating $F(x)$ with respect to MSE was calculated. We now do something similar but in the end, m will not depend on any unknown parameters.

With Eq. (5.2), it holds that

$$\text{Bias}\left[\hat{F}_{m,n}(x)\right] = B_m(x) - F(x) \tag{5.14}$$

and in Eq. (5.25), it was calculated that

$$\text{Var}\left[\hat{F}_{m,n}(x)\right] = \frac{1}{n}\left[\sum_{k=0}^{m} F\left(\frac{k}{m}\right)P_{k,m}^2(x) + 2\sum\sum_{0\le k<l\le m} F\left(\frac{k}{m}\right)P_{k,m}(x)P_{l,m}(x) - B_m^2(x)\right].$$

Using these two equations, the MSE can be calculated by

$$\mathrm{MSE}\left[\hat{F}_{m,n}(x)\right] = \mathrm{Var}\left[\hat{F}_{m,n}(x)\right] + \mathrm{Bias}\left[\hat{F}_{m,n}(x)\right]^2.$$

The proposal in [25] is now to replace the unknown distribution function F by the kernel estimator $F_{h,n}$ in the bias and by the EDF F_n in the variance. This gives

$$\begin{aligned} \mathrm{Bias}^*\left[\hat{F}_{m,n}(x)\right] &= B_{h,m}(x) - F_{h,n}(x), \qquad (5.15)\\ \mathrm{Var}^*\left[\hat{F}_{m,n}(x)\right] &= \frac{1}{n}\left[\sum_{k=0}^{m} F_n\left(\frac{k}{m}\right) P_{k,m}^2(x) + 2\sum\sum_{0\le k<l\le m} F_n\left(\frac{k}{m}\right) P_{k,m}(x)P_{l,m}(x) - \hat{F}_{m,n}^2(x)\right],\\ \mathrm{MSE}^*\left[\hat{F}_{m,n}(x)\right] &= \mathrm{Var}^*\left[\hat{F}_{m,n}(x)\right] + \mathrm{Bias}^*\left[\hat{F}_{m,n}(x)\right]^2, \end{aligned}$$

where

$$B_{h,m}(x) = \sum_{k=0}^{m} F_{h,n}\left(\frac{k}{m}\right) P_{k,m}(x).$$

The question that arises now is, why the distribution function is replaced by two different terms in bias and variance. Following [25], the goal is to obtain

$$\mathrm{MSE}^*\left[\hat{F}_{m,n}(x)\right] = \mathrm{MSE}\left[\hat{F}_{m,n}(x)\right] + o(r_n) \text{ a.s.},$$

where $r_n \to 0$ for $n \to \infty$ faster than $\mathrm{MSE}\left[\hat{F}_{m,n}(x)\right]$. In Theorem 5.3, it was shown that the bias converges to zero in $o(m^{-1})$. Replacing F by F_n in the bias would not lead to the same result as F_n is discrete. Hence, F is replaced by the continuous kernel estimator.

For the variance, simulations in [25] show that it is enough to use the EDF to replace F.

As can be seen later in the proof of Theorem 5.8, the goal in choosing the bandwidth h for the kernel estimator is not that it optimally estimates F. Instead, the goal is to obtain the relation

$$\mathrm{Bias}^*\left[\hat{F}_{m,n}(x)\right]^2 - \mathrm{Bias}\left[\hat{F}_{m,n}(x)\right]^2 = o\left(\frac{1}{m^2}\right).$$

In [25], it is stated that $h = n^{-1/9}$ fulfills this equation. Let h satisfy this from now on in Section 5.7.

In Corollary 5.2, it is shown that

$$n\,\mathrm{MSE}\left[\hat{F}_{m_{opt},n}(x)\right] \to \sigma^2(x) \qquad (5.16)$$

if m is a multiple of $n^{2/3}$.

In the next theorem, which follows [25, Theorem 1], the convergence of $\mathrm{MSE}^*\left[\hat{F}_{m,n}(x)\right] - \mathrm{MSE}\left[\hat{F}_{m,n}(x)\right]$ is examined. The result can be expected from Eq. (5.16).

Theorem 5.8. *Let F fulfill Assumption 5.1 and additionally be absolutely continuous. Then*

$$\mathrm{MSE}^*\left[\hat{F}_{m,n}(x)\right] - \mathrm{MSE}\left[\hat{F}_{m,n}(x)\right] = o\left(\frac{1}{n}\right) \quad a.s.$$

for $c_1 n^{2/3} \le m \le c_2 n^{2/3}$, where c_1 and c_2 are two constants. End Theorem

For the proof, see Proofs Bernstein.

Now, the proposal given in [25] is to choose m so that it minimizes $\mathrm{MSE}^*\left[\hat{F}_{m,n}(x)\right]$ under the condition that $m \in I_n = [c_1 n^{2/3}, c_2 n^{2/3}]$, where $0 < c_1 < c_2$ are constants. We call the resulting number $\widehat{m}$, which is defined by

$$\widehat{m} = \left\lceil \mathrm{argmin}_{k\in I_n} \mathrm{MSE}^*\left[\hat{F}_{m,n}(x)\right]\right\rceil. \qquad (5.17)$$

A simulation in [25] suggests that $\left[\frac{4b^2(x)}{V(x)}\right]^{2/3}$, the coefficient in m_{opt}, varies between 0 and 1500 so that $I_n = [\frac{1}{10}n^{2/3}, 1500n^{2/3}]$.

5.7.2 Global Choice of m

Now, the goal is to find the global optimum of m. As in Section 5.4, this is done with the MISE. As the support of F is on $[0,1]$, the MISE is defined by

$$\text{MISE}\left[\hat{F}_{m,n}\right] = \int_0^1 \text{MSE}\left[\hat{F}_{m,n}(x)\right] \mathrm{d}x$$

and following [25], an estimator of the MISE is

$$\text{MISE}^*\left[\hat{F}_{m,n}\right] = \int_0^1 \text{MSE}^*\left[\hat{F}_{m,n}(x)\right] \mathrm{d}x.$$

Using $n\,\text{MSE}\left[\hat{F}_{m_{opt},n}(x)\right] \to \sigma^2(x)$ from Corollary 5.2 again together with Fatou's lemma, it holds that

$$\liminf_{n\to\infty} n\,\text{MISE}\left[\hat{F}_{m_{opt},n}\right] \geq \int_0^1 \liminf_{n\to\infty} n\,\text{MSE}\left[\hat{F}_{m_{opt},n}(x)\right] \mathrm{d}x > 0.$$

Similar to the theorem above, the next theorem, which can be found in [25, Theorem 2] deals with the convergence rate of $\text{MISE}^*\left[\hat{F}_{m,n}\right] - \text{MISE}\left[\hat{F}_{m,n}\right]$.

Theorem 5.9. *Under the same conditions as in Theorem 5.8, it holds that*

$$\text{MISE}^*\left[\hat{F}_{m,n}\right] - \text{MISE}\left[\hat{F}_{m,n}\right] = o\left(\frac{1}{n}\right)$$

for $n \to \infty$. End Theorem

For the proof, see Proofs Bernstein.

The proposition in [25] is now very similar to the one for the local choice of m. The parameter is chosen to minimize the MISE and is called $\widehat{\widehat{m}}$, i.e.

$$\widehat{\widehat{m}} = \left\lceil \text{argmin}_{k\in I_n} \text{MISE}^*\left[\hat{F}_{m,n}\right]\right\rceil, \tag{5.18}$$

where I_n is the interval as defined above.

5.7.3 Asymptotic Properties

Here, the asymptotic properties of the optimal parameters proposed above are stated. More precisely, the properties are stated for the degree $\tilde{m}$ of the Bernstein estimator chosen by a data-based algorithm. The parameters $\widehat{m}$ and $\widehat{\widehat{m}}$ are special cases of $\tilde{m}$.

The following theorem gives the rates of pointwise and uniform convergence of $F_{\tilde{m},n}$ to F under i.i.d. assumption. It follows [25, Theorem 3].

Theorem 5.10. *If $X_1, X_2, \ldots$ are i.i.d. as already assumed in the beginning of the section and $\tilde{m} \geq cn^{\delta+1/2}$ for $c, \delta > 0$, then for all $0 \leq x \leq 1$ it holds for $n \to \infty$ that*

(a) $\left\|\hat{F}_{\tilde{m},n} - F\right\| = o(n^{-1/3})$ *a.s. and* $\text{Bias}\left[\hat{F}_{\tilde{m},n}\right] = o(n^{-1/3})$,

(b) $\limsup \left(\frac{2n}{\log\log(n)}\right)^{1/2} \left\|\hat{F}_{\tilde{m},n} - F\right\| \leq 1$ *a.s., and*

(c) $\frac{\left\|\hat{F}_{\tilde{m},n} - F\right\|}{\left\|F_n - F\right\|} \leq 1 + o_P(1)$ *a.s.,*

where $\|\cdot\|$ is the sup-norm defined in Theorem 5.2 and $X_n = o_P(1)$ means that $\lim_{n\to\infty}(|X_n| \geq \epsilon) = 0$ for all $\epsilon > 0$. End Theorem

For the proof, see Proofs Bernstein.

Sup-norm convergence implies point-wise convergence. As the support of F is on $[0, 1]$, the sup-norm convergence of $\hat{F}_{\tilde{m},n}$ also implies that $\int_0^1 \left|\hat{F}_{\tilde{m},n}(x) - F(x)\right| \mathrm{d}x$ and MISE $= \int_0^1 \mathbb{E}\left[\hat{F}_{\tilde{m},n}(x) - F(x)\right]^2 \mathrm{d}x$ converge to zero for $n \to \infty$.

Part (b) says that the Chung-Smirnov property that was shown for nonrandom m in Eq. (5.4) still holds for random $m = \tilde{m}$.

In part (c), one can see that for large n, the probability that the EDF is more accurate than the estimator $F_{\tilde{m},n}$, is very small.

5.7.4 Simulation

The simulation in [25] compares the Monte Carlo estimates of the MSE coming from different kernel and Bernstein estimators and the EDF. The bandwidths used for the kernel estimators are from [10], [26], [27], and h_{opt} from Eq. (3.2) while the degrees for the Bernstein estimators are $\widehat{m}, \widehat{\widehat{m}}$ and m_{opt} from Eq. (5.11).

The sample size is chosen to be $n = 100$ and 12 different distributions are tested. Nine of them are beta distributions with different parameters and three are $AR(1)$-processes. The MSE is calculated for five points: the 5th, 25th, median, 75th, and 95th percentiles.

The main observations that were found in [25] are the following.

(a) Even in the case that the density is given, m_{opt} and h_{opt} are not always the best choices for m and h.

(b) $\widehat{m}$ compares very well with m_{opt} for the 5th and 95th percentile. In many cases it performs even better.

(c) m_{opt} varies a lot depending on the underlying distribution, $\widehat{m}$ and $\widehat{\widehat{m}}$ not so much.

(d) In the tail region, i.e., the 5th and 95th percentile, [10],[26], and $\widehat{m}$ seem to outperform the empirical EDF or the other kernel estimators in almost all the examples. But none of them can be said to be the best of the three.

(e) In the interquartile region, [27], $\widehat{\widehat{m}}$ and $\widehat{m}$ seem to be better than the other estimators.

(f) In the tail regions of the AR(1)-models, $\widehat{m}$ seems to be the most accurate estimator.

5.8 Properties of $P_{k,m}$

We now present a few properties of the Bernstein polynomials that are needed for some proofs in this thesis. The following lemma can be found in [17, Lemma 2 and Lemma 3].

Lemma 5.3. *Define*

$$L_m(x) = \sum_{k=0}^{m} P_{k,m}^2(x)$$

and

$$R_{j,m}(x) = m^{-j} \sum\sum_{0 \leq k < l \leq m} (k - mx)^j P_{k,m}(x) P_{l,m}(x)$$

for $j \in \{0, 1, 2\}$. *It trivially holds that*

1. $L_m(0) = L_m(1) = 1$,
2. $0 \leq L_m(x) \leq 1$ *for* $x \in [0, 1]$, *and*
3. $R_{j,m}(0) = R_{j,m}(1) = 0$ *for* $j \in \{0, 1, 2\}$.

In addition, defining $\phi_1(x) = [4\pi x(1-x)]^{-1/2}$ *and* $\phi_2(x) = \left[\frac{x(1-x)}{2\pi}\right]^{1/2}$ *and letting* g *be a continuous function on* $[0, 1]$ *leads to the properties*

(a) $0 \leq R_{2,m}(x) \leq (4m)^{-1}$ *for* $x \in (0, 1)$,

(b) $L_m(x) = m^{-1/2}\left[\phi_1(x) + o_x(1)\right]$ *for* $x \in (0, 1)$,

(c) $R_{1,m}(x) = m^{-1/2}\left[-\phi_2(x) + o_x(1)\right]$ *for* $x \in (0, 1)$,

(d) $m^{1/2} \int_0^1 L_m(x)\,\mathrm{d}x = \int_0^1 \phi_1(x)\,\mathrm{d}x + O(m^{-1}) = \frac{\sqrt{\pi}}{2} + O(m^{-1})$, *and*

(e) $m^{1/2} \int_0^1 g(x) R_{1,m}(x)\,\mathrm{d}x = -\int_0^1 g(x)\phi_2(x)\,\mathrm{d}x + o(1)$.

End Lemma

Proof. This proof follows the proof of Theorem 1 in [17]. For part (d), [28] helps to understand some equations.
For the proof, define

$$T_{j,m}(x) = \sum_{k=0}^{m} (k - mx)^j P_{k,m}(x) \text{ and } H_{j,m}(x) = \sum_{k=0}^{m} |k - mx|^j P_{k,m}(x). \tag{5.19}$$

Then it holds with [19, Section 1.5] that

$$T_{j+1,m}(x) = x(1-x)\left[T_{j,m}'(x) + mjT_{j-1,m}(x)\right].$$

With this, it is easy to see that

$$\begin{aligned} &T_{1,m}(x) = 0, T_{2,m}(x) = mx(1-x), T_{3,m}(x) = mx(1-x)(1-2x), \\ &T_{4,m}(x) = 3m(m-2)x^2(1-x)^2 + mx(1-x). \end{aligned} \tag{5.20}$$

Now, we proof the lemma.

(a) The positivity of $R_{2,m}(x)$ is clear. For the second inequality, note that

$$\begin{aligned} R_{2,m}(x) &= m^{-2} \sum\sum_{0\le k<l\le m} (k-mx)^2 P_{k,m}(x) P_{l,m}(x) \\ &\le m^{-2} \sum_{k=0}^{m}\sum_{l=0}^{m} (k-mx)^2 P_{k,m}(x) P_{l,m}(x) \\ &= m^{-2} T_{2,m}(x) \\ &= \frac{x(1-x)}{m} \le \frac{1}{4m}. \end{aligned}$$

(b) Let $U_i, W_j, i, j \in \{1, .., m\}$ be i.i.d. Bernoulli random variables with $\mathbb{P}(U_1 = 1) = x$ and

$$R_i = \frac{U_i - W_i}{\sqrt{2x(1-x)}}.$$

Then it holds that $\mathbb{E}[R_i] = 0, \mathrm{Var}[R_i] = 1$ and R_i has a lattice distribution with span

$$h = \frac{1}{\sqrt{2x(1-x)}}.$$

Note that

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{m} R_i = 0\right) &= \mathbb{P}\left(\sum_{i=1}^{m} U_i = \sum_{i=1}^{m} W_i\right) \\ &= \mathbb{P}\left(\sum_{i=1}^{m} U_i = \sum_{i=1}^{m} W_i = 0\right) + ... + \mathbb{P}\left(\sum_{i=1}^{m} U_i = \sum_{i=1}^{m} W_i = m\right) \\ &= \sum_{k=0}^{m} P_{k,m}(x)^2. \end{aligned}$$

With [29, XV.5, p. 517, Theorem 3], it holds that

$$\frac{\sqrt{m}}{h} \sum_{k=0}^{m} P_{k,m}^2(x) - \frac{1}{\sqrt{2\pi}} \to 0$$

and it follows that

$$\sqrt{4\pi m x(1-x)} \sum_{k=0}^{m} P_{k,m}^2(x) \to 1$$

from which the claim follows.

(c) In [30, Theorem 1], it was shown that

$$\sum_{l=k}^{m} P_{l,m}(x) = 1 - \Phi(\delta_k - G_x(\delta_{k-1/2})) + O_x(m^{-1}),$$

where $O_x(m^{-1})$ is independent of k, Φ is the standard normal distribution function,

$$\delta_k = (k - mx)[mx(1-x)]^{-1/2},$$

and

$$G_x(t) = \left[\frac{1}{2} + \frac{1}{6}(1-2x)(t^2-1)\right][mx(1-x)]^{-1/2}.$$

Expanding $\Phi(t)$ about $t = 0$ leads to

$$\Phi(t) = \frac{1}{2} + \frac{t}{\sqrt{2\pi}} + o(|t|),$$

from where we get that

$$\sum_{l=k+1}^{m} P_{l,m}(x) = \frac{1}{2} - \frac{\delta_{k+1} - G_x(\delta_{k+1/2})}{\sqrt{2\pi}} + o_x(|\delta_{k+1} - G_x(\delta_{k+1/2})|) + O_x(m^{-1}),$$

where $O_x(m^{-1})$ is again independent of k.

Now, note that

$$\begin{aligned}
\delta_{k+1} - G_x(\delta_{k+1/2}) &= (k+1-mx)A - \left[\frac{1}{2} + \frac{1}{6}(1-2x)(\delta_{k+1/2}^2 - 1)\right] A \\
&= \delta_k + A - \left[\frac{1}{2} + \frac{1}{6}(1-2x)(\delta_k^2 + \delta_k A + \frac{1}{4}A^2 - 1)\right] A \\
&= \frac{1}{3}(2-x)A + \left[1 - \frac{1}{6}(1-2x)A^2\right]\delta_k - \frac{1}{6}(1-2x)A\delta_k^2 + \frac{1}{24}(1-2x)A^3 \\
&= \frac{1}{3}(2-x)[mx(1-x)]^{-1/2} + \left[1 - \frac{1}{6}(1-2x)[mx(1-x)]^{-1}\right]\delta_k \\
&\quad - \frac{1}{6}(1-2x)[mx(1-x)]^{-1/2}\delta_k^2 + O_x(m^{-3/2}),
\end{aligned}$$

where $A = [mx(1-x)]^{-1/2}$.

From this it follows that

$$\begin{aligned}
R_{1,m}(x) &= m^{-1}\sum_{k=0}^{m}(k-mx)P_{k,m}(x)\left[\sum_{l=k+1}^{m} P_{l,m}(x)\right] \\
&= \left[\frac{1}{2} - \frac{1}{3}(2-x)[2\pi mx(1-x)]^{-1/2}\right] m^{-1}T_{1,m}(x) \\
&\quad - [2\pi mx(1-x)]^{-1/2}m^{-1}T_{2,m}(x) \\
&\quad + o_x(m^{-3/2}H_{1,m}(x)) + o_x(m^{-3/2}H_{2,m}(x)) + O_x(m^{-5/2}H_{3,m}(x)).
\end{aligned}$$

Inserting Eq. (5.20), it holds that

$$\begin{aligned}
R_{1,m}(x) &= -x(1-x)[2\pi mx(1-x)]^{-1/2} + o_x(m^{-1/2}) \\
&\quad + o_x(m^{-3/2}H_{1,m}(x)) + O_x(m^{-5/2}H_{3,m}(x)).
\end{aligned} \tag{5.21}$$

The Cauchy-Schwarz inequality implies that

$$\begin{aligned}
m^{-3/2}H_{1,m}(x) &\le m^{-3/2}[T_{2,m}(x)]^{1/2} \\
&= m^{-3/2}[mx(1-x)]^{1/2} \le \frac{1}{2m} = O(m^{-1})
\end{aligned} \tag{5.22}$$

and

$$\begin{aligned}
m^{-5/2}H_{3,m}(x) &\le m^{-5/2}[T_{2,m}(x)T_{4,m}(x)]^{1/2} \\
&= m^{-5/2}[3m^2(m-2)x^3(1-x)^3 + m^2x^2(1-x)^2]^{1/2} = O(m^{-1}).
\end{aligned}$$

Inserting this into Eq. (5.21), it follows that

$$\begin{aligned} R_{1,m}(x) &= -x(1-x)[2\pi m x(1-x)]^{-1/2} + o_x(m^{-1/2}) + O_x(m^{-1}) \\ &= m^{-1/2}\left[-\frac{\sqrt{x(1-x)}}{\sqrt{2\pi}} + o_x(1)\right]. \end{aligned}$$

(d) It holds that

$$\begin{aligned} \sum_{k=0}^{m}\int_0^1 P_{k,m}^2(x) &= \sum_{k=0}^{m}\left(\frac{\Gamma(m+1)}{\Gamma(k+1)\Gamma(m-k+1)}\right)^2 \int_0^1 x^{2k}(1-x)^{2(m-k)}\,\mathrm{d}x \\ &= \sum_{k=0}^{m}\left(\frac{\Gamma(m+1)}{\Gamma(k+1)\Gamma(m-k+1)}\right)^2 \frac{\Gamma(2k+1)\Gamma(2(m-k)+1)}{\Gamma(2m+2)} \\ &= \frac{\Gamma(m+1)^2}{\Gamma(2m+2)}\sum_{k=0}^{m}\binom{2k}{k}\binom{2(m-k)}{m-k} \\ &= \frac{\Gamma(m+1)^2}{\Gamma(2m+2)}4^m \\ &= \frac{\sqrt{\pi}}{2m+1}\frac{\Gamma(m+1)}{\Gamma(m+1/2)}, \end{aligned}$$

where the second equality uses the Betafunction and the last equality follows from the fact that

$$4^m = \frac{\Gamma(2m)2\sqrt{\pi}}{\Gamma(m)\Gamma(m+1/2)},$$

that can be found in [31, p. 256, 6.1.18]. It follows with [31, p. 257] that

$$\begin{aligned} m^{1/2}\int_0^1 L_m(x)\,\mathrm{d}x &= \frac{\sqrt{\pi m}}{2m+1}\frac{\Gamma(m+1)}{\Gamma(m+1/2)} \\ &= \frac{\sqrt{\pi m}}{2m+1}\left[1+\frac{1}{8m}+O(m^{-2})\right] = \frac{\sqrt{\pi}}{2} + O(m^{-1}) \\ &= \int_0^1 \phi_1(x)\,\mathrm{d}x + O(m^{-1}). \end{aligned}$$

This concludes the proof.

(e) With part (c) we know that $G_m(x) := m^{1/2}R_{1,m} \to -\phi_2(x) =: G(x)$ on the unit interval. It also holds that

$$\begin{aligned} |G_m(x)| &= m^{-1/2}\left|\sum_{k=0}^{m}(k-mx)P_{k,m}(x)\left[\sum_{l=k+1}^{m}P_{l,m}(x)\right]\right| \\ &\le m^{-1/2}\sum_{k=0}^{m}|k-mx|P_{k,m}(x)\left[\sum_{l=k+1}^{m}P_{l,m}(x)\right] \\ &\le m^{-1/2}H_{1,m}(x) \le \frac{1}{2} \end{aligned}$$

for $x \in [0, 1]$, where the last inequality comes from Eq. (5.22). Hence, the sequence is uniformly bounded on the unit interval and with this also uniformly integrable. With [32, Thm 16.14 on pp. 217-218], it holds that

$$\int_0^1 |G_m(x) - G(x)| \, \mathrm{d}x = o(1),$$

which implies that

$$\left| \int_0^1 g(x) G_m(x) \, \mathrm{d}x - \int_0^1 g(x) G(x) \, \mathrm{d}x \right| \leq \sup_{x \in [0,1]} |g(x)| \int_0^1 |G_m(x) - G(x)| \, \mathrm{d}x = o(1).$$

End Proof

5.9 Proofs Bernstein

Proof of Theorem 5.3. This proof follows the proof of Theorem 1 in [17]. The first part follows directly from Lemma 5.1 because of $\mathbb{E}\left[\hat{F}_{m,n}(x)\right] = B_m(x)$.

For the second part, define

$$\Delta_i(x) = \mathbb{I}(X_i \leq x) - F(x)$$

for $x \in [0,1]$, where $\mathbb{I}$ is the indicator function. It follows that the sequence of random variables $\Delta_1, ..., \Delta_n, n \in \mathbb{N}$, are i.i.d. with mean zero. Defining

$$Y_{i,m} = \sum_{k=0}^{m} \Delta_i\left(\frac{k}{m}\right) P_{k,m}(x),$$

it is easy to see that

$$\begin{aligned}
\hat{F}_{m,n}(x) - B_m(x) &= \sum_{k=0}^{m}\left[F_n\left(\frac{k}{m}\right) - F\left(\frac{k}{m}\right)\right] P_{k,m}(x) \\
&= \frac{1}{n}\sum_{k=0}^{m}\sum_{i=1}^{n}\left[\mathbb{I}\left(X_i \leq \frac{k}{m}\right) - F\left(\frac{k}{m}\right)\right] P_{k,m}(x) \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_{i,m}. \qquad (5.23)
\end{aligned}$$

The random variables $Y_{1,m}, ..., Y_{n,m}$ are also i.i.d. with mean zero for given m. This means that the variance can be calculated by

$$\begin{aligned}
\mathrm{Var}\left[\hat{F}_{m,n}(x)\right] &= \mathrm{Var}\left[\hat{F}_{m,n}(x) - B_m(x)\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[Y_{i,m}] \\
&= \frac{1}{n}\mathrm{Var}[Y_{1,m}] \\
&= \frac{1}{n}\mathbb{E}[Y_{1,m}^2]. \qquad (5.24)
\end{aligned}$$

It also holds for $x, y \in [0,1]$ that

$$\begin{aligned}
\mathbb{E}[\Delta_1(x)\Delta_1(y)] &= \mathbb{E}[(\mathbb{I}(X_1 \leq x) - F(x))(\mathbb{I}(X_1 \leq y) - F(y))] \\
&= \mathbb{E}[\mathbb{I}(X_1 \leq x)\mathbb{I}(X_1 \leq y)] - F(x)F(y) \\
&= \mathbb{E}[\mathbb{I}(X_1 \leq \min(x,y))] - F(x)F(y) \\
&= \min(F(x), F(y)) - F(x)F(y),
\end{aligned}$$

which implies

$$\begin{aligned}
\mathbb{E}[Y_{1,m}^2] &= \sum_{k=0}^{m}\sum_{l=0}^{m}\mathbb{E}\left[\Delta_1\left(\frac{k}{m}\right)\Delta_1\left(\frac{l}{m}\right)\right] P_{k,m}(x)P_{l,m}(x) \\
&= \sum_{k=0}^{m}\sum_{l=0}^{m}\min\left(F\left(\frac{k}{m}\right), F\left(\frac{l}{m}\right)\right) P_{k,m}(x)P_{l,m}(x) - \sum_{k=0}^{m}\sum_{l=0}^{m} F\left(\frac{k}{m}\right) F\left(\frac{l}{m}\right) P_{k,m}(x)P_{l,m}(x) \\
&= \sum_{k=0}^{m} F\left(\frac{k}{m}\right) P_{k,m}^2(x) + 2\sum\sum_{0\leq k<l\leq m} F\left(\frac{k}{m}\right) P_{k,m}(x)P_{l,m}(x) - B_m^2(x). \qquad (5.25)
\end{aligned}$$

Now the goal is to find an asymptotic expression of this formula. Using Taylor's theorem we get

$$F\left(\frac{k}{m}\right) = F(x) + O\left(\left|\frac{k}{m} - x\right|\right).$$

Hence, the first term of Eq. (5.25) can be written as

$$\sum_{k=0}^{m} F\left(\frac{k}{m}\right) P_{k,m}^2(x) = F(x)L_m(x) + O(I_m(x)), \tag{5.26}$$

where

$$I_m(x) = \sum_{k=0}^{m} \left|\frac{k}{m} - x\right| P_{k,m}^2(x)$$

and L_m is defined as in Lemma 5.3. We use Taylor's theorem again to obtain

$$F\left(\frac{k}{m}\right) = F(x) + \left(\frac{k}{m} - x\right) f(x) + O\left(\left(\frac{k}{m} - x\right)^2\right).$$

Note that

$$1 = \sum_{k=0}^{m}\sum_{l=0}^{m} P_{k,m}(x)P_{l,m}(x) = 2\sum\sum_{0\le k<l\le m} P_{k,m}(x)P_{l,m}(x) + \sum_{k=0}^{m} P_{k,m}^2(x) = 2R_{0,m}(x) + L_m(x),$$

from which we get that

$$R_{0,m} = \frac{1}{2}[1 - L_m(x)].$$

Using this result, the Taylor expansion and the fact that $R_{2,m} \le (4m)^{-1}$ (see Lemma 5.3) leads to the transformation

$$\begin{aligned}\sum\sum_{0\le k<l\le m} F\left(\frac{k}{m}\right) P_{k,m}(x)P_{l,m}(x) &= F(x)R_{0,m}(x) + f(x)R_{1,m}(x) + O(R_{2,m}(x)) \\ &= \frac{1}{2}F(x)(1 - L_m(x)) + f(x)R_{1,m}(x) + O(m^{-1})\end{aligned} \tag{5.27}$$

of the second term of Eq. (5.25). We are now applying Lemma 5.1 to get

$$\begin{aligned}F(x) - B_m^2(x) &= F(x) - \left(F(x) + m^{-1}b(x) + O(m^{-1})\right)^2 \\ &= F(x) - \left(F(x) + O(m^{-1})\right)^2 \\ &= F(x) - F(x)^2 - 2F(x)O(m^{-1}) - O(m^{-1})^2 \\ &= \sigma^2(x) + O(m^{-1})\end{aligned}$$

and plugging Eq. (5.26) and Eq. (5.27) into Eq. (5.25) leads to

$$\begin{aligned}\mathbb{E}[Y_{1,m}^2] &= F(x)L_m(x) + O(I_m(x)) + F(x)(1 - L_m(x)) + 2f(x)R_{1,m}(x) + O(m^{-1}) - B_m^2(x) \\ &= \sigma^2(x) + 2f(x)R_{1,m}(x) + O(I_m(x)) + O(m^{-1}).\end{aligned} \tag{5.28}$$

Now, using the fact that $0 \le P_{k,m}(x) \le 1$ and applying the Cauchy-Schwarz inequality gives

$$\begin{aligned}I_m(x) &\le \left[\sum_{k=0}^{m}\left(\frac{k}{m} - x\right)^2 P_{k,m}(x)\right]^{1/2}\left[\sum_{k=0}^{m} P_{k,m}^3(x)\right]^{1/2} \\ &\le \left[\frac{T_{2,m}}{m^2} L_m(x)\right]^{1/2} \\ &\le \left[\frac{L_m(x)}{4m}\right]^{1/2},\end{aligned} \tag{5.29}$$

where $T_{2,m}(x) = \sum_{k=0}^{m} (k - mx)^2 P_{k,m}(x) = mx(1-x) \leq \frac{m}{4}$ for $x \in [0,1]$. With Lemma 5.3 (b), this leads to

$$I_m(x) \leq \left(\frac{1}{4m^{3/2}} [\phi_1(x) + o_x(1)] \right)^{1/2} = o_x(m^{-3/4}) = O_x(m^{-3/4}). \tag{5.30}$$

Finally we can calculate with Lemma 5.3 (c) that

$$\begin{aligned} \mathbb{E}[Y_{1,m}^2] &= \sigma^2(x) + 2f(x)R_{1,m}(x) + O(I_m(x)) + O(m^{-1}) \\ &= \sigma^2(x) + 2f(x)m^{-1/2}[-\phi_2(x) + o_x(1)] + O(I_m(x)) + O(m^{-1}) \\ &= \sigma^2(x) - m^{-1/2}V(x) + o_x(m^{-1/2}), \end{aligned} \tag{5.31}$$

where the last part comes from Eq. (5.30). With Eq. (5.24) this leads to the desired asymptotic expression

$$\text{Var}\left[\hat{F}_{m,n}(x)\right] = n^{-1}\sigma^2(x) - m^{-1/2}n^{-1}V(x) + o_x(m^{-1/2}n^{-1})$$

for the variance. End Proof

Proof of Theorem 5.4. This proof follows the proof of Theorem 2 in [17]. For fixed m we know from the proof of Theorem 5.3 that

$$\hat{F}_{m,n}(x) - B_m(x) = \frac{1}{n} \sum_{i=1}^{n} Y_{i,m},$$

where the $Y_{i,m}$ are i.i.d. random variables with mean 0. Define $\gamma_m^2 = \mathbb{E}[Y_{1,m}^2]$. We use the central limit theorem for double arrays (see [33], Section 1.9.3) to show the claim.

Defining

$$A_n = \mathbb{E}\left[\sum_{i=1}^{n} Y_{i,m}\right] = 0 \text{ and } B_n^2 = \text{Var}\left[\sum_{i=1}^{n} Y_{i,m}\right] = n\gamma_m^2,$$

it says that

$$\frac{\sum_{i=1}^{n} Y_{i,m} - A_n}{B_n} \xrightarrow{D} \mathcal{N}(0,1)$$

if and only if the Lindeberg condition

$$\frac{n\mathbb{E}[\mathbb{I}(|Y_{1,m}| > \epsilon B_n)Y_{1,m}^2]}{B_n^2} \to 0 \text{ for } n \to \infty \text{ and all } \epsilon > 0$$

is satisfied. With Eq. (5.31) we know that $\gamma_m \to \sigma(x)$ for $m \to \infty$ (which follows from $n \to \infty$) and it holds for $n \to \infty$ that

$$\begin{aligned} &\frac{\sum_{i=1}^{n} Y_{i,m} - A_n}{B_n} \xrightarrow{D} \mathcal{N}(0,1) \\ \Leftrightarrow\ &\frac{\sum_{i=1}^{n} Y_{i,m}}{\sqrt{n} \cdot \gamma_m} \xrightarrow{D} \mathcal{N}(0,1) \\ \Leftrightarrow\ &\frac{\sqrt{n}}{\gamma_m}\left(\hat{F}_{m,n}(x) - B_m(x)\right) \xrightarrow{D} \mathcal{N}(0,1) \\ \Leftrightarrow\ &\sqrt{n}\left(\hat{F}_{m,n}(x) - B_m(x)\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right), \end{aligned}$$

which is the claim of Theorem 5.4. In our case the Lindeberg condition has the form

$$\frac{\mathbb{E}[\mathbb{I}(|Y_{1,m}| > \epsilon\sqrt{n}\gamma_m)Y_{1,m}^2]}{\gamma_m^2} \to 0 \text{ for } n \to \infty \text{ and all } \epsilon > 0.$$

This is what has to be shown to proof the theorem. Using the fact that

$$|Y_{1,m}| \leq \sum_{k=0}^{m} \left| \Delta_1 \left(\frac{k}{m} \right) \right| P_{k,m}(x) \leq \sum_{k=0}^{m} P_{k,m}(x) = 1$$

leads to

$$\mathbb{I}(|Y_{1,m}| > \epsilon\sqrt{n}\gamma_m) \leq \mathbb{I}(1 > \epsilon\sqrt{n}\gamma_m) \to 0,$$

which gives the desired result. End Proof

Proof of Theorem 5.6. This proof follows the proof of Theorem 3 in [17]. Using a part of Eq. (5.29), Jensen's inequality, and Lemma 5.3 (d) leads to

$$\int_0^1 I_m(x)\,\mathrm{d}x \leq \left[\frac{1}{4m} \int_0^1 L_m(x)\,\mathrm{d}x \right]^{1/2} = \left[\frac{1}{4m^{3/2}} \left(\frac{\sqrt{\pi}}{2} + O(m^{-1}) \right) \right]^{1/2} = O(m^{-3/4}).$$

Using Eq. (5.23),Lemma 5.1, and Eq. (5.28)

$$\begin{aligned} \mathrm{MISE}\left[\hat{F}_{m,n}\right] &= \int_0^1 \mathrm{Var}\left[\hat{F}_{m,n}(x)\right] + \mathrm{Bias}\left[\hat{F}_{m,n}(x)\right]^2 \mathrm{d}x \\ &= \frac{1}{n}\int_0^1 \sigma^2(x) + 2f(x)R_{1,m}(x) + O(I_m(x)) + O(m^{-1})\,\mathrm{d}x + \frac{1}{m^2}\int_0^1 b^2(x) + o(1)\,\mathrm{d}x \\ &= \frac{1}{n}\int_0^1 \sigma^2(x) + 2f(x)R_{1,m}(x)\,\mathrm{d}x + \frac{1}{m^2}\int_0^1 b^2(x)\,\mathrm{d}x + O(m^{-3/4}n^{-1}) + o(m^{-2}), \end{aligned}$$

because of [34, Section 4.2.1] and the fact that I_m is positive. Now, we get

$$\mathrm{MISE}\left[\hat{F}_{m,n}\right] = n^{-1}C_1 - n^{-1}m^{-1/2}C_2 + m^{-2}C_3 + o(m^{-2}) + o(m^{-1/2}n^{-1})$$

with $2f(x)\phi_2(x) = V(x)$ and Lemma 5.3 (e). End Proof

Proof of Theorem 5.7. This proof follows the proof of Theorem 4 in [17].

We only show the claims for the global variable $i_G(n)$ because the local part can be shown almost analogously using $\theta(x) = \frac{V(x)}{\sigma^2(x)}$ and $\gamma(x) = \frac{b^2(x)}{\sigma^2(x)}$ instead of $\tilde{\theta}$ and $\tilde{\gamma}$. With the dependency on x, the error terms are x-dependent as well.

For simplicity we write $i(n) = i_G(n)$. By the definition of $i(n)$ we know that $\lim_{n\to\infty} i(n) = \infty$ and

$$\begin{aligned} &\mathrm{MISE}\left[F_{i(n)}\right] \leq \mathrm{MISE}\left[\hat{F}_{m,n}\right] \leq \mathrm{MISE}\left[F_{i(n)-1}\right] \\ \Leftrightarrow\ & i(n)^{-1}C_1 \leq n^{-1}C_1 - m^{-1/2}n^{-1}C_2 + m^{-2}C_3 + o(m^{-1/2}n^{-1}) + o(m^{-2}) \leq (i(n)-1)^{-1}C_1 \\ \Leftrightarrow\ & 1 \leq \frac{i(n)}{n}\left[1 - m^{-1/2}\tilde{\theta} + m^{-2}n\tilde{\gamma} + o(m^{-1/2}) + o(m^{-2}n)\right] \leq \frac{i(n)}{i(n)-1}, \end{aligned} \tag{5.32}$$

where $\tilde{\theta} = \frac{C_2}{C_1}$ and $\tilde{\gamma} = \frac{C_3}{C_1}$. Now, if $mn^{-1/2} \to \infty$ ($\Leftrightarrow m^{-2}n \to 0$), taking the limit $n \to \infty$ leads to

$$\frac{i(n)}{n} \to 1,$$

so that

$$i(n) = n + o(n) = n(1 + o(1)).$$

(a) We assume that that $mn^{-2/3} \to \infty$ and $mn^{-2} \to 0$. Rewrite Eq. (5.32) as

$$\begin{aligned} & m^{-1/2}n^{-1}\tilde{\theta} \le A_{1,n} + m^{-2}\tilde{\gamma} + o(m^{-1/2}n^{-1}) + o(m^{-2}) \le m^{-1/2}n^{-1}\tilde{\theta} + A_{2,n} \\ \Leftrightarrow\ & \tilde{\theta} \le m^{1/2}nA_{1,n} + m^{-3/2}n\tilde{\gamma} + o(1) + o(m^{-3/2}n) \le \tilde{\theta} + m^{1/2}nA_{2,n}, \end{aligned} \tag{5.33}$$

where

$$A_{1,n} = \frac{1}{n} - \frac{1}{i(n)} \text{ and } A_{2,n} = \frac{1}{i(n)-1} - \frac{1}{i(n)}.$$

It holds that

$$\lim_{n\to\infty} m^{1/2}nA_{1,n} = \left(\lim_{n\to\infty} \frac{i(n)-n}{m^{-1/2}n}\right)\left(\lim_{n\to\infty} \frac{n}{i(n)}\right) = \lim_{n\to\infty} \frac{i(n)-n}{m^{-1/2}n},$$

and because $m^{1/2}n^{-1} = (mn^{-2})^{1/2} \to 0$

$$\lim_{n\to\infty} m^{1/2}nA_{2,n} = \left(\lim_{n\to\infty} m^{1/2}n^{-1}\right)\left(\lim_{n\to\infty} \frac{n}{i(n)}\right)\left(\lim_{n\to\infty} \frac{n}{i(n)-1}\right) = 0.$$

We also know that $m^{-3/2}n = (mn^{-2/3})^{-3/2} \to 0$ so that

$$\lim_{n\to\infty} \frac{i(n)-n}{m^{-1/2}n} = \tilde{\theta} \Rightarrow \frac{i(n)-n}{m^{-1/2}n} = \tilde{\theta} + o(1)$$

follows from Eq. (5.33).

(b) The second part can be proven with very similar arguments. If $mn^{-2/3} \to c$ it also holds that $m^{-2}n = (mn^{-2/3})^{-3/2}m^{-1/2} \to 0$ and $m^{1/2}n^{-1} = (mn^{-2/3})^{1/2}n^{-2/3} \to 0$ so that

$$\lim_{n\to\infty} \frac{i(n)-n}{m^{-1/2}n} = \tilde{\theta} - c^{-3/2}\tilde{\gamma}$$

and with

$$\lim_{n\to\infty} \frac{i(n)-n}{m^{-1/2}n} = \left(\lim_{n\to\infty} \frac{i(n)-n}{n^{2/3}}\right)\left(\lim_{n\to\infty} m^{1/2}n^{-1/3}\right) = c^{1/2} \lim_{n\to\infty} \frac{i(n)-n}{n^{2/3}}$$

the claim

$$c^{1/2}\frac{i(n)-n}{n^{2/3}} = \tilde{\theta} - c^{-3/2}\tilde{\gamma} + o(1)$$

holds.

End Proof

Proof of Theorem 5.8. This proof follows the proof of Theorem 1 in [25]. It holds that

$$\begin{aligned} \left|\mathrm{Var}^*\left[\hat{F}_{m,n}(x)\right] - \mathrm{Var}\left[\hat{F}_{m,n}(x)\right]\right| \le & \frac{1}{n}\left[\sum_{k=0}^{m}\left|F_n\left(\frac{k}{m}\right) - F\left(\frac{k}{m}\right)\right| P_{k,m}^2(x)\right. \\ & +2\sum\sum_{0\le k<l\le m}\left|F_n\left(\frac{k}{m}\right) - F\left(\frac{k}{m}\right)\right| P_{k,m}(x)P_{l,m}(x) \\ & \left.+\sum_{k=0}^{m}\left|F_n\left(\frac{k}{m}\right) - F\left(\frac{k}{m}\right)\right| P_{k,m}(x)\right] \\ \le & \frac{4}{n}\|F_n - F\|. \end{aligned}$$

This means that $\left|\mathrm{Var}^*\left[\hat{F}_{m,n}(x)\right] - \mathrm{Var}\left[\hat{F}_{m,n}(x)\right]\right| = o\left(n^{-1}\right)$. For the bias it follows with Lemma 5.1 that

$$\begin{aligned}
\mathrm{Bias}\left[\hat{F}_{m,n}(x)\right]^2 &= \left[\frac{x(1-x)}{2m}f'(x) + o(m^{-1})\right]^2 \\
&= \left[\frac{x(1-x)}{2m}f'(x)\right]^2 + \frac{x(1-x)}{2m}f'(x)o(m^{-1}) + o(m^{-2}) \\
&= \left[\frac{x(1-x)}{2m}f'(x)\right]^2 + o(m^{-2})
\end{aligned}$$

and it holds with very similar arguments as in the proof of Lemma 5.1 and as in the equation before with $F''_{h,n}(x) = f'_n(x)$ that

$$\mathrm{Bias}^*\left[\hat{F}_{m,n}(x)\right]^2 = \left[\frac{x(1-x)}{2m}f'_n(x)\right]^2 + o(m^{-2}) \text{ a.s.,}$$

where

$$f'_n(x) = \frac{1}{nh^2}\sum_{i=1}^{n} K'\left(\frac{x-X_i}{h}\right).$$

Under the assumption that h is a multiple of $n^{-1/9}$ it holds that

$$f'_n(x) \to f'(x) \text{ a.s.} \tag{5.34}$$

for $n \to \infty$ which is shown at the end of the proof. With this, it holds that

$$\mathrm{Bias}^*\left[\hat{F}_{m,n}(x)\right]^2 - \mathrm{Bias}\left[\hat{F}_{m,n}(x)\right]^2 = o(m^{-2})$$

for $m \to \infty$. The claim follows with

$$\begin{aligned}
&\mathrm{MSE}^*\left[\hat{F}_{m_{opt},n}(x)\right] - \mathrm{MSE}\left[\hat{F}_{m_{opt},n}(x)\right] \\
&= \mathrm{Var}^*\left[\hat{F}_{m,n}(x)\right] - \mathrm{Var}\left[\hat{F}_{m,n}(x)\right] + \mathrm{Bias}^*\left[\hat{F}_{m,n}(x)\right]^2 - \mathrm{Bias}\left[\hat{F}_{m,n}(x)\right]^2 \\
&= o(n^{-1}) + o(m^{-2}) = o(n^{-2})
\end{aligned}$$

for $n \to \infty$, where the last step holds because of the conditions on m.

Now, we show Eq. (5.34). It holds that

$$\mathbb{E}\left[f'_n(x)\right] = \mathbb{E}\left[F''_{h,n}(x)\right] = \mathbb{E}\left[f'_{h,n}(x)\right] = f'(x) + o(1)$$

for $n \to \infty$ and hence, for all $\epsilon > 0$ there exists an N so that for all $n > N$, we get

$$\left|\mathbb{E}\left[f'_n(x)\right] - f'(x)\right| < \frac{\epsilon}{2}.$$

Then the triangle inequality yields for $n > N$ that

$$\mathbb{P}(|f'_n(x) - f'(x)| > \epsilon) \leq \mathbb{P}\left(|\mathbb{E}[f'_n(x)] - f'_n(x)| > \frac{\epsilon}{2}\right) = \mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n} Y_{ni}\right| > \frac{\epsilon h^2}{2}\right)$$

where $Y_{ni} = K'\left(\frac{x-X_i}{h}\right) - \mathbb{E}\left[K'\left(\frac{x-X_i}{h}\right)\right], i \in \{1, ..., n\}$, are independent mean-zero random variables. It holds that $\sup_i \|Y_{ni}\|_\infty \leq \|K'\|$ where constants are omitted. Now, using [35], it holds for $n \geq 4$ that

$$\mathbb{P}(|f'_n(x) - f'(x)| > \epsilon) \leq exp\left(-\frac{C\epsilon^2 h^4 n}{\|K'\|^2 + \epsilon h^2\|K'\| log(n) log(log(n))}\right)$$

where some constants are omitted again and for h being a multiple of $n^{-1/9}$ we know that

$$\mathbb{P}(|f_n'(x) - f'(x)| > \epsilon) \leq exp\left(-\frac{C_1\epsilon^2 n^{5/9}}{log(n)log(log(n))}\right), C_1 > 0.$$

This can be used to show with the Borel-Cantelli theorem that $f_n'(x) \to f'(x)$ a.s. for $n \to \infty$.

End Proof

Proof of Theorem 5.9. This proof follows the proof of Theorem 2 in [25].

It is enough to show that $\left|\text{MSE}^*\left[\hat{F}_{m_{opt},n}(x)\right] - \text{MSE}\left[\hat{F}_{m_{opt},n}(x)\right]\right|$ can be bounded by a constant free of n and x because then it follows that the integral exists and hence

$$\begin{aligned}\text{MISE}^*\left[\hat{F}_{m_{opt},n}\right] - \text{MISE}\left[\hat{F}_{m_{opt},n}\right] &= \int_0^1 \text{MSE}^*\left[\hat{F}_{m_{opt},n}(x)\right] - \text{MSE}\left[\hat{F}_{m_{opt},n}(x)\right] \mathrm{d}x \\ &= \int_0^1 o(n^{-1})\,\mathrm{d}x = o(n^{-1}).\end{aligned}$$

We know from the proof before that

$$\left|\text{Var}^*\left[\hat{F}_{m,n}(x)\right] - \text{Var}\left[\hat{F}_{m,n}(x)\right]\right| \leq \frac{4}{n}\|F_n - F\| \leq \frac{4}{n} \leq 4.$$

With Eq. (5.15) and Eq. (5.14) we know that the absolute values of $\text{Bias}^*\left[\hat{F}_{m,n}(x)\right]$ and $\text{Bias}\left[\hat{F}_{m,n}(x)\right]$ are both bounded by 1. Summarizing the former results leads to

$$\left|\text{MSE}^*\left[\hat{F}_{m_{opt},n}(x)\right] - \text{MSE}\left[\hat{F}_{m_{opt},n}(x)\right]\right| \leq 4 + 2 = 6,$$

which is a constant boundary.

End Proof

Proof of Theorem 5.10. As the proofs before, this one follows from [25, Theorem 3].

(a) With the Dvoretzky, Kiefer and Wolfowitz inequality that can be found in [33, p. 59], there exists a finite positive constant C not depending on F such that

$$\mathbb{P}(\|F_n - F\| > d) \leq Ce^{-2nd^2}, d > 0$$

for all $n = 1, 2, \ldots$. It follows that

$$\mathbb{P}(n^{1/3}\|F_n - F\| > d) \leq Ce^{-2d^2 n^{1/3}}$$

for all $d > 0$ and with the Borel-Cantelli theorem, it follows that

$$\|F_n - F\| = o(n^{-1/3}) \text{ a.s..}$$

It holds that

$$\begin{aligned}\left|\hat{F}_{\tilde{m},n}(x) - B_{\tilde{m}}(x)\right| &= \left|\sum_{k=0}^{\tilde{m}} F_n\left(\frac{k}{\tilde{m}}\right) P_{k,\tilde{m}}(x) - \sum_{k=0}^{\tilde{m}} F\left(\frac{k}{\tilde{m}}\right) P_{k,\tilde{m}}(x)\right| \\ &\leq \sum_{k=0}^{\tilde{m}} \left|F_n\left(\frac{k}{\tilde{m}}\right) - F\left(\frac{k}{\tilde{m}}\right)\right| P_{k,\tilde{m}}(x) \leq \|F_n - F\|\end{aligned}$$

and with this
$$\left\|\hat{F}_{\tilde{m},n} - B_{\tilde{m}}\right\| \le \|F_n - F\| \text{ a.s.},$$
as the right side does not depend on x.

It is easy to see that
$$B_{\tilde{m}}(x) = \sum_{k=0}^{\tilde{m}} F\left(\frac{k}{\tilde{m}}\right) P_{k,\tilde{m}}(x) = \mathbb{E}\left[F\left(\frac{Z}{\tilde{m}}\right)\Big|A_n\right],$$
where $A_n = \sigma(X_1, ...X_n)$ and Z has the conditional distribution $\text{Bin}(\tilde{m}, x)$. Using the Taylor Theorem it holds that
$$\begin{aligned}
|B_{\tilde{m}}(x) - F(x)| &= \left|\mathbb{E}\left[F\left(\frac{Z}{\tilde{m}}\right)\Big|A_n\right] - F(x)\right| \\
&= \left|\mathbb{E}\left[F(x) + \left(\frac{Z}{\tilde{m}} - x\right) f(x) + \frac{1}{2}\left(\frac{Z}{\tilde{m}} - x\right)^2 f'(x)\Big|A_n\right] - F(x)\right| \\
&= \left|f(x)\mathbb{E}\left[\frac{Z}{\tilde{m}} - x\Big|A_n\right] + \frac{1}{2}f'(x)\mathbb{E}\left[\left(\frac{Z}{\tilde{m}} - x\right)^2\Big|A_n\right]\right| \\
&\le \frac{1}{2}\|f'\| \operatorname{Var}\left[\frac{Z}{\tilde{m}} - x\Big|A_n\right] = \frac{1}{2}\|f'\|\frac{1}{\tilde{m}^2}\tilde{m}x(1-x) \le \frac{\|f'\|}{2cn^{\delta+1/2}}.
\end{aligned}$$
It follows that
$$\|B_{\tilde{m}}(x) - F(x)\| \le \frac{\|f'\|}{2cn^{\delta+1/2}}$$
and in total
$$\begin{aligned}
\left\|\hat{F}_{\tilde{m},n} - F\right\| &\le \left\|\hat{F}_{\tilde{m},n} - B_{\tilde{m}}\right\| + \|B_{\tilde{m}} - F\| \\
&\le \|F_n - F\| + \frac{\|f'\|}{2cn^{\delta+1/2}} = o(n^{-1/3}) + o(n^{-\delta-1/2}) = o(n^{-1/3}) \text{ a.s.,}
\end{aligned} \tag{5.35}$$
which proves the first claim of the first part.

For the second claim, note that $\left\|\hat{F}_{\tilde{m},n} - F\right\| \le 1$ and with the same arguments as in the proof of Theorem 5.9,
$$\mathbb{E}\left\|\hat{F}_{\tilde{m},n} - F\right\| = o(n^{-1/3}).$$
From this, the claim follows because
$$\mathbb{E}\left[\sup_{x\in[0,1]}\left|\hat{F}_{\tilde{m},n}(x) - F(x)\right|\right] \ge \mathbb{E}\left[\hat{F}_{\tilde{m},n}(x) - F(x)\right] = \text{Bias}\left[\hat{F}_{m,n}(x)\right].$$

(b) Note that with Eq. (5.35),
$$\begin{aligned}
&\limsup_{n\to\infty}\left(\frac{2n}{\log\log n}\right)^{1/2}\left\|\hat{F}_{\tilde{m},n} - F\right\| \\
&\le \limsup_{n\to\infty}\left(\frac{2n}{\log\log n}\right)^{1/2}\|F_n - F\| + \limsup_{n\to\infty}\left(\frac{2n}{\log\log n}\right)^{1/2}\frac{\|f'\|}{2cn^{\delta+1/2}},
\end{aligned}$$
where the first part can be bounded by one because of Eq. (5.4) and the second part converges to 0.

(c) Again with Eq. (5.35), it is easy to see that

$$\frac{\left\|\hat{F}_{\tilde{m},n} - F\right\|}{\|F_n - F\|} \leq 1 + \frac{\|f'\|}{2cn^{\delta+1/2}\|F_n - F\|} \text{ a.s..}$$

With Theorem 2.1, we know that

$$n^{1/2}(F_n(x) - F(x)) \xrightarrow{D} \mathcal{N}(0, \sigma^2(x))$$

and with the continuous mapping theorem, it is easy to see that $n^{1/2}\|F_n - F\|$ converges in distribution to a positive-valued random variable. With this, $\frac{\|f'\|}{2cn^{\delta+1/2}\|F_n - F\|}$ converges in distribution to 0 which is a constant so that it also converges in probability. This proves the claim.

End Proof

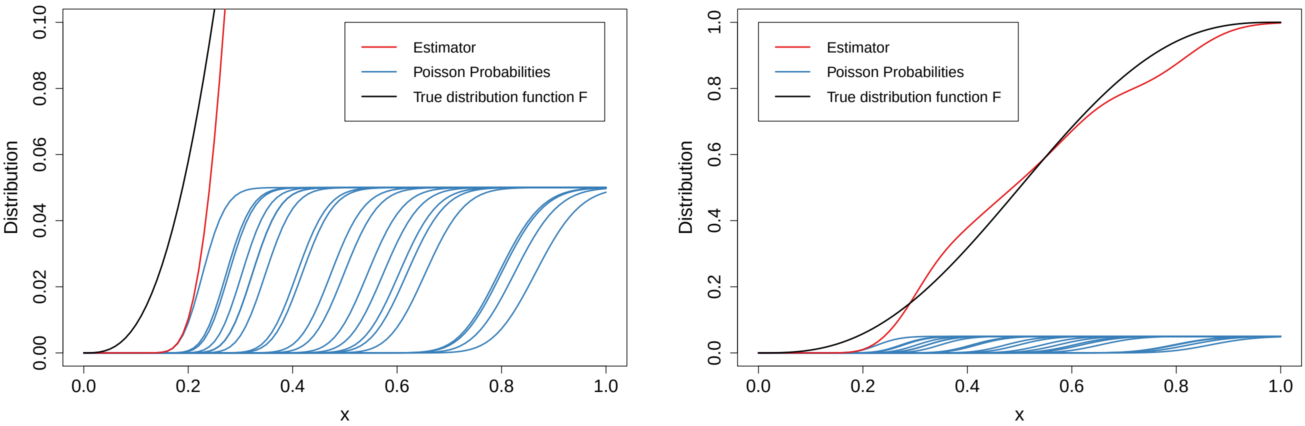

Figure 7: Illustration of the Szasz estimator for $n = 20$.

6 Szasz Distribution Function Estimation on Half Line

As already mentioned in Section 5, it is not enough to consider distribution functions on $[0, 1]$. This is the reason why in this section, we deal with distributions on $[0, \infty)$.

Similar to Section 5, let $X_1, X_2, ...$ be a sequence of i.i.d. random variables with unknown distribution function F and unknown density function f supported on the interval $[0, \infty)$. We assume that a finite random sample $X_1, ..., X_n, n \in \mathbb{N}$, is available.

In order to estimate a distribution function on $[0, \infty)$, a very similar technique to the Bernstein estimator is used. Instead of the Bernstein basis polynomials, we here use the functions

$$V_{k,m}(x) = e^{-mx} \frac{(mx)^k}{k!}.$$

The first person to do so was Mirakyan in 1941. In 1950, Szasz wrote the famous paper [36], where he expanded the Bernstein polynomials to the Szasz-Mirakyan operator

$$S_m(x) = S_m(u; x) = \sum_{k=0}^{\infty} u\left(\frac{k}{m}\right) e^{-mx} \frac{(mx)^k}{k!} = \sum_{k=0}^{\infty} u\left(\frac{k}{m}\right) V_{k,m}(x)$$

for a function u being continuous on $(0, \infty)$.

He proved the following theorem, which can be found in [36].

Theorem 6.1. *If u is a continuous function on $(0, \infty)$ with a finite limit at infinity, then as $m \to \infty$,*

$$S_m(u; x) = \sum_{k=0}^{\infty} u\left(\frac{k}{m}\right) e^{-mx} \frac{(mx)^k}{k!} \to u(x)$$

uniformly for $x \in (0, \infty)$. End Theorem

One can expand Theorem 6.1 to a function u being continuous on $[0, \infty)$ with $u(0) = 0$. Then, $S_m(u; 0) = 0$ and with the continuity it holds that $S_m(u; x) \to u(x)$ for $x \in [0, \infty)$.

Similar to Vitale, this is used to estimate a density function f in [37] with

$$\hat{f}^S_{m,n}(x) = \frac{m}{n} \sum_{m=0}^{\infty} B^{(n)}_{k,m} e^{-mx} \frac{(mx)^k}{k!},$$

where f is supported on $[0, \infty)$ and

$$B^{(n)}_{k,m} = \text{Number of } X_i \text{ in } \left[\frac{k}{m}, \frac{k+1}{m}\right), k \in \mathbb{N}_0.$$

Now, we turn our attention to the distribution function. With Theorem 6.1 we know that a distribution function F on $[0, \infty)$ can be represented by

$$S_m(F; x) = \sum_{k=0}^{\infty} F\left(\frac{k}{m}\right) e^{-mx} \frac{(mx)^k}{k!}, \tag{6.1}$$

which converges to F uniformly for $x \in [0, \infty)$. This follows from $F(0) = 0$ and the remark after the theorem. An idea to estimate this distribution function F on $[0, \infty)$ is

$$\hat{F}^S_{m,n}(x) = \sum_{k=0}^{\infty} F_n\left(\frac{k}{m}\right) e^{-mx} \frac{(mx)^k}{k!}.$$

The intuition is that we replace the unknown distribution function F in the Szasz-Mirakyan operator Eq. (6.1) by the empirical distribution function (EDF) F_n. We call this estimator $\hat{F}^S_{m,n}$ the Szasz estimator. As before, we assume that $m = m_n$ depends on n.

Note here that the sum is infinite, which is not a desirable property. The sum cannot be truncated because then the poisson probabilities do not add up to one anymore. This yields an estimator not approaching one for $x \to \infty$. However, later on we will see that the estimator can easily be rewritten, so that the sum is finite.

In order to work with this estimator we assume the following.

Assumption 6.1. *The distribution function F is continuous. The derivatives f and f' are continuous and bounded on* $[0, \infty)$. End Assumption

Note that if only the convergence itself is important and we are not interested in deriving the convergence rate, it is enough to assume these properties on $(0, \infty)$.

In the following, we state and prove some important properties of the Szasz estimator.

6.1 General Properties

Here, some important properties of the Szasz estimator $\hat{F}^S_{m,n}(x)$ are shown. The behavior at the boundary is very good as can be seen now. We know that

$$\hat{F}^S_{m,n}(0) = 0 = F(0) = S_m(F; 0) \text{ and } \lim_{x\to\infty} \hat{F}^S_{m,n}(x) = 1 = \lim_{x\to\infty} F(x) = \lim_{x\to\infty} S_m(F; x) \tag{6.2}$$

with probability one for all m. This means that bias and variance in the point $x = 0$ are zero.

To show that the limit is one, the following functions are needed. The gamma function is defined as

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx.$$

The upper and lower incomplete gamma functions are defined by

$$\Gamma(z, s) = \int_s^{\infty} x^{z-1} e^{-x} dx,$$

$$\gamma(z, s) = \int_0^{s} x^{z-1} e^{-x} dx$$

respectively. The limit is one since

$$\begin{aligned}
\hat{F}^S_{m,n}(x) &= \sum_{k=0}^{\infty} F_n\left(\frac{k}{m}\right) e^{-mx}\frac{(mx)^k}{k!} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{\infty} \mathbb{I}\{k \geq mX_i\} e^{-mx}\frac{(mx)^k}{k!} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=\lceil mX_i\rceil}^{\infty} e^{-mx}\frac{(mx)^k}{k!} \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(Y \geq \lceil mX_i\rceil) \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{\gamma(\lceil mX_i\rceil, mx)}{\Gamma(\lceil mX_i\rceil)} \xrightarrow{x\to\infty} 1,
\end{aligned}$$

where $Y \sim \text{Poi}(mx)$ is a random variable. Since the above representation only contains a finite number of summands, it can be used to simulate the estimator. Now, it is also possible to illustrate the estimator, see Figure 7.

The expectation of the Szasz operator is of course given by the expression $\mathbb{E}[F^S_{m,n}(x)] = S_m(F;x)$ for $x \in [0,\infty)$.

It holds that $\hat{F}_{m,n}(x)$ yields a proper continuous distribution function with probability one and for all values of m. The continuity of $\hat{F}_{m,n}(x)$ is obvious. Moreover, it follows from Eq. (6.2) and the next theorem that $0 \leq \hat{F}_{m,n}(x) \leq 1$ for $x \in [0,\infty)$.

Theorem 6.2. *The function $\hat{F}^S_{m,n}(x)$ is increasing in x on $[0,\infty)$.* End Theorem

Proof. This proof is similar to the one of Theorem 5.1. Let

$$g_n(0) = 0 \text{ and } g_n\left(\frac{k}{m}\right) = F_n\left(\frac{k}{m}\right) - F_n\left(\frac{k-1}{m}\right), k = 1,2,...,$$

and

$$U_k(m,x) = \sum_{j=k}^{\infty} V_{j,m}(x) = \frac{1}{\Gamma(k)}\int_0^{mx} t^{k-1}e^{-t}\,\mathrm{d}t.$$

The last equation holds because

$$U_k(m,x) = 1 - \sum_{j=0}^{k-1} V_{j,m}(x) = 1 - \frac{\Gamma(k,mx)}{\Gamma(k)} = \frac{\gamma(k,mx)}{\Gamma(k)}.$$

It follows that $\hat{F}^S_{m,n}$ can be written as

$$\hat{F}^S_{m,n}(x) = \sum_{k=0}^{\infty} g_n\left(\frac{k}{m}\right) U_k(m,x)$$

because

$$\begin{aligned}
\sum_{k=0}^{\infty} g_n\left(\frac{k}{m}\right) U_k(m,x) &= \sum_{k=1}^{\infty}\left[F_n\left(\frac{k}{m}\right) - F_n\left(\frac{k-1}{m}\right)\right]\sum_{j=k}^{\infty} V_{j,m}(x) \\
&= \sum_{k=1}^{\infty}\sum_{j=k}^{\infty} F_n\left(\frac{k}{m}\right) V_{j,m}(x) - \sum_{k=0}^{\infty}\sum_{j=k}^{\infty} F_n\left(\frac{k}{m}\right) V_{j,m}(x) + \sum_{k=0}^{\infty} F_n\left(\frac{k}{m}\right) V_{k,m}(x) \\
&= \hat{F}^S_{m,n}(x).
\end{aligned}$$

The claim follows since $g_n\left(\frac{k}{m}\right)$ is non-negative for at least one k and $U_k(m,x)$ is increasing. End Proof

The next theorem shows that $\hat{F}^S_{m,n}(x)$ is uniformly strongly consistent.

Theorem 6.3. *If F is a continuous probability distribution function on $[0,\infty)$, then*

$$\left\|\hat{F}^S_{m,n} - F\right\| \to 0 \ \ a.s.$$

for $m,n \to \infty$. We use the notation $\|G\| = \sup_{x\in[0,\infty)} |G(x)|$ for a bounded function G on $[0,\infty)$.

End Theorem

Proof. The proof follows the proof of Theorem 5.2. It holds that

$$\left\|\hat{F}^S_{m,n} - F\right\| \leq \left\|\hat{F}^S_{m,n} - S_m\right\| + \|S_m - F\|$$

and

$$\left\|\hat{F}^S_{m,n} - S_m\right\| = \|\sum_{k=0}^{\infty} [F_n(k/m) - F(k/m)]\, V_{k,m}\| \leq \|F_n - F\| \cdot \|\sum_{k=0}^{\infty} V_{k,m}\| = \|F_n - F\|.$$

We know with the Glivenko-Cantelli theorem that $\|F_n - F\| \to 0$ a.s. for $n \to \infty$ so that the claim follows with Theorem 6.1. End Proof

In the sequel, the bias and the variance of the estimator are calculated.

6.1.1 Bias and Variance

We now calculate the bias and the variance of the Szasz estimator $\hat{F}^S_{m,n}$ on the inner interval $(0,\infty)$, as we already know that bias and variance are zero for $x = 0$. In the following lemma, we first find a different expression of S_m that is similar to Lemma 5.1.

Lemma 6.1. *It holds for $x \in (0,\infty)$ that*

$$S_m(F;x) = S_m(x) = F(x) + m^{-1}b^S(x) + o_x(m^{-1}),$$

where $b^S(x) = \frac{xf'(x)}{2}$.

End Lemma

Proof. Following the proof of Lemma 5.1, Taylor's theorem gives

$$\begin{aligned} S_m(x) &= \sum_{k=0}^{\infty} F\left(\frac{k}{m}\right) V_{k,m}(x) \\ &= F(x) + \underbrace{\sum_{k=0}^{\infty} \left(\frac{k}{m} - x\right) f(x) V_{k,m}(x)}_{S_2} + \underbrace{\frac{1}{2} f'(x) \sum_{k=0}^{\infty} \left(\frac{k}{m} - x\right)^2 V_{k,m}(x)}_{S_3} \\ &\quad + \underbrace{\sum_{k=0}^{\infty} o\left(\left(\frac{k}{m} - x\right)^2\right) V_{k,m}(x)}_{S_4}. \end{aligned}$$

The second summand S_2 simplifies to $S_2 = xf(x) - xf(x) = 0$ because for $x \in [0,\infty)$, it holds that

$$\sum_{k=0}^{\infty} \frac{k}{m} V_{k,m}(x) = \frac{1}{m}\mathbb{E}[Y] = x,$$

where $Y \sim \mathrm{Poi}(mx)$.

The third part S_3 can be written as

$$\sum_{k=0}^{\infty} \left(\frac{k}{m} - x\right)^2 V_{k,m}(x) = \frac{1}{m^2} \mathrm{Var}[Y] = \frac{x}{m}. \tag{6.3}$$

For the last summand S_4 we know that

$$\begin{aligned} S_4 &= \sum_{k=0}^{\infty} o\left(\left(\frac{k}{m} - x\right)^2\right) V_{k,m}(x) \\ &= o\left(\sum_{k=0}^{\infty} \left(\frac{k}{m} - x\right)^2 V_{k,m}(x)\right) \\ &= o\left(\frac{x}{m}\right) = o_x(m^{-1}) \end{aligned}$$

with Eq. (6.3). End Proof

The following theorem establishes asymptotic expressions for the bias and the variance of the Szasz estimator $\hat{F}^S_{m,n}$ as $m, n \to \infty$ are established. The statement is similar to Theorem 5.3.

Theorem 6.4. *For each $x \in (0, \infty)$, the bias has the representation*

$$\begin{aligned} \mathrm{Bias}\left[\hat{F}^S_{m,n}(x)\right] &= \mathbb{E}\left[\hat{F}_{m,n}\right] - F(x) = m^{-1} \frac{x f'(x)}{2} + o_x(m^{-1}) \\ &= m^{-1} b^S(x) + o_x(m^{-1}). \end{aligned}$$

For the variance it holds that

$$\mathrm{Var}\left[\hat{F}^S_{m,n}(x)\right] = n^{-1}\sigma^2(x) - m^{-1/2} n^{-1} V^S(x) + o_x(m^{-1/2} n^{-1}),$$

where

$$\sigma^2(x) = F(x)(1 - F(x)), \; V^S(x) = f(x) \left[\frac{2x}{\pi}\right]^{1/2}$$

and $b^S(x)$ is defined as in Lemma 6.1. End Theorem

For the proof, see Proofs Szasz.

In the following, we talk about the asymptotic behavior of the Szasz estimator.

6.2 Asymptotic Behavior

Here, we turn our attention to the asymptotic behavior of the Szasz estimator. The next theorem is similar to Theorem 5.4 and shows the asymptotic normality of this estimator.

Theorem 6.5. *Let $x \in (0, \infty)$, such that $0 < F(x) < 1$. Then, for $m, n \to \infty$ it holds that*

$$n^{1/2}\left(\hat{F}^S_{m,n}(x) - \mathbb{E}[\hat{F}^S_{m,n}(x)]\right) = n^{1/2}\left(\hat{F}^S_{m,n}(x) - S_m(x)\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right),$$

where $\sigma^2(x) = F(x)(1 - F(x))$. End Theorem

The idea for the proof is to use the central limit theorem for double arrays, see Proofs Szasz for more details.

Note that as in the settings before, this result holds for all choices of m with $m \to \infty$ without any restrictions.

We now take a closer look at the asymptotic behavior of $\hat{F}^S_{m,n}(x) - F(x)$, where the behavior of m is restricted. With Lemma 6.1, it is easy to see that

$$n^{1/2}\left(\hat{F}^S_{m,n}(x) - F(x)\right) = n^{1/2}\left(\hat{F}^S_{m,n}(x) - S_m(x)\right) + m^{-1}n^{1/2}b^S(x) + o_x(m^{-1}n^{1/2}). \tag{6.4}$$

This leads directly to the following corollary, which is similar Corollary 5.1 but on $(0, \infty)$.

Corollary 6.1. *Let $m, n \to \infty$. Then, for $x \in (0, \infty)$ with $0 < F(x) < 1$, it holds that*

(a) if $mn^{-1/2} \to \infty$, then

$$n^{1/2}\left(\hat{F}^S_{m,n}(x) - F(x)\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right),$$

(b) if $mn^{-1/2} \to c$, where c is a positive constant, then

$$n^{1/2}\left(\hat{F}^S_{m,n}(x) - F(x)\right) \xrightarrow{D} \mathcal{N}\left(c^{-1}b^S(x), \sigma^2(x)\right), \tag{6.5}$$

where $\sigma^2(x)$ and $b^S(x)$ are defined as in Theorem 6.4. End Corollary

Next, we derive the asymptotically optimal m with respect to MSE.

6.3 Asymptotically Optimal m with Respect to MSE

For the estimator $\hat{F}^S_{m,n}$, it is also interesting to calculate the MSE

$$\text{MSE}\left[\hat{F}^S_{m,n}(x)\right] = \mathbb{E}\left[\left(\hat{F}^S_{m,n}(x) - F(x)\right)^2\right]$$

and the asymptotically optimal m with respect to MSE. In the point $x = 0$, the MSE is zero. For $(0, \infty)$, the next theorem shows the MSE.

Theorem 6.6. *The MSE of the Szasz estimator is of the form*

$$\begin{aligned}\text{MSE}\left[\hat{F}^S_{m,n}(x)\right] &= \text{Var}\left[\hat{F}^S_{m,n}(x)\right] + \text{Bias}\left[\hat{F}^S_{m,n}(x)\right]^2 \\ &= n^{-1}\sigma^2(x) - m^{-1/2}n^{-1}V^S(x) + m^{-2}\left(b^S(x)\right)^2 + o_x(m^{-2}) + o_x(m^{-1/2}n^{-1})\end{aligned} \tag{6.6}$$

for $x \in (0, \infty)$. End Theorem

Proof. This follows directly from Theorem 6.4. End Proof

To calculate the optimal m with respect to the MSE, one has to take the derivative with respect to m of the above equation and set it to zero. This can be done analogously to Eq. (5.10). The next corollary, which is similar to Corollary 5.2, follows.

Corollary 6.2. *Assuming that $f(x) \neq 0$ and $f'(x) \neq 0$, the asymptotically optimal choice of m for estimating $F(x)$ with respect to MSE is*

$$m_{opt} = n^{2/3}\left[\frac{4(b^S(x))^2}{V^S(x)}\right]^{2/3}. \tag{6.7}$$

Therefore, the associated MSE can be written as

$$\mathrm{MSE}\left[\hat{F}^S_{m_{opt},n}(x)\right] = n^{-1}\sigma^2(x) - \underbrace{\frac{3}{4}n^{-4/3}\left[\frac{(V^S(x))^4}{4(b^S(x))^2}\right]^{1/3}}_{S_2} + o_x(n^{-4/3}) \tag{6.8}$$

for $x \in (0,\infty)$, where $\sigma^2(x), b^S(x)$, and $V^S(x)$ are defined as in Theorem 6.4. End Corollary

In [37], it is stated that the optimal m to estimate the density function with respect to the MSE is $O(n^{2/5})$. We just established that for the distribution function, $m \in O(n^{2/3})$. The same phenomenon as in Section 5.5 can be observed here. When using $m \in O(n^{2/5})$ for the distribution estimation, it lies outside of any confidence band of F. This holds because of the fact that from $mn^{-2/5} \to c$ it follows that $mn^{-1/2} \to 0$. Together with $f'(x) \neq 0$ and Eq. (6.4), it holds that

$$\mathbb{P}\left(n^{1/2}\left|\hat{F}^S_{m,n}(x) - F(x)\right| > \epsilon\right) \to 1$$

for all $\epsilon > 0$. This shows that for this choice of m, $\hat{F}^S_{m,n}(x)$ does not converge to a limiting distribution centred at $F(x)$ with proper rescaling. Therefore, $\hat{F}^S_{m,n}$ lies outside of any confidence band based on F_n with probability going to one.

As before, we now take a closer look at the asymptotically optimal m with respect to MISE.

6.4 Asymptotically Optimal m with Respect to MISE

We now focus on the MISE, using a different definition than before. As we deal with an infinite integral, we use a non-negative weight function ω. Here, the weight function is chosen as $\omega(x) = e^{-ax}f(x)$. Following [27], the MISE is defined as

$$\mathrm{MISE}\left[\hat{F}^S_{m,n}\right] = \mathbb{E}\left[\int_0^\infty \left(\hat{F}^S_{m,n}(x) - F(x)\right)^2 e^{-ax}f(x)\,\mathrm{d}x\right].$$

Note that $\mathrm{MISE}\left[\hat{F}^S_{m,n}\right]$ cannot be calculated by integrating the expression of $\mathrm{MSE}\left[\hat{F}^S_{m,n}\right]$ obtained in Eq. (6.6) as the asymptotic expressions depend on x.

The next theorem gives the MISE of the Szasz operator and is similar to Theorem 5.6.

Theorem 6.7. *It holds that*

$$\mathrm{MISE}\left[\hat{F}^S_{m,n}\right] = n^{-1}C^S_1 - m^{-1/2}n^{-1}C^S_2 + m^{-2}C^S_3 + o(m^{-1/2}n^{-1}) + o(m^{-2})$$

with

$$C^S_1 = \int_0^\infty \sigma^2 e^{-ax}f(x)\,\mathrm{d}x \ , \ C^S_2 = \int_0^\infty V^S(x)e^{-ax}f(x)\,\mathrm{d}x \ \text{ and } \ C^S_3 = \int_0^\infty (b^S(x))^2 e^{-ax}f(x)\,\mathrm{d}x.$$

The definitions of $\sigma^2(x), b^S(x)$, and $V^S(x)$ can be found in Theorem 6.4. End Theorem

For the proof, see Proofs Szasz.

Very similar to Corollary 5.3, the next corollary gives the asymptotically optimal m for estimating F with respect to MISE.

Corollary 6.3. *It follows that the asymptotically optimal m for estimating F with respect to MISE is*

$$m_{opt} = n^{2/3}\left[\frac{4C_3^S}{C_2^S}\right]^{2/3},$$

which leads to

$$\mathrm{MISE}\left[\hat{F}^S_{m_{opt},n}\right] = n^{-1}C_1^S \underbrace{- \frac{3}{4}n^{-4/3}\left[\frac{(C_2^S)^4}{4C_3^S}\right]^{1/3}}_{S_2} + o(n^{-4/3}), \tag{6.9}$$

the optimal MISE. End Corollary

If we compare the optimal MSE and optimal MISE of the Szasz estimator with those of the EDF, we observe the same behavior as for the Bernstein estimator. The second summand S_2 in Eq. (6.8) and Eq. (6.9) is always positive so that the Szasz estimator seems to outperform the EDF. This is proven in the following.

6.5 Deficiency

We now measure the local and global performance of the Szasz estimator with the help of the deficiency. Let

$$i_L^S(n,x) = \min\left\{k \in \mathbb{N} : \mathrm{MSE}[F_k(x)] \le \mathrm{MSE}\left[\hat{F}^S_{m,n}(x)\right]\right\}, \text{ and}$$
$$i_G^S(n) = \min\left\{k \in \mathbb{N} : \mathrm{MISE}[F_k] \le \mathrm{MISE}\left[\hat{F}^S_{m,n}\right]\right\}$$

be the local and global numbers of observations that F_n needs to perform at least as well as $\hat{F}^S_{m,n}$. The next theorem deals with these quantities and is similar to Theorem 5.7.

Theorem 6.8. *Let $x \in (0,\infty)$ and $m, n \to \infty$. If $mn^{-1/2} \to \infty$ it holds that*

$$i_L^S(n,x) = n[1 + o_x(1)] \text{ and } i_G^S(n) = n[1 + o(1)].$$

In addition, the following statements are true.

(a) If $mn^{-2/3} \to \infty$ and $mn^{-2} \to 0$, then

$$i_L^S(n,x) - n = m^{-1/2}n[\theta^S(x) + o_x(1)], \text{ and}$$
$$i_G^S(n) - n = m^{-1/2}n[C_2^S/C_1^S + o(1)].$$

(b) If $mn^{-2/3} \to c$, where c is a positive constant, then

$$i_L^S(n,x) - n = n^{2/3}[c^{-1/2}\theta^S(x) - c^{-2}\gamma^S(x) + o_x(1)], \text{ and}$$
$$i_G^S(n) - n = n^{2/3}[c^{-1/2}C_2^S/C_1^S - c^{-2}C_3^S/C_1^S + o(1)],$$

where

$$\theta^S(x) = \frac{V^S(x)}{\sigma^2(x)} \text{ and } \gamma^S(x) = \frac{(b^S(x))^2}{\sigma^2(x)}.$$

Here, $V^S(x), \sigma^2(x)$, and $b^S(x)$ are defined as in Theorem 6.4 and C_1^S, C_2^S, and C_3^S are defined as in Theorem 6.7. End Theorem

For the proof, see Proofs Szasz.

This theorem shows under which conditions the Szasz estimator outperforms the EDF. The asymptotic deficiency goes to infinity as n grows. This means that for increasing n, the number of extra observations has to increase so that the EDF outperforms the Szasz estimator. Hence, the EDF is asymptotically deficient to the Szasz estimator.

It is again possible to maximize the deficiency to get an optimal m. The following statement is similar to Lemma 5.2.

Lemma 6.2. *The optimal m with respect to the global deficiency in the case $mn^{-2/3} \to c$ is of the same order as in Corollary 6.3.* End Lemma

Proof. The proof is the same as for Lemma 5.2. In the case $mn^{-2/3} \to c$, the deficiency $i_G(n) - n$ is asymptotically positive only when

$$c > \left[\frac{C_3^S}{C_2^S}\right]^{2/3} = c^*.$$

Then, the optimal c maximizing $g(c) = c^{-1/2}C_2^S/C_1^S - c^{-2}C_3^S/C_1^S$ is

$$c_{opt} = \left[\frac{4C_3^S}{C_2^S}\right]^{2/3} = 2^{4/3}c^*.$$

Hence, the optimal order of the Szasz estimator with respect to the deficiency satisfies

$$m_{opt}n^{-2/3} \to c_{opt} \Leftrightarrow m_{opt} = n^{2/3}[c_{opt} + o(1)],$$

which shows the claim. End Proof

6.6 Properties of $V_{k,m}$

We now present a few properties of $V_{k,m}$ that are needed for the proofs. The following lemma is similar to Lemma 5.3.

Lemma 6.3. *Define*

$$L_m^S(x) = \sum_{k=0}^{\infty} V_{k,m}^2(x)$$

and

$$R_{j,m}^S(x) = m^{-j} \sum\sum_{0 \le k < l \le \infty} (k - mx)^j V_{k,m}(x) V_{l,m}(x)$$

for $j \in \{0, 1, 2\}$, *and* $V_{k,m}(x) = e^{-mx}\frac{(mx)^k}{k!}$. *It trivially holds that* $0 \le L_m^S(x) \le 1$ *for* $x \in [0, \infty)$. *In addition, the following properties hold. Let* g *be a continuous and bounded function on* $[0, \infty)$. *This leads to*

(a) $L_m^S(0) = 1$ *and* $\lim_{x\to\infty} L_m^S(x) = 0$,

(b) $R_{j,m}^S(0) = 0$ *for* $j \in \{0, 1, 2\}$,

(c) $0 \le R_{2,m}^S(x) \le \frac{x}{m}$ *for* $x \in (0, \infty)$,

(d) $L_m^S(x) = m^{-1/2}\left[(4\pi x)^{-1/2} + o_x(1)\right]$ *for* $x \in (0, \infty)$,

(e) $R_{1,m}^S(x) = m^{-1/2}\left[-\frac{\sqrt{x}}{\sqrt{2\pi}} + o_x(1)\right]$ *for* $x \in (0, \infty)$,

(f) $m^{1/2}\int_0^\infty L_m^S(x)e^{-ax}\,\mathrm{d}x = \int_0^\infty (4\pi x)^{-1/2}e^{-ax}\,\mathrm{d}x + o(1) = \frac{1}{2\sqrt{a}} + o(1)$ *for* $a \in (0, \infty)$,

(g) $m^{1/2}\int_0^\infty xL_m^S(x)e^{-ax}\,\mathrm{d}x = \int_0^\infty x^{1/2}(4\pi)^{-1/2}e^{-ax}\,\mathrm{d}x + o(1) = \frac{1}{4a^{3/2}} + o(1)$ *for* $a \in (0, \infty)$,

(h) $m^{1/2}\int_0^\infty g(x)R_{1,m}^S(x)e^{-ax}\,\mathrm{d}x = -\int_0^\infty g(x)\frac{\sqrt{x}}{\sqrt{2\pi}}e^{-ax}\,\mathrm{d}x + o(1)$ *for* $a \in (0, \infty)$.

End Lemma

Proof. The proof is similar to the proof of Lemma 5.3.

(a) $L_m^S(0) = 1$ is clear. Using the mode of the poisson distribution it holds for the limit that

$$\lim_{x\to\infty} L_m^S(x) \le \lim_{x\to\infty} \max_k V_{k,m} \sum_{k=0}^{\infty} V_{k,m} = \lim_{x\to\infty} P(Y = \lfloor mx \rfloor)) = 0,$$

where $Y \sim \mathrm{Poi}(mx)$.

(b) $R_{j,m}^S(0) = 0$ holds trivially.

(c) The non-negativity is clear. For the other inequality, it holds that

$$\begin{aligned} R_{2,m}^S(x) &\le m^{-2}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(k - mx)^2 V_{k,m}(x)V_{l,m}(x) \\ &= m^{-2}\sum_{k=0}^{\infty}(k - mx)^2 V_{k,m}(x) = m^{-2}\,\mathrm{Var}[Y] = \frac{x}{m}, \end{aligned}$$

where $Y \sim \mathrm{Poi}(mx)$.

(d) Let $U_i, W_j, i, j \in \{1, ..., m\}$, be i.i.d. random variables with distribution Poi(x), hence,

$$\mathbb{P}(U_1 = k) = e^{-x}\frac{x^k}{k!}.$$

Define

$$R_i = \frac{U_i - W_i}{\sqrt{2x}}.$$

Then, we know that $\mathbb{E}[R_i] = 0, \mathrm{Var}[R_i] = 1$ and R_i has a lattice distribution with span

$$h = \frac{1}{\sqrt{2x}}.$$

Note that with the independence it holds that

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{m} R_i = 0\right) &= \mathbb{P}\left(\sum_{i=1}^{m} U_i = \sum_{i=1}^{m} W_i\right)\\ &= \mathbb{P}\left(\sum_{i=1}^{m} U_i = \sum_{i=1}^{m} W_i = 0\right) + \mathbb{P}\left(\sum_{i=1}^{m} U_i = \sum_{i=1}^{m} W_i = 1\right) + ...\\ &= \sum_{k=0}^{\infty} V_{k,m}^2(x).\end{aligned}$$

With [29, XV.5, p. 517, Theorem 3], we get that

$$\frac{\sqrt{m}}{h}\sum_{k=0}^{\infty} V_{k,m}^2(x) - \frac{1}{\sqrt{2\pi}} \to 0$$

and it follows that

$$\sqrt{4\pi m x}\sum_{k=0}^{\infty} V_{k,m}^2(x) \to 1,$$

from which the claim follows.

(e) With [38, Eq. (2.2), Example 2], we know that for a random variable $Y \sim$ Poi(mx),

$$\sum_{l=k+1}^{\infty} V_{l,m}(z) = 1 - \mathbb{P}(Y \le k) = 1 - \Phi(z + \Delta_2(k)) + O((mx)^{-3/2}),$$

where $z = (k - mx)/\sqrt{mx}$ and

$$\Delta_2(k) = (mx)^{-1/2}\frac{1}{6}(4 - z^2) + (mx)^{-1}\frac{1}{72}z(5z^2 - 14).$$

Expandig $\Phi(t)$ about $t = 0$ leads to

$$\Phi(t) = \frac{1}{2} + \frac{t}{\sqrt{2\pi}} + o(|t|)$$

and therefore

$$\sum_{l=k+1}^{\infty} V_{l,m}(z) = \frac{1}{2} - \frac{k - mx}{\sqrt{2\pi m x}} - \frac{\Delta_2(k)}{\sqrt{2\pi}} + o_x\left(\left|\frac{k - mx}{\sqrt{mx}} + \Delta_2(k)\right|\right) + O\left((mx)^{-3/2}\right).$$

It holds that

$$\Delta_2(k) = O_x(m^{1/2})$$

and hence,

$$\begin{aligned}
R_{1,m}^S(x) &= m^{-1}\sum_{k=0}^{\infty}(k-mx)V_{k,m}(x)\left[\sum_{l=k+1}^{\infty}V_{l,m}(x)\right] \\
&= m^{-1}\sum_{k=0}^{\infty}(k-mx)V_{k,m}(x)\left[\frac{1}{2}-\frac{k-mx}{\sqrt{2\pi mx}}+O_x(m^{1/2})\right. \\
&\quad \left. +o_x\left(\left|\frac{k-mx}{\sqrt{2\pi mx}}+O_x(m^{1/2})\right|\right)+O(m^{-3/2})\right] \\
&= -m^{-1}\sum_{k=0}^{\infty}(k-mx)^2\frac{V_{k,m}(x)}{\sqrt{2\pi mx}} \\
&\quad + m^{-1}\sum_{k=0}^{\infty}(k-mx)V_{k,m}(x)\left[o_x\left(\left|\frac{k-mx}{\sqrt{2\pi mx}}+O_x(m^{1/2})\right|\right)\right] \\
&= -\frac{mx}{m\sqrt{2\pi mx}}+o_x\left(\frac{1}{m\sqrt{2\pi mx}}\sum_{k=0}^{\infty}(k-mx)^2V_{k,m}(x)\right) \\
&= m^{-1/2}\left(-\frac{\sqrt{x}}{\sqrt{2\pi}}+o_x(1)\right),
\end{aligned}$$

where we used that $\sum_{k=0}^{\infty}(k-mx)V_{k,m}(x)=0$.

(f) The goal is to calculate

$$m^{1/2}\int_0^{\infty}L_m^S(x)e^{-ax}\,\mathrm{d}x = m^{1/2}\int_0^{\infty}\sum_{k=0}^{\infty}V_{k,m}^2(x)e^{-ax}\,\mathrm{d}x = m^{1/2}\sum_{k=0}^{\infty}\int_0^{\infty}V_{k,m}^2(x)e^{-ax}\,\mathrm{d}x.$$

For the integral we know that

$$\begin{aligned}
\int_0^{\infty}V_{k,m}^2(x)e^{-ax}\,\mathrm{d}x &= \int_0^{\infty}\left(e^{-mx}\frac{(mx)^k}{k!}\right)^2e^{-ax}\,\mathrm{d}x \\
&= \frac{m^{2k}}{(k!)^2}\int_0^{\infty}x^{2k}e^{-(2m+a)x}\,\mathrm{d}x \\
&= \frac{m^{2k}}{(k!)^2(2m+a)^{2k+1}}\int_0^{\infty}y^{2k}e^{-y}dy = \frac{m^{2k}}{(k!)^2(2m+a)^{2k+1}}\Gamma(2k+1).
\end{aligned}$$

Calculating the sum leads to

$$\begin{aligned}
m^{1/2}\sum_{k=0}^{\infty}\frac{m^{2k}}{(k!)^2(2m+a)^{2k+1}}\Gamma(2k+1) &= m^{1/2}\sum_{k=0}^{\infty}\left(\frac{1}{2m+a}\right)^{2k+1}m^{2k}\binom{2k}{k} \\
&= \sqrt{\frac{m}{a(a+4m)}} \\
&= \frac{1}{2\sqrt{a}}+\frac{1}{\sqrt{a}}\left[\sqrt{\frac{m}{a+4m}}-\frac{1}{2}\right] = \frac{1}{2\sqrt{a}}+o(1).
\end{aligned}$$

It holds that

$$\int_0^{\infty}(4\pi x)^{-1/2}e^{-ax}\,\mathrm{d}x = \frac{1}{2\sqrt{a}}$$

and hence,

$$m^{1/2}\int_0^\infty L_m^S(x)e^{-ax}\,\mathrm{d}x = \frac{1}{2\sqrt{a}} + o(1) = \int_0^\infty (4\pi x)^{-1/2}e^{-ax}\,\mathrm{d}x + o(1).$$

(g) Similar to (f), we get

$$\int_0^\infty xV_{k,m}^2(x)e^{-ax}\,\mathrm{d}x = \frac{m^{2k}}{(k!)^2(2m+a)^{2k+2}}\Gamma(2k+2),$$

leading to

$$\begin{aligned} m^{1/2}\sum_{k=0}^\infty \frac{m^{2k}}{(k!)^2(2m+a)^{2k+2}}\Gamma(2k+2) &= \frac{\sqrt{m}(a+2m)}{(a(a+4m))^{3/2}} \\ &= \frac{1}{4a^{3/2}} + \frac{1}{a^{3/2}}\left[\frac{\sqrt{m}(a+2m)}{(a+4m)^{3/2}} - \frac{1}{4}\right] \\ &= \frac{1}{4a^{3/2}} + o(1). \end{aligned}$$

(h) Define $G_m^S(x) = m^{1/2}R_{1,m}^S(x)e^{-ax}$ and $G^S(x) = -\frac{\sqrt{x}}{\sqrt{2\pi}}e^{-ax}$. Then with part (e) we know that $G_m^S(x) \xrightarrow{m\to\infty} G^S(x)$.

Note that

$$\begin{aligned} R_{1,m}(x) &= m^{-1}e^{-2mx}\sum_{0\le k<l\le\infty}\sum (k-mx)\frac{(mx)^{k+l}}{k!l!} \\ &= m^{-1}e^{-2mx}\sum_{k=0}^\infty (k-mx)\frac{(mx)^k}{k!}\left[\sum_{l=k+1}^\infty \frac{(mx)^l}{l!}\right] \\ &= m^{-1}e^{-mx}\sum_{k=0}^\infty (k-mx)\frac{(mx)^k}{k!}\left(1-\frac{\Gamma(1+k,mx)}{\Gamma(1+k)}\right) \\ &= -m^{-1}e^{-mx}\sum_{k=0}^\infty (k-mx)\frac{(mx)^k}{k!}\frac{\Gamma(1+k,mx)}{\Gamma(1+k)}. \end{aligned}$$

Using $\frac{\Gamma(1+k,mx)}{\Gamma(1+k)} = \mathbb{P}(Y\le k)\in[0,1]$ for $Y\sim \mathrm{Poi}(mx)$, the above calculation yields

$$\begin{aligned} |G_m^S(x)| &\le m^{-1/2}e^{-(a+m)x}\sum_{k=0}^\infty |k-mx|\frac{(mx)^k}{k!}\frac{\Gamma(1+k,mx)}{\Gamma(1+k)} \\ &\le m^{-1/2}e^{-ax}\sum_{k=0}^\infty |k-mx|V_{k,m}(x) \\ &\le m^{-1/2}e^{-ax}\left(\sum_{k=0}^\infty (k-mx)^2V_{k,m}(x)\right)^{1/2} \\ &= m^{-1/2}e^{-ax}\sqrt{mx} = \sqrt{x}e^{-ax}. \end{aligned}$$

This is integrable since

$$\int_0^\infty \sqrt{x}e^{-ax}\,\mathrm{d}x = \frac{\sqrt{\pi}}{2a^{3/2}}.$$

With the dominated convergence theorem it follows that

$$\int_0^\infty |G_m^S(x) - G^S(x)| \,\mathrm{d}x = o(1)$$

and

$$\left| \int_0^\infty g(x) G_m^S(x) \,\mathrm{d}x - \int_0^\infty g(x) G^S(x) \,\mathrm{d}x \right| \leq \sup_{x \in [0,\infty)} |g(x)| \int_0^\infty |G_m^S(x) - G^S(x)| \,\mathrm{d}x = o(1),$$

as g is bounded.

End Proof

6.7 Proofs Szasz

Proof of Theorem 6.4. This proof is similar to that of Theorem 5.3. The bias follows directly from Lemma 6.1. For the proof of the variance, let

$$Y_{i,m}^S = \sum_{k=0}^{\infty} \Delta_i \left(\frac{k}{m}\right) V_{k,m}(x),$$

where

$$\Delta_i(x) = \mathbb{I}(X_i \le x) - F(x)$$

for $x \in [0, \infty)$. We know that $\Delta_1, ..., \Delta_n$ are i.i.d. with mean zero. Hence,

$$\begin{aligned}\hat{F}_{m,n}^S(x) - S_m(x) &= \sum_{k=0}^{\infty} \left[F_n\left(\frac{k}{m}\right) - F\left(\frac{k}{m}\right)\right] V_{k,m}(x) \\ &= \frac{1}{n}\sum_{k=0}^{\infty}\sum_{i=1}^{n} \left[\mathbb{I}\left(X_i \le \frac{k}{m}\right) - F\left(\frac{k}{m}\right)\right] V_{k,m}(x) \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_{i,m}^S. \end{aligned} \tag{6.10}$$

Note that $Y_{i,m}^S < \infty$ a.s. and that, for given m, $Y_{1,m}^S, ..., Y_{n,m}^S$ are i.i.d. with mean zero. It follows with the same arguments as in the proof of Theorem 5.3 that

$$\operatorname{Var}\left[\hat{F}_{m,n}^S(x)\right] = \frac{1}{n}\mathbb{E}[(Y_{1,m}^S)^2] \tag{6.11}$$

and for $x, y \in [0, \infty)$ that

$$\mathbb{E}[\Delta_1(x)\Delta_1(y)] = \min(F(x), F(y)) - F(x)F(y),$$

from which we get that

$$\mathbb{E}[(Y_{1,m}^S)^2] = \sum_{k=0}^{\infty} F\left(\frac{k}{m}\right) V_{k,m}^2(x) + 2\sum\sum_{0\le k<l\le\infty} F\left(\frac{k}{m}\right) V_{k,m}(x)V_{l,m}(x) - S_m^2(x). \tag{6.12}$$

We want to find an asymptotic expression for the previous expression. For the first part of Eq. (6.12), write

$$F\left(\frac{k}{m}\right) = F(x) + O\left(\left|\frac{k}{m} - x\right|\right)$$

from which follows that

$$\sum_{k=0}^{\infty} F\left(\frac{k}{m}\right) V_{k,m}^2(x) = F(x)L_m^S(x) + O(I_m^S(x)), \tag{6.13}$$

where

$$I_m^S(x) = \sum_{k=0}^{\infty} \left|\frac{k}{m} - x\right| V_{k,m}^2(x)$$

and L_m^S is defined as in Lemma 6.3. For the second term of Eq. (6.12), use Taylor's theorem to get

$$F\left(\frac{k}{m}\right) = F(x) + \left(\frac{k}{m} - x\right) f(x) + O\left(\left(\frac{k}{m} - x\right)^2\right).$$

We know that

$$R_{0,m}^S = \frac{1}{2}[1 - L_m^S(x)]$$

because

$$1 = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} V_{k,m}(x)V_{l,m}(x) = 2\sum\sum_{0\le k<l\le\infty} V_{k,m}(x)V_{l,m}(x) + \sum_{k=0}^{\infty} V_{k,m}^2(x) = 2R_{0,m}^S(x) + L_m^S(x).$$

Now, Lemma 6.3, which states that $R_{2,m}^S(x) \le \frac{x}{m}$, leads to

$$\begin{aligned}\sum\sum_{0\le k<l\le\infty} F\left(\frac{k}{m}\right) V_{k,m}(x)V_{l,m}(x) &= F(x)R_{0,m}^S(x) + f(x)R_{1,m}^S(x) + O\left(R_{2,m}^S(x)\right)\\ &= \frac{1}{2}F(x)(1 - L_m^S(x)) + f(x)R_{1,m}^S(x) + O_x(m^{-1}). \qquad (6.14)\end{aligned}$$

With Lemma 6.1, we get that

$$\begin{aligned}F(x) - S_m^2(x) &= F(x) - \left(F(x) + m^{-1}\frac{xf'(x)}{2} + O_x(m^{-1})\right)^2\\ &= F(x) - \left(F(x) + O_x(m^{-1})\right)^2\\ &= F(x) - F(x)^2 - 2F(x)O_x(m^{-1}) - O_x(m^{-1})^2\\ &= \sigma^2(x) + O_x(m^{-1})\end{aligned}$$

and using Eq. (6.13) and Eq. (6.14) for Eq. (6.12), it holds that

$$\begin{aligned}\mathbb{E}[(Y_{1,m}^S)^2] &= F(x)L_m^S(x) + O(I_m^S(x)) + F(x)(1 - L_m^S(x)) + 2f(x)R_{1,m}^S(x) + O_x(m^{-1}) - S_m^2(x)\\ &= \sigma^2(x) + 2f(x)R_{1,m}^S(x) + O(I_m^S(x)) + O_x(m^{-1}). \qquad (6.15)\end{aligned}$$

Note that Cauchy-Schwarz can be applied here so that with Lemma 6.3 (d), we get that

$$\begin{aligned}I_m^S(x) &\le \left[\sum_{k=0}^{\infty}\left(\frac{k}{m} - x\right)^2 V_{k,m}(x)\right]^{1/2}\left[\sum_{k=0}^{m} V_{k,m}^3(x)\right]^{1/2}\\ &\le \left[\frac{T_{2,m}^S}{m^2}L_m^S(x)\right]^{1/2}\\ &\le \left[\frac{x}{m}L_m^S(x)\right]^{1/2}\\ &\le \left[\frac{x}{m}m^{-1/2}\left[(4\pi x)^{-1/2} + o_x(1)\right]\right]^{1/2}\\ &= o_x(m^{-3/4}), \qquad (6.16)\end{aligned}$$

where $T_{2,m}^S(x) = \sum_{k=0}^{\infty}(k - mx)^2 V_{k,m}(x) = mx$ for $x \in [0,\infty)$. Now with Lemma 6.3 (e), it holds that

$$\begin{aligned}\mathbb{E}[(Y_{1,m}^S)^2] &= \sigma^2(x) + 2f(x)R_{1,m}^S(x) + O(I_m^S(x)) + O_x(m^{-1})\\ &= \sigma^2(x) + 2f(x)m^{-1/2}\left[-\frac{\sqrt{x}}{\sqrt{2\pi}} + o_x(1)\right] + O(I_m^S(x)) + O_x(m^{-1})\\ &= \sigma^2(x) - m^{-1/2}\frac{\sqrt{2x}f(x)}{\sqrt{\pi}} + o_x(m^{-1/2}). \qquad (6.17)\end{aligned}$$

With Eq. (6.11), the desired claim

$$\operatorname{Var}\left[\hat{F}^S_{m,n}(x)\right] = n^{-1}\sigma^2(x) - m^{-1/2}n^{-1}\frac{\sqrt{2x}f(x)}{\sqrt{\pi}} + o_x(m^{-1/2}n^{-1}) \tag{6.18}$$

holds. End Proof

Proof of Theorem 6.5. This proof follows the proof of Theorem 5.4. For fixed m we know from the proof of Theorem 5.3 that

$$\hat{F}^S_{m,n}(x) - S_m(x) = \frac{1}{n}\sum_{i=1}^{n} Y^S_{i,m},$$

where the $Y^S_{i,m}$ are i.i.d. random variables with mean 0. Define $(\gamma^S_m)^2 = \mathbb{E}[(Y^S_{1,m})^2]$. We now use the central limit theorem for double arrays (see [33], Section 1.9.3) to show the claim. In the proof of Theorem 5.4, it was already explained why it is enough to show

$$\frac{\mathbb{E}[\mathbb{I}(|Y^S_{1,m}| > \epsilon\sqrt{n}\gamma^S_m)(Y^S_{1,m})^2]}{(\gamma^S_m)^2} \to 0 \text{ for } n \to \infty \text{ and all } \epsilon > 0,$$

because with Eq. (6.17), it holds that $\gamma^s_m \to \sigma(x)$.

We know that

$$|Y^S_{1,m}| \leq 1$$

which leads with

$$\mathbb{I}\left(|Y^S_{1,m}| > \epsilon\sqrt{n}\gamma^S_m\right) \leq \mathbb{I}(1 > \epsilon\sqrt{n}\gamma^S_m) \to 0$$

to the desired result. End Proof

Proof of Theorem 6.7. This proof follows the proof of Theorem 5.6. With a part of Eq. (6.16), Jensen's inequality for expected values, and Lemma 6.3 (g) leads to

$$\begin{aligned}
\int_0^\infty I^S_m(x)e^{-ax}f(x)\,\mathrm{d}x &\leq \left[\frac{1}{m}\int_0^\infty xL^S_m(x)e^{-2ax}f(x)\,\mathrm{d}x\right]^{1/2} \\
&\leq \left[\frac{\|f\|}{m}\int_0^\infty xL^S_m(x)e^{-2ax}\,\mathrm{d}x\right]^{1/2} \\
&= \left[\frac{\|f\|}{m^{3/2}}\left(\frac{1}{4(2a)^{3/2}} + o(1)\right)\right]^{1/2} = O(m^{-3/4}).
\end{aligned}$$

Using Eq. (6.10),Lemma 6.1, and Eq. (6.15) leads to

$$\begin{aligned}
\operatorname{MISE}\left[\hat{F}^S_{m,n}\right] &= \int_0^\infty \left[\operatorname{Var}\left[\hat{F}^S_{m,n}(x)\right] + \operatorname{Bias}\left[\hat{F}^S_{m,n}(x)\right]^2\right] e^{-ax}f(x)\,\mathrm{d}x \\
&= \frac{1}{n}\int_0^\infty \left[\sigma^2(x) + 2f(x)R^S_{1,m}(x) + O(I^S_m(x)) + O\left(\frac{x}{m}\right)\right] e^{-ax}f(x)\,\mathrm{d}x \\
&+ \int_0^\infty \left[m^{-1}b^S(x) + o\left(\frac{x}{m}\right)\right]^2 e^{-ax}f(x)\,\mathrm{d}x \\
&= \frac{1}{n}\int_0^\infty \left[\sigma^2(x) + 2f(x)R^S_{1,m}(x)\right] e^{-ax}f(x)\,\mathrm{d}x + \int_0^\infty m^{-2}(b^S(x))^2e^{-ax}f(x)\,\mathrm{d}x \\
&+ O(m^{-3/4}n^{-1}) + o(m^{-2}).
\end{aligned}$$

Now, with $2f(x)\frac{\sqrt{x}}{\sqrt{2\pi}} = V^S(x)$ and Lemma 6.3 (h), we get

$$\text{MISE}\left[\hat{F}^S_{m,n}\right] = n^{-1}C_1^S - n^{-1}m^{-1/2}C_2^S + m^{-2}C_3^S + o(m^{-2}) + o(m^{-1/2}n^{-1}).$$

The integrals C_i^S exist for $i = 1, 2, 3$ because f and $(f')^2$ are positive and bounded on $[0, \infty)$. It follows that

$$C_1^S = \int_0^\infty F(x)(1-F(x))e^{-ax}f(x)\,\mathrm{d}x \leq \|f\| \int_0^\infty e^{-ax}\,\mathrm{d}x = \frac{\|f\|}{a} < \infty,$$

$$C_2^S = \int_0^\infty f(x)\left[\frac{2x}{\pi}\right]^{1/2} e^{-ax}f(x)\,\mathrm{d}x \leq \frac{\sqrt{2}\|f\|^2}{\sqrt{\pi}} \int_0^\infty \sqrt{x}e^{-ax}\,\mathrm{d}x = \frac{\|f\|^2}{\sqrt{2}a^{3/2}} < \infty,$$

and

$$C_3^S = \int_0^\infty \left(\frac{xf'(x)}{2}\right)^2 e^{-ax}f(x)\,\mathrm{d}x \leq \frac{\|(f')^2\| \cdot \|f\|}{4} \int_0^\infty x^2e^{-ax}\,\mathrm{d}x = \frac{\|f'\|^2\|f\|}{2a^3} < \infty,$$

where the norm is again defined by $\|g\| = \sup_{x\in[0,\infty)} |g(x)|$ for a bounded $g : [0, \infty) \to \mathbb{R}$. End Proof

Proof of Theorem 6.8. This proof follows the proof of Theorem 5.7. We only present the proof for the local part. For simplicity, write $i(n) = i_L^S(n, x)$.

By the definition of $i(n)$ we know that $\lim_{n\to\infty} i(n) = \infty$ and

$$\begin{aligned}
&\text{MSE}\left[F_{i(n)}(x)\right] \leq \text{MSE}\left[\hat{F}^S_{m,n}(x)\right] \leq \text{MSE}\left[F_{i(n)-1}(x)\right] \\
\Leftrightarrow\ & i(n)^{-1}\sigma^2(x) \leq n^{-1}\sigma^2(x) - m^{-1/2}n^{-1}V^S(x) + m^{-2}(b^S(x))^2 \\
&+ o_x(m^{-1/2}n^{-1}) + o_x(m^{-2}) \leq (i(n)-1)^{-1}\sigma^2(x) \\
\Leftrightarrow\ & 1 \leq \frac{i(n)}{n}\left[1 - m^{-1/2}\theta^S(x) + m^{-2}n\gamma^S(x) + o_x(m^{-1/2}) + o_x(m^{-2}n)\right] \leq \frac{i(n)}{i(n)-1},
\end{aligned} \tag{6.19}$$

where $\theta^S(x) = \frac{V^S(x)}{\sigma^2(x)}$ and $\gamma^S(x) = \frac{(b^S(x))^2}{\sigma^2(x)}$. Now, if $mn^{-1/2} \to \infty$ ($\Leftrightarrow m^{-2}n \to 0$), taking the limit $n \to \infty$ leads to

$$\frac{i(n)}{n} \to 1, \tag{6.20}$$

so that

$$i(n) = n + o_x(n) = n(1 + o_x(1)).$$

(a) We assume that $mn^{-2/3} \to \infty$ and $mn^{-2} \to 0$. Rewrite Eq. (6.19) as

$$\begin{aligned}
&m^{-1/2}n^{-1}\theta^S(x) \leq A_{1,n} + m^{-2}\gamma^S(x) + o_x(m^{-1/2}n^{-1}) + o_x(m^{-2}) \leq m^{-1/2}n^{-1}\theta^S(x) + A_{2,n} \\
\Leftrightarrow\ & \theta^S(x) \leq m^{1/2}nA_{1,n} + m^{-3/2}n\gamma^S(x) + o_x(1) + o_x(m^{-3/2}n) \leq \theta^S(x) + m^{1/2}nA_{2,n},
\end{aligned} \tag{6.21}$$

where

$$A_{1,n} = \frac{1}{n} - \frac{1}{i(n)} \quad \text{and} \quad A_{2,n} = \frac{1}{i(n)-1} - \frac{1}{i(n)}.$$

It holds that

$$\lim_{n\to\infty} m^{1/2}nA_{1,n} = \left(\lim_{n\to\infty} \frac{i(n)-n}{m^{-1/2}n}\right)\left(\lim_{n\to\infty} \frac{n}{i(n)}\right) = \lim_{n\to\infty} \frac{i(n)-n}{m^{-1/2}n},$$

and because $m^{1/2}n^{-1} = (mn^{-2})^{1/2} \to 0$,

$$\lim_{n\to\infty} m^{1/2}nA_{2,n} = \left(\lim_{n\to\infty} m^{1/2}n^{-1}\right)\left(\lim_{n\to\infty} \frac{n}{i(n)}\right)\left(\lim_{n\to\infty} \frac{n}{i(n)-1}\right) = 0.$$

We also know that $m^{-3/2}n = (mn^{-2/3})^{-3/2} \to 0$, hence

$$\lim_{n\to\infty} \frac{i(n)-n}{m^{-1/2}n} = \theta^S(x) \Rightarrow \frac{i(n)-n}{m^{-1/2}n} = \theta^S(x) + o_x(1)$$

follows from Eq. (6.21).

(b) The second part can be proven with similar arguments. If $mn^{-2/3} \to c$ it also holds that $m^{-2}n = (mn^{-2/3})^{-3/2}m^{-1/2} \to 0$ and $m^{1/2}n^{-1} = (mn^{-2/3})^{1/2}n^{-2/3} \to 0$ so that we get that

$$\lim_{n\to\infty} \frac{i(n)-n}{m^{-1/2}n} = \theta^S(x) - c^{-3/2}\gamma^S(x)$$

and with

$$\lim_{n\to\infty} \frac{i(n)-n}{m^{-1/2}n} = \left(\lim_{n\to\infty} \frac{i(n)-n}{n^{2/3}}\right)\left(\lim_{n\to\infty} m^{1/2}n^{-1/3}\right) = c^{1/2} \lim_{n\to\infty} \frac{i(n)-n}{n^{2/3}}$$

the claim

$$c^{1/2}\frac{i(n)-n}{n^{2/3}} = \theta^S(x) - c^{-3/2}\gamma^S(x) + o_x(1)$$

holds.

The global part can be proved analogously to the proof of Theorem 5.7 with $\tilde{\theta}^S = \frac{C_2^S}{C_1^S}$ and $\tilde{\gamma}^S = \frac{C_3^S}{C_1^S}$. End Proof

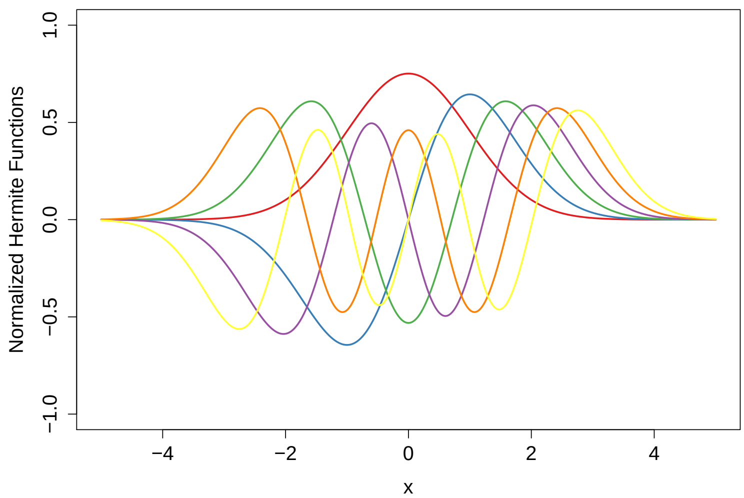

Figure 8: The normalized Hermite functions for $k \in \{0, ..., 5\}$.

7 Hermite Distribution Function Estimation

In this section, let $X_1, X_2, ...$ again be a sequence of i.i.d. random variables with unknown distribution function F and unknown density function f. We receive a finite random sample $X_1, ..., X_n, n \in \mathbb{N}$, to estimate the distribution function F with the so-called Hermite polynomials that are introduced in Section 7.1.

We first do this for f supported on $[0, \infty)$ in Section 7.3, following [39] and then for $(-\infty, \infty)$ in Section 7.4, following [40].

During this section, we want the following assumption to hold.

Assumption 7.1. *The density function f is in L_2.* End Assumption

Note that for many results to hold, the additional assumption of $\left(x - \frac{d}{\mathrm{d}x}\right)^r f \in L_2$ where $r > 1$ needs to be fulfilled. This assures that the function f is rapidly decreasing, which makes sense as the Hermite functions share this property, see [41].

One result will be that the proposed estimator is inferior to the kernel distribution estimator in terms of the asymptotic rate of convergence but there is one clear advantage. For the empirical distribution function (EDF) and the kernel estimator, the sequential (online) estimation (i.e., to process the observations sequentially so that the storage of all observations is not necessary) of the distribution function is only possible at a particular x (see for example [42]). For the estimator introduced in this section, the sequential estimation is possible at an arbitrary x. In addition, the time that it takes to update the estimate is $O(1)$ and hence, does not grow with the number of samples.

Throughout this section we sometimes use the notation $a(x) \sim b(x), x \to \infty$, which means that

$$\lim_{x \to \infty} \frac{a(x)}{b(x)} = 1.$$

We now introduce the Hermite polynomials.

7.1 Hermite Polynomials

The so-called Hermite polynomials H_k are defined by

$$H_k(x) = (-1)^k e^{x^2} \frac{d^k}{\mathrm{d}x^k} e^{-x^2}.$$

An explicit expression for the polynomials is

$$H_k(x) = k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m}{m!(k-2m)!}(2x)^{k-2m}.$$

These polynomials are orthogonal under e^{-x^2}, which means that

$$\int_{-\infty}^{\infty} e^{-x^2} H_k(x) H_l(x)\, \mathrm{d}x = \sqrt{\pi} 2^k k! \delta_{kl},$$

where δ_{kl} is the Kronecker delta function. The normalized Hermite functions are defined by

$$h_k(x) = (2^k k! \sqrt{\pi})^{-1/2} e^{\frac{-x^2}{2}} H_k(x). \tag{7.1}$$

They form an orthonormal basis for L_2 and hence, fulfill

$$\int_{-\infty}^{\infty} h_k(x) h_l(x)\, \mathrm{d}x = \delta_{kl}.$$

The normalized Hermite functions are illustrated in Figure 8 for $k \in \{0, ..., 5\}$.

The Hermite polynomials satisfy the inequality

$$(2^k k! \sqrt{\pi})^{-1/2} |H_k(x))| e^{\frac{-x^2}{2}} \leq c_a (k+1)^{-1/4}, |x| \leq a$$

for some constant c_a and non-negative a and the inequality

$$(2^k k! \sqrt{\pi})^{-1/2} |x^{-1/3} H_k(x))| e^{\frac{-x^2}{2}} \leq d_b (k+1)^{-1/4}, |x| \geq b$$

for some constant d_b and positive b (see [43, Theorem 8.91.3], used in [44, p. 176, p. 177]).

In this section, we use the Gauss-Hermite expansion to estimate the density function and the distribution function.

7.2 Gauss-Hermite Expansion

The Gauss-Hermite expansion is for example defined in [45]. It has good convergence properties and is robust to outliers, see [46]. In the following, we show the necessary steps to obtain this expansion. We define

$$Z(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}},$$

$$\alpha_k = \frac{\sqrt{\pi}}{2^{k-1} k!},$$

and

$$a_k = \int_{-\infty}^{\infty} f(x) h_k(x)\, \mathrm{d}x.$$

As already mentioned before, the normalized Hermite functions h_k, defined in Eq. (7.1), form an orthonormal basis for L_2. Using this result, it makes sense that for $f \in L_2$,

$$f(x) = \sum_{k=0}^{\infty} a_k h_k(x) \tag{7.2}$$

$$= \sum_{k=0}^{\infty} \sqrt{\alpha_k} \cdot a_k H_k(x) Z(x). \tag{7.3}$$

From now on, the expressions in Eq. (7.2) and Eq. (7.3) are used interchangeably and are called the Gauss-Hermite expansion. The equality of the two expressions holds because of

$$\begin{aligned}\sum_{k=0}^{\infty} a_k h_k(x) &= \sum_{k=0}^{\infty} a_k (2^k k! \sqrt{\pi})^{-1/2} e^{\frac{-x^2}{2}} H_k(x)\\ &= \sum_{k=0}^{\infty} \sqrt{\frac{\sqrt{\pi}}{2^{k-1}k!}} a_k \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} H_k(x)\\ &= \sum_{k=0}^{\infty} \sqrt{\alpha_k} \cdot a_k \cdot H_k(x) Z(x).\end{aligned}$$

The expression Eq. (7.2) shows that the density of the normal distribution can be estimated with only the first summand $a_0 h_0$. For the standard normal distribution, the first summand is of the form

$$a_0 h_0(x) = \frac{1}{\pi^{1/4}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2} \, \mathrm{d}x \cdot \frac{1}{\pi^{1/4}} e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

All the other summands are zero in this case.

The infinite sum in Eq. (7.2) and Eq. (7.3) is not desirable. A truncation of the sum leads to the $N+1$ truncated expansion

$$\begin{aligned}f_N(x) &= \sum_{k=0}^{N} a_k h_k(x)\\ &= \sum_{k=0}^{N} \sqrt{\alpha_k} \cdot a_k H_k(x) Z(x).\end{aligned}$$

The coefficients a_k are chosen so that the L_2-distance between f and f_N is minimized. A detailed explanation can be found in [47, 2.3 Orthogonal sequences].

In the following, if $N = N(n)$ depends on n, which is always the case unless explicitly mentioned, we assume that $N \to \infty$ for $n \to \infty$.

In the following, we deal with Hermite density estimation.

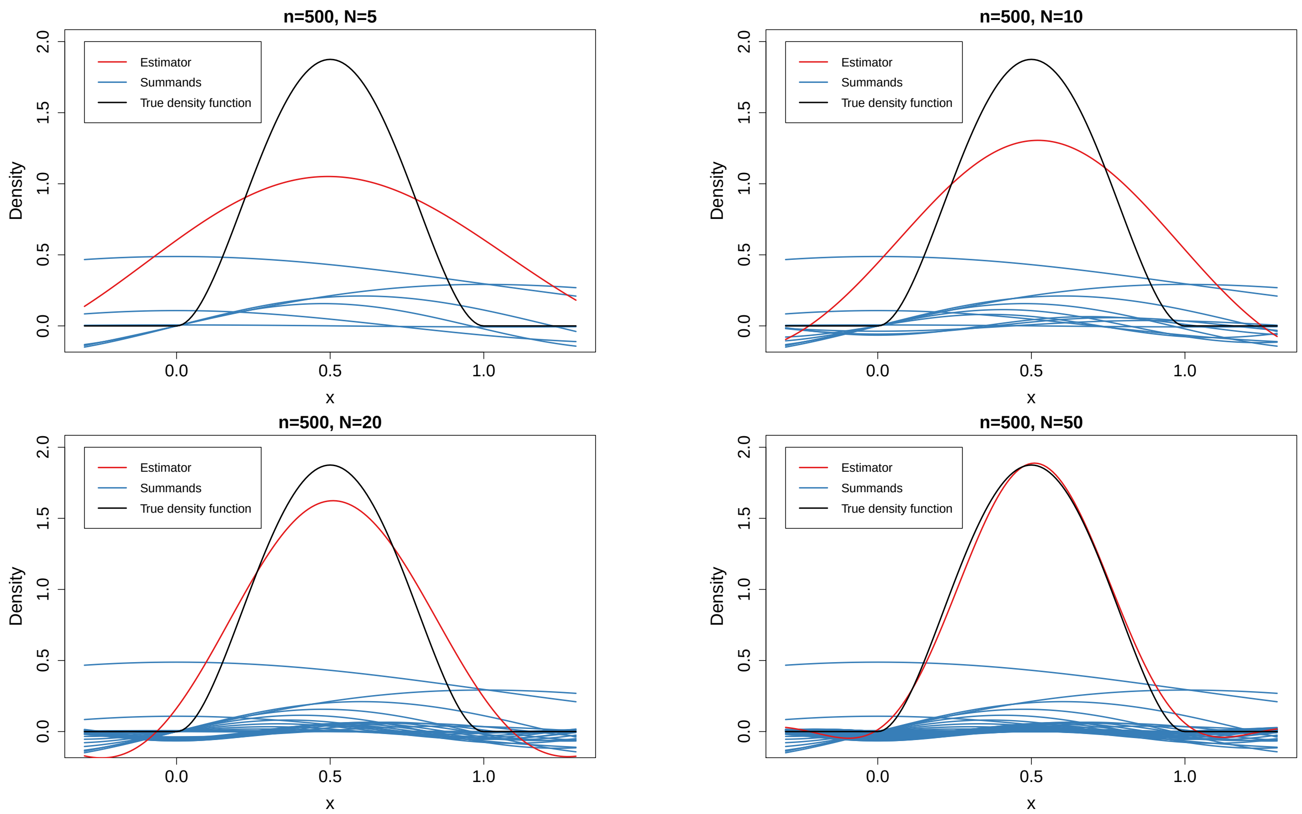

Figure 9: Illustration of the Hermite density estimator for different parameters.

7.2.1 Density Estimation

With the expressions defined above we now define a density estimator. The coefficients a_k can also be written as $a_k = \mathbb{E}[h_k(Y)]$, where Y is a random variable with density function f. Following [39], a_k can be estimated by

$$\hat{a}_k = \frac{1}{n}\sum_{i=1}^{n} h_k(X_i) \tag{7.4}$$

with the law of large numbers. Therefore, the above mentioned density estimator is defined as

$$\begin{aligned} \hat{f}_{N,n}(x) &= \sum_{k=0}^{N} \hat{a}_k h_k(x) \\ &= \sum_{k=0}^{N} \sqrt{\alpha_k} \cdot \hat{a}_k H_k(x) Z(x). \end{aligned} \tag{7.5}$$

This estimator is illustrated in Figure 9.

The next step is to calculate the MISE. With Parseval's identity and the fact that the functions h_k form an orthonormal basis for L_2, we obtain

$$\begin{aligned} \text{MISE}\left[\hat{f}_{N,n}\right] &= \mathbb{E}\left[\int_{-\infty}^{\infty} \left(\hat{f}_{N,n}(x) - f(x)\right)^2 \mathrm{d}x\right] \\ &= \mathbb{E}\left[\|\hat{f}_{N,n} - f\|_{L^2}^2\right] = \mathbb{E}\left[\sum_{k=0}^{\infty} | < \hat{f}_{N,n} - f, h_k > |^2\right] \\ &= \underbrace{\mathbb{E}\left[\sum_{k=0}^{N} (\hat{a}_k - a_k)^2\right]}_{S_1} + \underbrace{\sum_{k=N+1}^{\infty} a_k^2}_{S_2} \end{aligned} \tag{7.6}$$

because

$$
\begin{aligned}
< \hat{f}_{N,n} - f, h_k > &= \int_{-\infty}^{\infty} \left[\sum_{l=0}^{N} \hat{a}_l h_l(x) - \sum_{r=0}^{\infty} a_r h_r(x) \right] h_k(x) \, \mathrm{d}x \\
&= \sum_{l=0}^{N} \hat{a}_l \int_{-\infty}^{\infty} h_l(x) h_k(x) \, \mathrm{d}x - \sum_{r=0}^{\infty} a_r \int_{-\infty}^{\infty} h_r(x) h_k(x) \, \mathrm{d}x \\
&= \begin{cases} \hat{a}_k - a_k, & \text{if } k \leq N, \\ -a_k, & \text{if } k > N. \end{cases}
\end{aligned}
$$

The first term S_1 in Eq. (7.6) is the integrated variance term that gives the error that we get from using the estimates $\hat{a}_k$ instead of a_k. The second term S_2 in Eq. (7.6) is the integrated squared bias term that represents the error due to truncation. In [44, Theorem 1], the MISE consistency of the density estimator was proven under the condition $N^{5/6}/n \to 0$.

For many density functions, the Gauss-Hermite estimation yields good estimates, see [46, p. 208]. The downside is that through the truncation, the results can get negative for certain values of x. Furthermore, for distributions that deviate strongly from the Gaussian distribution, it might be necessary to choose large N to get a satisfactory fit.

Note that it is not possible to estimate the distribution function in the same way as the density function because F is not in L_2. This is easy to see with

$$
\int_{-\infty}^{\infty} F^2(x) \, \mathrm{d}x \leq \int_{-\infty}^{\infty} F(x) \, \mathrm{d}x,
$$

which is not bounded. Hence, the distribution function has to be estimated in a different way that is explained in the next sections for the real half line and the real line.

Before doing so, the next section is a small excursion that deals with a way to estimate the distribution function with Hermite polynomials that is not the focus of this section. The approach is needed later in Lemma 7.9. It is similar to the Gauss-Hermite estimator.

7.2.2 Gram-Charlier Series of Type A

Here, we quickly define the Gram-Charlier Series of Type A that can be used to estimate the distribution function and the density function, following [40]. Afterwards, we turn our attention to the Gauss-Hermite expansion again that is the topic of the rest of Section 7.

We first need to define the Chebyshev-Hermite polynomials

$$
H_{e_k}(x) = 2^{-\frac{k}{2}} H_k\left(\frac{x}{\sqrt{2}}\right)
$$

and the function

$$
Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.
$$

Now, if a density function f can formally be expanded as

$$
f(x) = \sum_{k=0}^{\infty} c_k H_{e_k}(x) Z(x)
$$

with

$$
c_k = \frac{1}{k!} \int_{-\infty}^{\infty} f(x) H_{e_k}(x) \, \mathrm{d}x,
$$

this can be used to estimate the density function and hence, also the distribution function. The truncated version has the form

$$f(x) = \sum_{k=0}^{N} c_k H_{e_k}(x) Z(x)$$

and following [39], the distribution function can be estimated by

$$\hat{F}_{N,n}^{GC}(x) = \sum_{k=0}^{N} \hat{c}_k \int_{-\infty}^{x} H_{e_k}(x) Z(y) dy, \tag{7.7}$$

where

$$\hat{c}_k = \frac{1}{k! \cdot n} \sum_{i=1}^{n} H_{e_k}(X_i).$$

This is the end of the excursion.

The next section deals with the Hermite distribution function estimator on the real half line.

7.3 The Distribution Function Estimator on the Real Half Line

We now consider distributions supported on $[0, \infty)$ and introduce the Gauss-Hermite distribution estimator on the real half line. Furthermore, we calculate the MSE and MISE of the Gauss-Hermite distribution estimator. Among other things, we will establish that they both directly depend on MISE $\left[\hat{f}_{N,n}\right]$.

Two examples for distributions fitting here are the chi-squared distribution and the exponential distribution. They are both challenging for the Gauss-Hermite estimator as they considerably differ from the normal distribution for certain parameters. The exponential distribution is even more challenging as the mode is zero.

In the case of the real half line, following [39], the Gauss-Hermite distribution estimator is calculated by

$$\hat{F}_{N,n}^{H}(x) = \int_{0}^{x} \hat{f}_{N,n}(t)\, \mathrm{d}t,$$

which leads to

$$\begin{aligned}
\hat{F}_{N,n}^{H}(x) &= \int_{0}^{x} \hat{f}_{N,n}(t)\, \mathrm{d}t \\
&= \sum_{k=0}^{N} \sqrt{\alpha_k} \hat{a}_k \int_{0}^{x} H_k(t) Z(t)\, \mathrm{d}t \\
&= \sum_{k=0}^{N} \sqrt{\alpha_k} \hat{a}_k k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m 2^{k-2m}}{m!(k-2m)!\sqrt{2\pi}} \int_{0}^{x} t^{k-2m} e^{\frac{-t^2}{2}}\, \mathrm{d}t \\
&= \sum_{k=0}^{N} \sqrt{\alpha_k} \hat{a}_k k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m 2^{k-2m}}{m!(k-2m)!\sqrt{2\pi}} 2^{\frac{k}{2}-m-\frac{1}{2}} \gamma\left(-m + \frac{k}{2} + \frac{1}{2}, \frac{x^2}{2}\right),
\end{aligned}$$

where γ is the lower incomplete gamma function. To summarize, the Gauss-Hermite distribution function estimator on $[0, \infty)$ has the form

$$\hat{F}_{N,n}^{H}(x) = \sum_{k=0}^{N} \hat{a}_k \sqrt{k!} \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m 2^{k-3m-\frac{1}{2}} \gamma\left(-m + \frac{k}{2} + \frac{1}{2}, \frac{x^2}{2}\right)}{m!(k-2m)!\pi^{\frac{1}{4}}}, \quad x \geq 0. \tag{7.8}$$

We now give the asymptotic bias and variance of the Gauss-Hermite distribution estimator.

7.3.1 Bias and Variance

The next lemma gives the asymptotic behavior of the variance.

Lemma 7.1. *Suppose f is supported on $[0,\infty)$ and $\mathbb{E}[|X|^{2/3}] < \infty$ for a random variable X with density function $f \in L_2$. Then, we have that*

$$\mathbb{E}\left[\left|\hat{F}^H_{N,n}(x) - \mathbb{E}\left[\hat{F}^H_{N,n}(x)\right]\right|^2\right] = O_x\left(\frac{N^{3/2}}{n}\right)$$

uniformly in x as $n \to \infty$. In addition,

$$\mathbb{E}\left[\left|\hat{F}^H_{N,n}(x) - \mathbb{E}\left[\hat{F}^H_{N,n}(x)\right]\right|^2\right] \to 0$$

for $\frac{N^{3/2}(n)}{n} \to 0$. End Lemma

For the proof, see Proofs Hermite.

Now, the squared bias can be calculated.

Lemma 7.2. *Suppose $f \in L_2$ is supported on $[0,\infty)$ and $\mathbb{E}[|X|^{2/3}] < \infty$ for a random variable X with density function f. Then,*

$$\left|\mathbb{E}\left[\hat{F}^H_{N,n}(x)\right] - F(x)\right|^2 = O_x\left(N^{-r+1/2}\right),$$

if $r \geq 1$ derivatives of f exist and $\left(x - \frac{d}{\mathrm{d}x}\right)^r f \in L_2$. End Lemma

For the proof, see Proofs Hermite.

It would be possible to calculate the MSE and MISE with Lemma 7.1 and Lemma 7.2 but this would lead to worse results than the following. In the next section, we focus on the MSE of the Gauss-Hermite distribution estimator. One result is that the estimator is MSE consistent.

7.3.2 MSE

The next lemma follows from [39, Proposition 1].

Lemma 7.3. *If f is supported on $[0,\infty)$ and $f \in L_2$, then*

$$\mathrm{MSE}\left[\hat{F}^H_{N,n}(x)\right] = \mathbb{E}\left|\hat{F}^H_{N,n}(x) - F(x)\right|^2 \leq x\,\mathrm{MISE}\left[\hat{f}_{N,n}\right]$$

for fixed x. End Lemma

Proof. As we are on the real half line, it holds that

$$\left|\hat{F}^H_{N,n}(x) - F(x)\right| = \left|\int_0^x \left[\hat{f}_{N,n}(y) - f(y)\right] dy\right|.$$

With the Cauchy-Schwarz inequality we know that

$$\begin{aligned}\left|\hat{F}^H_{N,n}(x) - F(x)\right|^2 &\leq \left[\int_0^x \left|\hat{f}_{N,n}(y) - f(y)\right|^2 dy\right]\left[\int_0^x 1 dy\right]\\ &= x\left[\int_0^x \left|\hat{f}_{N,n}(y) - f(y)\right|^2 dy\right]\\ &\leq x\left[\int_0^\infty \left|\hat{f}_{N,n}(y) - f(y)\right|^2 dy\right].\end{aligned}$$

It follows that

$$\mathbb{E}\left[\left|\hat{F}^H_{N,n}(x) - F(x)\right|^2\right] \leq x\,\mathbb{E}\left[\int_0^\infty \left|\hat{f}_{N,n}(y) - f(y)\right|^2 dy\right]$$
$$\leq x\,\mathbb{E}\left[\int_{-\infty}^\infty \left|\hat{f}_{N,n}(y) - f(y)\right|^2 dy\right] = x\,\mathrm{MISE}\left[\hat{f}_{N,n}\right]$$

holds. End Proof

With [40], we also know that in the case where f has support $[a, b]$, the very similar result

$$\mathbb{E}\left[\left|\hat{F}^H_{N,n}(x) - F(x)\right|^2\right] \leq (x-a)\,\mathrm{MISE}\left[\hat{f}_{N,n}\right]$$

holds. The proof is the same as the proof for Lemma 7.3 with $\int_a^x 1dy = x - a$.

The next theorem follows from [39, Theorem 1] and shows the consistency with respect to MSE.

Theorem 7.1. *Suppose the support of f is $[0, \infty)$ and $f \in L_2$. Then, we have that*

$$\mathbb{E}\left[\left|\hat{F}^H_{N,n}(x) - F(x)\right|^2\right] \to 0$$

if $\frac{N^{1/2}(n)}{n} \to 0$ as $n \to \infty$ and $\mathbb{E}[|X|^{2/3}] < \infty$. Here, X is a random variable with density function f. End Theorem

Proof. With the proof of [44, Theorem 2], we know that $\mathrm{MISE}\left[\hat{f}_{N,n}\right] \to 0$ under the given conditions. Hence, the desired result follows from Lemma 7.3. End Proof

Following from [39, Theorem 2], the next result provides information about the asymptotic behavior of the MSE.

Theorem 7.2. *Suppose f is supported on $[0, \infty)$, $f \in L_2$, $r \geq 1$ derivatives of f exist, and $\left(x - \frac{d}{dx}\right)^r f \in L_2$. If $\mathbb{E}[|X|^{2/3}] < \infty$, we have*

$$\mathrm{MSE}\left[\hat{F}^H_{N,n}(x)\right] = x\left[O\left(\frac{N^{1/2}}{n}\right) + O\left(N^{-r}\right)\right]$$

for $x \in [0, \infty)$. End Theorem

Proof. Under the given conditions, we know with the proof of [44, Theorem 2] that

$$\mathrm{MISE}\left[\hat{f}_{N,n}\right] = O\left(\frac{N^{1/2}}{n}\right) + O\left(N^{-r}\right).$$

The theorem follows again with Lemma 7.3. End Proof

To find the asymptotically optimal $N(n)$ with respect to MSE, set $\frac{N^{1/2}}{n} = N^{-r}$ and solve this equation for N, which leads to $N(n) \sim n^{2/(2r+1)}$. This approach is motivated by the fact that $\mathrm{MISE}\left[\hat{f}_{N,n}\right] = O\left(\frac{N^{1/2}}{n}\right) + O\left(N^{-r}\right) = O\left(\max\left(\frac{N^{1/2}}{n}, N^{-r}\right)\right)$. Now, we have to find the minimum of this maximum (dark blue in Figure 10) which is found at the intersection (orange) of the two functions depending on N because the squared bias (bright blue) decreases with growing N while the variance increases (black). This leads to the following corollary.

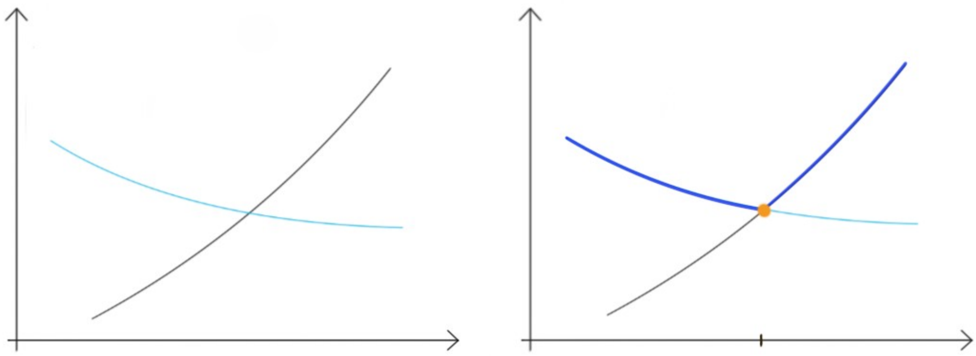

Figure 10: How to find the optimal N.

Corollary 7.1. *Suppose that f is supported on $[0,\infty)$, $f \in L_2$, $r \geq 1$ derivatives of f exist, $\left(x - \frac{d}{dx}\right)^r f \in L_2$, and $\mathbb{E}[|X|^{2/3}] < \infty$. The asymptotically optimal N for estimating F with respect to MSE is*

$$N_{opt} \sim n^{2/(2r+1)},$$

which leads to

$$\text{MSE}\left[\hat{F}^H_{N,n}(x)\right] = xO(n^{-\frac{2r}{2r+1}})$$

for $x \in [0,\infty)$. End Corollary

For $r = 1$, the rate that is achieved here is $O(n^{-2/3})$. Compared to the MSE rate $O(n^{-1})$ of the kernel distribution estimator that can be found in Corollary 3.2, this is suboptimal. However, for $r \to \infty$, the rate approaches $O(n^{-1})$ under the conditions that $r > 1$ derivatives exist and $\left(x - \frac{d}{dx}\right)^r f \in L_2$. These conditions hold for the following examples that can be found in [39] and [40].

7.3.3 Examples

Power-Law Distribution An important class of distributions for which the Gauss-Hermite estimator can be used is the class of power-law distributions. The density function is given by

$$f(x) = \frac{\alpha - 1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\alpha},$$

where $\alpha > 1$. As $x_{\min} > 0$ is assumed, f is supported on $[x_{\min}, \infty) \subset [0,\infty)$, $f \in L_2$ and all derivatives of f exist for $x \geq x_{\min}$. In the case of $\alpha > 2$, we have finite mean $\mathbb{E}[X] < \infty$.

With

$$(x - D)^r f(x) = \frac{\alpha - 1}{x_{\min}^{1-\alpha}}(x^r + ... + D^r)x^{-\alpha} = O\left(x^{-(\alpha - r)}\right),$$

we know that $[(x - D)^r f(x)]^2 = O\left(x^{-2(\alpha - r)}\right)$ and hence, we must have $2(\alpha - r) > 1$ to assure integrability of $[(x - D)^r f(x)]^2$. It follows that $r < \alpha - \frac{1}{2}$ and hence, $r = \lceil \alpha - \frac{1}{2} \rceil - 1$.

This means that with $\frac{N^{1/2}(n)}{n} \to 0$, Corollary 7.1 and Corollary 7.2, we get

$$\text{MSE}\left[\hat{F}^H_{N,n}(x)\right] = xO\left(n^{-\frac{2\lceil \alpha - \frac{1}{2} \rceil - 2}{2\lceil \alpha - \frac{1}{2} \rceil - 1}}\right),$$

and

$$\text{MISE}\left[\hat{F}^H_{N,n}\right] = O\left(n^{-\frac{2\lceil \alpha - \frac{1}{2} \rceil - 2}{2\lceil \alpha - \frac{1}{2} \rceil - 1}}\right).$$

Now, for finite mean, i.e., $\alpha > 2$, the rate is $O(n^{-2/3})$ or better. For finite variance, i.e., $\alpha > 3$, the rate is $O(n^{-4/5})$ or better. As α and hence, the number of finite moments goes to infinity, the rate approaches $O(n^{-1})$.

Power Law distribution with exponential cutoff The density function of the power law distribution with exponential cutoff has the form

$$f(x) = \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha, \lambda x_{\min})} x^{-\alpha} e^{-\lambda x},$$

where $\alpha, \lambda > 0$. The function f is supported on $[x_{\min}, \infty), x_{\min} > 0$. This distribution has similar properties to a power law distribution before this behavior is overwhelmed by the exponential decay factor. It helps in modeling the inter-event waiting times of processes arising from human dynamics (for example waiting times between phone calls).

On $[x_{\min}, \infty)$, it is easy to see that $r \geq 1$ derivatives of f exist and by induction,

$$\left(x - \frac{d}{\mathrm{d}x}\right)^r f(x) = \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha, \lambda x_{\min})} \sum_{k=-r}^{r} g_k(\alpha, \lambda) x^{-\alpha+k} e^{-\lambda x},$$

where $g_k(\alpha, \lambda)$ are polynomial functions of α and γ. Then it holds that $\left(x - \frac{d}{\mathrm{d}x}\right)^r f \in L_2$ because

$$\begin{aligned}
&\int_{x_{\min}}^{\infty} \left[\left(x - \frac{d}{\mathrm{d}x}\right)^r \lambda e^{-\lambda x}\right]^2 \mathrm{d}x \\
&= \left(\frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha, \lambda x_{\min})}\right)^2 \sum_{k=-r}^{r} g_k(\alpha, \lambda) g_l(\alpha, \lambda) \int_{x_{\min}}^{\infty} x^{-2\alpha+k+l} e^{-2\lambda x} \mathrm{d}x \\
&= \left(\frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha, \lambda x_{\min})}\right)^2 \sum_{k=-r}^{r} g_k(\alpha, \lambda) g_l(\alpha, \lambda) \frac{\Gamma(1-2\alpha+k+l, 2\lambda x_{\min})}{(2\lambda)^{1-2\alpha+k+l}} < \infty
\end{aligned}$$

for all $r \geq 1$. This shows that the conditions hold.

Exponential Distribution The density function of the exponential distribution supported on $[0, \infty)$ is given by

$$f(x) = \lambda e^{-\lambda x}, \lambda > 0.$$

On $[0, \infty)$, it is clear that $r \geq 1$ derivatives exist and it is easy to show by induction that

$$\left(x - \frac{d}{\mathrm{d}x}\right)^r f(x) = \sum_{k=0}^{r} g_k(\lambda) x^k e^{-\lambda x},$$

where $g_k(\lambda)$ are polynomial functions of λ. Then it follows that

$$\begin{aligned}
&\int_0^{\infty} \left[\left(x - \frac{d}{\mathrm{d}x}\right)^r \lambda e^{-\lambda x}\right]^2 \mathrm{d}x \\
&= \int_0^{\infty} \sum_{k,l=0}^{r} g_k(\lambda) g_l(\lambda) x^{k+l} e^{-2\lambda x} \mathrm{d}x \\
&= \sum_{k,l=0}^{r} g_k(\lambda) g_l(\lambda) \int_0^{\infty} x^{k+l} e^{-2\lambda x} \mathrm{d}x \\
&= \sum_{k,l=0}^{r} g_k(\lambda) g_l(\lambda) \frac{(k+l)!}{(2\lambda)^{k+l+1}} < \infty
\end{aligned}$$

for all finite $r \geq 1$ and hence, $\left(x - \frac{d}{dx}\right)^r f \in L_2$ and the conditions hold. Thus, the rate of the MSE for the Hermite distribution estimator approaches $O(n^{-1})$.

This distribution is especially interesting in a streaming scenario: If the given samples represent the inter-event waiting time of a Poisson process, this waiting time is exponentially distributed.

We now turn our attention to the MISE of the Hermite estimator on the real half line.

7.3.4 MISE

The MISE is defined by

$$\text{MISE}\left[\hat{F}^H_{N,n}\right] = \mathbb{E}\left[\int_0^\infty \left[\hat{F}^H_{N,n}(x) - F(x)\right]^2 f(x)\,\mathrm{d}x\right] = \int_0^\infty \mathbb{E}\left[\hat{F}^H_{N,n}(x) - F(x)\right]^2 f(x)\,\mathrm{d}x.$$

Note that we use the density function f as a weighting factor. Then we obtain the following result that one finds in [39, Proposition 2], which bounds the MISE of the distribution estimator by the MISE of the density estimator.

Lemma 7.4. *If $f \in L_2$ is supported on $[0, \infty)$, then*

$$\text{MISE}\left[\hat{F}^H_{N,n}\right] \leq \mu\,\text{MISE}\left[\hat{f}_{N,n}\right]$$

for finite mean $\mu = \int_0^\infty x f(x)\,\mathrm{d}x$. End Lemma

Proof. With Lemma 7.3, we have that

$$\text{MISE}\left[\hat{F}^H_{N,n}\right] \leq \text{MISE}\left[\hat{f}_{N,n}\right] \int_0^\infty x f(x)\,\mathrm{d}x = \mu\,\text{MISE}\left[\hat{f}_{N,n}\right]$$

for $\mu < \infty$. End Proof

The next theorem shows that the Gauss-Hermite estimator on the real half line is MISE consistent. The theorem follows from [39, Theorem 3].

Theorem 7.3. *For $f \in L_2$ supported on $[0, \infty)$ with mean $\mu < \infty$, we get*

$$\text{MISE}\left[\hat{F}^H_{N,n}\right] \to 0$$

for $\frac{N^{1/2}(n)}{n} \to 0$ as $n \to \infty$. End Theorem

Proof. From [44, Theorem 1] we know again that $\text{MISE}\left[\hat{f}_{N,n}\right] \to 0$ and the desired result follows from Lemma 7.4. For positive random variables, the fact that $\mathbb{E}[|X|^{2/3}] < \infty$ follows from $\mathbb{E}[X] < \infty$ and the Lyapunov inequality

$$\left(\mathbb{E}[|X|^s]\right)^{1/s} \leq \left(\mathbb{E}[|X|^t]\right)^{1/t}$$

for $0 < s < t$. End Proof

The next theorem, which follows from [39, Theorem 4], gives the asymptotic behavior of the MISE.

Theorem 7.4. *Let f be supported on $[0,\infty)$, $f \in L_2$, $r \geq 1$ derivatives of f exist, and $\left(x - \frac{d}{\mathrm{d}x}\right)^r f \in L_2$. Then*

$$\text{MISE}\left[\hat{F}^H_{N,n}\right] = \mu\left[O\left(\frac{N^{1/2}}{n}\right) + O\left(N^{-r}\right)\right]$$

for $\mu < \infty$. End Theorem

Proof. This follows from Lemma 7.4 and the proof of [44, Theorem 2], which says that

$$\text{MISE}\left[\hat{f}_{N,n}\right] = O\left(\frac{N^{1/2}}{n}\right) + O\left(N^{-r}\right)$$

under the given conditions that hold with the Lyapunov inequality. End Proof

As before, set $\frac{N^{1/2}}{n} = N^{-r}$ to find the asymptotically optimal $N(n)$ with respect to MISE, which leads to $N(n) \sim n^{2/(2r+1)}$. The next corollary follows.

Corollary 7.2. *Let f be supported on $[0,\infty)$, $f \in L_2$, $r \geq 1$ derivatives of f exist, and $\left(x - \frac{d}{\mathrm{d}x}\right)^r f \in L_2$. The asymptotically optimal N for estimating F with respect to MISE is*

$$N_{opt} \sim n^{2/(2r+1)},$$

which leads to

$$\text{MISE}\left[\hat{F}^H_{N,n}\right] = \mu O(n^{-\frac{2r}{2r+1}})$$

for $\mu < \infty$. End Corollary

Also note here that the rate of the MISE is worse than for the kernel estimator, as can be seen in Corollary 3.3.

We now prove the almost sure convergence of the estimator.

7.3.5 Almost Sure Convergence

The next theorem gives the asymptotic behavior of the difference between the estimator and the real distribution function F.

Theorem 7.5. *Suppose $f \in L_2$ is supported on $[0,\infty)$, is r times continuously differentiable, and $\left(x - \frac{d}{\mathrm{d}x}\right)^r f \in L_2$ where $r \geq 1$. If $\mathbb{E}[|X|^s] < \infty, s > \frac{8(r+1)}{3(2r+1)}$ and $N(n) \sim n^{\frac{2}{2r+1}}$, we have*

$$\left|\hat{F}^H_{N,n}(x) - F(x)\right| \to 0 \;\; a.s.$$

uniformly in x. End Theorem

For the proof, see Proofs Hermite.

The condition $s > \frac{8(r+1)}{3(2r+1)}$ that has to hold here is always satisfied for $s > \frac{16}{9}$ as stated in [44, Remark 2].

The next section deals with the asymptotic behavior of the Hermite estimator on the real half line.

7.3.6 Asymptotic Behavior

We now study the asymptotic behavior of the estimator and establish that the limit distribution is normal as for the estimators before.

Theorem 7.6. *For $x \in (0,\infty)$ with $0 < F(x) < 1$ and if f is differentiable in x, we obtain*

$$\sqrt{n}\left(\hat{F}^H_{N,n}(x) - \mathbb{E}\left[\hat{F}^H_{N,n}(x)\right]\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right),$$

for $n \to \infty$, where $\sigma^2(x)$ is defined as in Theorem 2.1. End Theorem

For the proof, see Proofs Hermite.

The next corollary deals with the asymptotic behavior of $\hat{F}^H_{N,n}(x) - F(x)$. With Lemma 7.2, it is easy to see that

$$n^{1/2}\left(\hat{F}^H_{N,n}(x) - F(x)\right) = n^{1/2}\left(\hat{F}^H_{N,n}(x) - \mathbb{E}\left[\hat{F}^H_{N,n}(x)\right]\right) + O_x(n^{1/2}N^{1/4-r/2}). \tag{7.9}$$

This leads directly to the following corollary.

Corollary 7.3. *Let $m, n \to \infty$. Then, for $x \in (0,\infty)$ with $0 < F(x) < 1$ and with the assumptions from Lemma 7.2, it holds that*

$$n^{1/2}\left(\hat{F}^H_{N,n}(x) - F(x)\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right),$$

if $n^{-1/2}N^{r/2-1/4} \to \infty$. End Corollary

The results above hold in the setting where $N = N(n)$. This means that they are not applicable to the online estimator defined in Section 7.5. It is, however, possible to say something about the MSE of the estimator with fixed N. In the present case of support on $[0,\infty)$, the online estimators with fixed N have MSE bounds determined by the integrated squared bias term of MISE $\left[\hat{f}_{N,n}\right]$ (see Eq. (7.6)). This bias stays the same for fixed N even as $n \to \infty$ but the simulation later shows that the online estimators are still useful.

With this knowledge, we now discuss what is important for the selection of N and which algorithm can be used to determine it.

7.3.7 Selection of *N*

The quality of the distribution estimator Eq. (7.8) is related to MISE $\left[\hat{f}_{N,n}\right]$ as can for example be seen in Lemma 7.3 and Lemma 7.4. According to [39], it is reasonable to minimize the MISE even though the optimal N is clearly different for the estimation of the density and the distribution function.

In Eq. (7.6), N controls the trade-off between the integrated variance and the integrated squared bias of the estimated density. Under certain conditions (see [44, Theorem 1]), the integrated variance term vanishes as $n \to \infty$ for N fixed. Hence, the bias term is now of interest. To minimize the bias, N should intuitively be as large as possible. However, larger N leads to an increase in both processing time and memory requirements. To find a balance here, one algorithm to find the optimal N is the Kronmal-Tarter stopping algorithm.

However, in line with [39], the choice of N does not really affect the effectiveness of the algorithm if N exceeds a certain size.

The next section deals with distributions on the real line.

7.4 The Distribution Function Estimator on the Real Line

We now take a look at distributions supported on $(-\infty, \infty)$. Then, the Gauss-Hermite distribution function estimator, following [40], is

$$\hat{F}^F_{N,n}(x) = \int_{-\infty}^{x} \hat{f}_{N,n}(t)\,\mathrm{d}t.$$

Using the definitions of H_k and Z, it follows for $x < 0$ that

$$\begin{aligned}
\hat{F}^F_{N,n}(x) &= \int_{-\infty}^{x} \hat{f}_{N,n}(t)\,\mathrm{d}t \\
&= \sum_{k=0}^{N} \sqrt{\alpha_k}\hat{a}_k \int_{-\infty}^{x} H_k(t)Z(t)\,\mathrm{d}t \\
&= \sum_{k=0}^{N} \sqrt{\alpha_k}\hat{a}_k k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m 2^{k-2m}}{m!(k-2m)!\sqrt{2\pi}} \int_{-\infty}^{x} t^{k-2m} e^{\frac{-t^2}{2}}\,\mathrm{d}t \\
&= \sum_{k=0}^{N} \sqrt{\alpha_k}\hat{a}_k k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m 2^{k-2m}}{m!(k-2m)!\sqrt{2\pi}} (-1)^{k-2m} 2^{\frac{k}{2}-m-\frac{1}{2}} \Gamma\left(-m + \frac{k}{2} + \frac{1}{2}, \frac{x^2}{2}\right).
\end{aligned}$$

We obtain the last equality through integration by substitution. Definining $z = \frac{t^2}{2}$, which means that $t = -\sqrt{2z}$ because $x < 0$, we get

$$\begin{aligned}
&\int_{-\infty}^{x} t^{k-2m} e^{\frac{-t^2}{2}}\,\mathrm{d}t \\
&= \int_{\infty}^{\frac{x^2}{2}} \left(-\sqrt{2z}\right)^{k-2m} e^{-z} \frac{1}{-\sqrt{2z}} dz \\
&= (-1)^{k-2m-1} 2^{\frac{k}{2}-m-\frac{1}{2}} \int_{\infty}^{\frac{x^2}{2}} z^{\frac{k}{2}-m-\frac{1}{2}} e^{-z} dz \\
&= (-1)^{k-2m} 2^{\frac{k}{2}-m-\frac{1}{2}} \int_{\frac{x^2}{2}}^{\infty} z^{\frac{k}{2}-m-\frac{1}{2}} e^{-z} dz.
\end{aligned}$$

For $x \geq 0$, it holds with similar arguments that

$$\begin{aligned}
\hat{F}^F_{N,n}(x) &= \int_{-\infty}^{x} \hat{f}_{N,n}(t)\,\mathrm{d}t \\
&= \int_{-\infty}^{\infty} \hat{f}_{N,n}(t)\,\mathrm{d}t - \int_{x}^{\infty} \hat{f}_{N,n}(t)\,\mathrm{d}t \\
&= \sum_{k=0}^{N} \sqrt{\alpha_k}\hat{a}_k k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m 2^{k-2m}}{m!(k-2m)!\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} t^{k-2m} e^{\frac{-t^2}{2}}\,\mathrm{d}t - \int_{x}^{\infty} t^{k-2m} e^{\frac{-t^2}{2}}\,\mathrm{d}t\right] \\
&= \sum_{k=0}^{N} \sqrt{\alpha_k}\hat{a}_k k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m 2^{\frac{3}{2}k-3m-\frac{1}{2}}}{m!(k-2m)!\sqrt{2\pi}} \left[\left[(-1)^{k-2m} + 1\right] \Gamma_1(k,m) - \Gamma_2(k,m)\right] \\
&= \sum_{k=0}^{N} \sqrt{\alpha_k}\hat{a}_k k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m 2^{\frac{3}{2}k-3m-\frac{1}{2}}}{m!(k-2m)!\sqrt{2\pi}} \left[(-1)^{k-2m} \Gamma_1(k,m) - \gamma_3(k,m)\right],
\end{aligned}$$

where

$$\Gamma_1(k,m) = \Gamma\left(-m + \frac{k}{2} + \frac{1}{2}\right),\ \Gamma_2(k,m) = \Gamma\left(-m + \frac{k}{2} + \frac{1}{2}, \frac{x^2}{2}\right),$$

and

$$\gamma_3(k,m) = \gamma\left(-m + \frac{k}{2} + \frac{1}{2}, \frac{x^2}{2}\right).$$

Summarizing this, we get

$$\hat{F}^F_{N,n}(x) = \begin{cases} \sum_{k=0}^{N} \hat{a}_k \sqrt{k!} \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m 2^{k-3m-\frac{1}{2}} \left[(-1)^{k-2m}\Gamma_1(k,m) + \gamma_3(k,m)\right]}{m!(k-2m)!\pi^{\frac{1}{4}}}, & x \geq 0, \\ \sum_{k=0}^{N} \hat{a}_k \sqrt{k!} \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^{-m+k} 2^{k-3m-\frac{1}{2}} \Gamma_2(k,m)}{m!(k-2m)!\pi^{\frac{1}{4}}}, & x < 0. \end{cases} \tag{7.10}$$

Another possibility is to set $\int_{-\infty}^{\infty} \hat{f}_{N,n}(t)\,\mathrm{d}t = 1$ and we achieve

$$\hat{F}^F_{N,n}(x) = \begin{cases} 1 - \sum_{k=0}^{N} \hat{a}_k \sqrt{k!} \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m 2^{k-3m-\frac{1}{2}} \Gamma\left(-m + \frac{k}{2} + \frac{1}{2}, \frac{x^2}{2}\right)}{m!(k-2m)!\pi^{\frac{1}{4}}}, & x \geq 0, \\ \sum_{k=0}^{N} \hat{a}_k \sqrt{k!} \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^{-m+k} 2^{k-3m-\frac{1}{2}} \Gamma\left(-m + \frac{k}{2} + \frac{1}{2}, \frac{x^2}{2}\right)}{m!(k-2m)!\pi^{\frac{1}{4}}}, & x < 0. \end{cases} \tag{7.11}$$

In the case where the support of f is $(-\infty, \infty)$, it is not possible to bound the MSE of the Gauss-Hermite distribution estimator by MISE $\left[\hat{f}_{N,n}\right]$. This is why the approach used in the section before for the real half line is not applicable. Hence, a new approach is pursued in [40], which is presented in the sequel.

We now take a closer look at the MSE and the MISE of the distribution estimator again. We will also present properties on the convergence behavior and the robustness.

7.4.1 MSE

For the squared bias, the next lemma holds. It follows from [40, Proposition 1].

Lemma 7.5. *If $f \in L_2$ is r times continuously differentiable and $\left(x - \frac{d}{\mathrm{d}x}\right)^r f \in L_2$ where $r > 2$, then*

$$\left|\mathbb{E}\left[\hat{F}^F_{N,n}(x)\right] - F(x)\right|^2 = O\left(N^{2-r}\right),$$

and

$$\left|\mathbb{E}\left[\hat{F}^F_{N,n}(x)\right] - F(x)\right|^2 \to 0$$

for $N \to \infty$. End Lemma

For the proof, see Proofs Hermite.

The asymptotic behavior of the variance is established in the next lemma, following from [40, Proposition 2].

Lemma 7.6. *Suppose $\mathbb{E}[|X|^{2/3}] < \infty$ for a random variable X with density function $f \in L_2$. Then, as $n \to \infty$,*

$$\mathbb{E}\left[\left|\hat{F}^F_{N,n}(x) - \mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]\right|^2\right] = O\left(\frac{N^{5/2}}{n}\right)$$

uniformly in x. In addition,

$$\mathbb{E}\left[\left|\hat{F}^F_{N,n}(x) - \mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]\right|^2\right] \to 0,$$

if $\frac{N^{5/2}(n)}{n} \to 0$. End Lemma

For the proof, see Proofs Hermite.

Now, the two previous results are used to get the asymptotic behavior of the MSE. The theorem follows from [40, Theorem 1].

Theorem 7.7. *If $f \in L_2$ is r times continuously differentiable and $\left(x - \frac{d}{\mathrm{d}x}\right)^r f \in L_2$ where $r > 2$ with $\mathbb{E}[|X|^{2/3}] < \infty$, it holds that*

$$\mathrm{MSE}\left[\hat{F}^F_{N,n}(x)\right] = \mathbb{E}\left[\left|\hat{F}^F_{N,n}(x) - F(x)\right|^2\right] = O\left(N^{-r+2}\right) + O\left(\frac{N^{5/2}}{n}\right)$$

uniformly in x. End Theorem

Proof. With Lemma 7.5 and Lemma 7.6, we know that

$$\begin{aligned}\mathrm{MSE}\left[\hat{F}^F_{N,n}(x)\right] &= \left|\mathbb{E}\left[\hat{F}^F_{N,n}(x)\right] - F(x)\right|^2 + \mathbb{E}\left[\left|\hat{F}^F_{N,n}(x) - \mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]\right|^2\right] \\ &= O\left(N^{-r+2}\right) + O\left(\frac{N^{5/2}}{n}\right)\end{aligned}$$

holds. End Proof

Note that the squared bias decreases with growing N while the variance increases. Hence, with the same explanation as for Corollary 7.1, the next corollary follows from setting $N^{-r+2} = \frac{N^{5/2}}{n}$.

Corollary 7.4. *Suppose $f \in L_2$ is r times continuously differentiable and $\left(x - \frac{d}{\mathrm{d}x}\right)^r f \in L_2$ where $r > 2$ with $\mathbb{E}[|X|^{2/3}] < \infty$. The asymptotically optimal N for estimating F with respect to MSE is*

$$N_{opt} \sim n^{2/(2r+1)},$$

which leads to

$$\mathrm{MSE}\left[\hat{F}^F_{N,n}(x)\right] = O\left(n^{-\frac{2(r-2)}{2r+1}}\right) = o(1)$$

uniformly in x, where $r > 2$ was used. End Corollary

Note that the convergence rate is worse compared to Corollary 7.1. On the other hand, the present rate is uniform. Furthermore, the rate is suboptimal compared to $O(n^{-1})$ of the smooth kernel distribution estimator (Corollary 3.2). However, for $r \gg 1$, the rate approaches optimal under the conditions that $r > 2$ derivatives exist and $\left(x - \frac{d}{\mathrm{d}x}\right)^r f \in L_2$.

We now take a look at the MISE of the Gauss-Hermite distribution estimator.

7.4.2 MISE

The MISE is defined by

$$\mathrm{MISE}\left[\hat{F}^F_{N,n}\right] = \int_{-\infty}^{\infty} \mathbb{E}\left[\left|\hat{F}^F_{N,n}(x) - F(x)\right|^2\right] f(x)\,\mathrm{d}x,$$

where f is again used as a weight function.

The next theorem gives the asymptotic behavior of the MISE and follows directly from Theorem 7.7.

Theorem 7.8. *Suppose $f \in L_2$ is r times continuously differentiable and $\left(x - \frac{d}{dx}\right)^r f \in L_2$ where $r > 2$. If additionally, $\mathbb{E}[|X|^{2/3}] < \infty$, we have*

$$\text{MISE}\left[\hat{F}^F_{N,n}\right] = O\left(N^{-r+2}\right) + O\left(\frac{N^{5/2}}{n}\right).$$

X is again a random variable with density function f. End Theorem

Setting $N^{-r+2} = \frac{N^{5/2}}{n}$ and solving the equation for N, the next corollary follows.

Corollary 7.5. *Suppose $f \in L_2$ is r times continuously differentiable and $\left(x - \frac{d}{dx}\right)^r f \in L_2$ where $r > 2$ with $\mathbb{E}[|X|^{2/3}] < \infty$. The asymptotically optimal N for estimating F with respect to MISE is*

$$N_{opt} \sim n^{2/(2r+1)},$$

which leads to

$$\text{MISE}\left[\hat{F}^F_{N,n}\right] = O\left(n^{-\frac{2(r-2)}{2r+1}}\right) = o(1)$$

uniformly in x. End Corollary

As in the $[0, \infty)$ case, the N in the theorems above depends on n. This is not the case in the algorithm in Section 7.5. Here, we cannot compare the MSE and the MISE of $\hat{F}^F_{N,n}$ to MISE $\left[\hat{f}_{N,n}\right]$. But Eq. (7.20) implies that for large and fixed N and $r \gg 1$, the Gauss-Hermite distribution estimator is approximately unbiased with bias $O(N^{-r/2+1})$. Eq. (7.21) says that the variance is $O\left(\frac{N^{5/2}}{n}\right)$, which goes to zero for $n \to \infty$. Hence, for fixed and large N and $r \gg 1$, the MSE approximately approaches zero for $n \to \infty$. The same follows for the MISE.

We now focus on the almost sure convergence of the estimator.

7.4.3 Almost Sure Convergence

The next theorem gives the asymptotic behavior of the difference between the estimator and the real distribution function F.

Theorem 7.9. *Suppose $f \in L_2$ is r times continuously differentiable and $\left(x - \frac{d}{dx}\right)^r f \in L_2$ where $r > 2$. If $\mathbb{E}[|X|^s] < \infty, s > \frac{8(r+1)}{3(2r+1)}$ and $N(n) \sim n^{\frac{2}{2r+1}}$, then we have*

$$\left|\hat{F}^F_{N,n}(x) - F(x)\right| = O\left(n^{-\frac{r-2}{2r+1}} \log n\right) \quad a.s.$$

uniformly in x and

$$\left|\hat{F}^F_{N,n}(x) - F(x)\right| \to 0 \;\; a.s.$$

uniformly in x. End Theorem

For the proof, see Proofs Hermite.

The condition $s > \frac{8(r+1)}{3(2r+1)}$ that has to hold here is always satisfied for $s > \frac{16}{9}$ as stated in [44, Remark 2].

The next theorem shows the almost sure convergence when $N(n)$ is a random variable with values in $\mathbb{N}$. $N(n)$ could be a measurable function of $X_i, i = 1, .., n$, and thus the theorem could be applied for a data-driven estimator of $N(n)$ under certain conditions. One such condition is that $\frac{N(n)}{n^{\frac{2}{2r+1}}} \to 0$ with $r > 0$.

Theorem 7.10. *Suppose $f \in L_2$ is r times continuously differentiable and $\left(x - \frac{d}{dx}\right)^r f \in L_2$ where $r > 2$. If $N(n) \to \infty$ a.s. and $\sum_{n=1}^{\infty} \mathbb{P}\left(\frac{N(n)}{n^\gamma} > \epsilon\right) < \infty, \epsilon > 0$, i.e., $\frac{N(n)}{n^\gamma} \to 0$ a.s. with $0 < \gamma < 6/17$, we get*

$$\left|\hat{F}^F_{N,n}(x) - F(x)\right| \to 0 \ \ a.s.$$

uniformly in x. This applies specifically if $\frac{N(n)}{n^{\frac{2}{2r+1}}} \to 0$ a.s. End Theorem

For the proof, see Proofs Hermite.

We now take a closer look at the asymptotic behavior of the estimator.

7.4.4 Asymptotic Behavior

As for the Hermite estimator on the real half line, we now study the asymptotic behavior of the estimator and establish that the limit distribution is normal again.

Theorem 7.11. *For $x \in (-\infty, \infty)$ with $0 < F(x) < 1$, we get that*

$$\sqrt{n}\left(\hat{F}^F_{N,n}(x) - \mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right)$$

for $n \to \infty$ if f is differentiable in x. End Theorem

For the proof, see Proofs Hermite.

The next corollary deals with the asymptotic behavior of $\hat{F}^F_{N,n}(x) - F(x)$. With Lemma 7.5, it is easy to see that

$$n^{1/2}\left(\hat{F}^F_{N,n}(x) - F(x)\right) = n^{1/2}\left(\hat{F}^F_{N,n}(x) - \mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]\right) + O_x(n^{1/2}N^{1-r/2}). \tag{7.12}$$

This leads directly to the following corollary.

Corollary 7.6. *Let $m, n \to \infty$. Then, for $x \in (-\infty, \infty)$ with $0 < F(x) < 1$, it holds with the assumptions of Lemma 7.5 that*

$$n^{1/2}\left(\hat{F}^F_{N,n}(x) - F(x)\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right)$$

for $n^{-1/2}N^{r/2-1} \to \infty$. End Corollary

We now talk about the robustness of the estimator that can be measured with the so-called influence function.

7.4.5 Robustness

A very important aspect of an estimator is its performance if there exists contamination of the data, i.e., outlying observations. One way to measure this is the influence function (see [48]) that gives the effect of an infinitesimal contamination at a particular point, x', on the estimator, standardized by the mass of the contamination (see [49]).

Assume a functional $T(x, F)$, evaluated at a point x and distribution F, is given. Then the influence function is defined by

$$\mathrm{IF}(x, x'; T, F) = \lim_{\epsilon \to 0} \frac{T(x, (1-\epsilon)F + \epsilon\delta_{x'}) - T(x, F)}{\epsilon}.$$

The distribution $\delta_{x'}$ is defined by $\mathbb{P}(Y = x') = 1$ for a random variable $Y \sim \delta_{x'}$.

The empirical influence function is defined by

$$\mathrm{IF}(x, x'; T, F_n) = \lim_{\epsilon \to 0} \frac{T(x, (1-\epsilon)F_n + \epsilon\delta_{x'}) - T(x, F_n)}{\epsilon},$$

where F_n is the EDF. For a linear functional, the equality

$$\lim_{n\to\infty} \mathrm{IF}(x, x'; T, F_n) = \mathrm{IF}(x, x'; T, F) \text{ a.s.}$$

holds. This is not true in general.

Now, the gross-error sensitivity, defined by

$$\sup_{x'} |\mathrm{IF}(x, x'; T, F)|,$$

gives an upper bound on the asymptotic bias of the estimator caused by contamination by outliers. If this number is finite, the maximal influence on the value of the estimator that an outlier at x' can have, is bounded. If the estimator is of the form $T(x, F_n)$, it is said to be bias-robust or B-robust if

$$\sup_{x'} |\mathrm{IF}(x, x'; T, F)| < \infty \text{ and } \sup_{x'} |\mathrm{IF}(x, x'; T, F_n)| < \infty$$

for all $n \in \mathbb{N}$.

In the sequel, we give bias-robustness results for the Hermite estimator for finite N, the smooth kernel distribution function estimator, and the Gram-Charlier A distribution function estimator given in Eq. (7.7). The results can be found in [40].

Lemma 7.7. *The Hermite distribution estimator Eq. (7.10) for fixed N is bias-robust.* End Lemma

For the proof, see Proofs Hermite.

Lemma 7.8. *The kernel distribution estimator $F_{h,n}(x)$ that is defined in Eq. (3.1) is bias-robust.* End Lemma

For the proof, see Proofs Hermite.

Lemma 7.9. *The Gram-Charlier A distribution function estimator given in Eq. (7.7) is not bias-robust.* End Lemma

For the proof, see Proofs Hermite.

We now quickly interpret the last three lemmas. The bias-robustness of the Hermite estimator was only shown for finite N and not for $N \to \infty$. Thus, the result here is weaker than that for the kernel estimator where we proved bias-robustness for all h including $h \to 0$. The last lemma, however, shows for a closely related distribution estimator with a finite number of terms that the property shown in Lemma 7.7 is not trivial.

The next section gives an algorithm for sequential calculation. For the considered estimators, N is fixed and constant.

7.5 Algorithm for Sequential Calculation

Following [39], we now give the definition of an algorithm that outputs the distribution functions in Eq. (7.8), Eq. (7.10) and Eq. (7.11) after getting a new sample without the need for recalculating the whole estimator. This is possible because the coefficients $\hat{a}_k$ in Eq. (7.4) can be updated with each new observation without recalculating the entire sum. The resulting algorithm has the following form.

1. For $k = 0, ..., N$, initialize $\hat{a}_k^{(1)} = h_k(X_1)$, where X_1 is the first observation.
2. For a new observation X_i, update $\hat{a}_0^{(i-1)}, ..., \hat{a}_N^{(i-1)}$ with
$$\hat{a}_k^{(i)} = \frac{1}{i}\left[(i-1)\hat{a}_k^{(i-1)} + h_k(X_i)\right].$$
3. Plug these updated coefficients into Eq. (7.5) Eq. (7.8), Eq. (7.10) and Eq. (7.11) to get the updated estimates.

The second step that updates the coefficients is constant $O(1)$ and does not depend on the number of previous observations. As both the density and the distribution function do not explicitly depend on the observations, updating them is also $O(1)$.

We used a fixed and constant N, which means that the density function estimator is biased. This leads to a biased distribution function estimator that is, however, sequential.

We now talk about a way to improve the Hermite estimator.

7.6 Standardizing

It is intuitive that the truncated Gauss-Hermite estimator yields better results if applied to standardized random variables. This means that we transform the random variable X with mean μ and standard deviation σ to
$$\tilde{X} = \frac{X - \mu}{\sigma}.$$

There are two approaches to use the Hermite estimator on the standardized data. The first one is to change the data and adapt the distribution and density function. Then, this new distribution function is estimated. The second approach is to change the estimator. In this case, the true distribution function to be estimated stays the same. In the present case this means that the Hermite polynomials are shifted and scaled appropriately.

The idea for the following adaptions that are also used in the simulation was gained in [50].

The first approach is used for the Hermite estimator on the real line. The density function transforms to
$$\tilde{f}(\tilde{x}) = \sigma f(\sigma\tilde{x} + \mu)$$
and the distribution function to
$$\tilde{F}(\tilde{x}) = F(\sigma\tilde{x} + \mu).$$
These are the new functions to be estimated.

For the Hermite estimator on the real half line we use the second approach. In this case we only scale and do not shift, i.e., $\tilde{X} = \frac{X-\mu}{\sigma}$ but set $\mu = 0$ later on, to make sure that no negative data arises that the estimator could not deal with. Note that if the support is far away from zero, some shifting could still be necessary. This holds for example for the power-law distribution in Section 7.3.3. In most cases, x_{min} is known so that the shifting can be done by setting x_{min} to the desired value. The truncated density gets the form
$$\tilde{f}(\tilde{x}) = \sum_{k=0}^{N} a_k h_k(\tilde{x}),$$

where

$$a_k = \int\limits_{-\infty}^{\infty} h_k(\tilde{x})\tilde{f}(\tilde{x})d\tilde{x} = \int_{-\infty}^{\infty} h_k(\tilde{x}) \cdot \sigma \cdot f(\sigma\tilde{x} + \mu)d\tilde{x} = \int_{-\infty}^{\infty} h_k\left(\frac{z-\mu}{\sigma}\right) \cdot f(z)dz.$$

Hence, the ideal estimated coefficient would be

$$\hat{a}_k = \hat{a}_k(\mu, \sigma) = \frac{1}{n}\sum_{i=1}^{n} h_k\left(\frac{X_i - \mu}{\sigma}\right),$$

but in general, mean, and variance are unknown. An alternative is to use estimators that lead to a biased coefficient $\hat{a}_k$. Still, standardizing improves the quality of the fit in many cases. Then, the original truncated density is of the form

$$f(\tilde{x}) = \frac{1}{\sigma}\tilde{f}\left(\frac{\tilde{x}-\mu}{\sigma}\right) = \frac{1}{\sigma}\sum_{k=0}^{N} a_k h_k\left(\frac{\tilde{x}-\mu}{\sigma}\right)$$

and the estimator is

$$\hat{f}_{N,n}(\tilde{x}) = \frac{1}{\sigma}\sum_{k=0}^{N} \hat{a}_k h_k\left(\frac{\tilde{x}-\mu}{\sigma}\right).$$

Integrating this expression leads to

$$\begin{aligned}
&\hat{F}^H_{N,n}(x)\\
&= \int\limits_0^x \hat{f}_{N,n}(t)\,\mathrm{d}t\\
&= \frac{1}{\sigma}\sum_{k=0}^{N}\sqrt{\alpha_k}\hat{a}_k k! \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m 2^{k-2m}}{m!(k-2m)!\sqrt{2\pi}} \int\limits_0^x \left(\frac{\tilde{x}-\mu}{\sigma}\right)^{k-2m} e^{\frac{-\left(\frac{\tilde{x}-\mu}{\sigma}\right)^2}{2}} d\tilde{x}\\
&= \begin{cases} \frac{1}{\sigma}\sum_{k=0}^{N}\hat{a}_k k! \sum_{m=0}^{\lfloor k/2 \rfloor} W_1(k,m)\left(\gamma_1(k,m) - \gamma_2(k,m)\right), & x \le \mu, \\ \frac{1}{\sigma}\sum_{k=0}^{N}\hat{a}_k k! \sum_{m=0}^{\lfloor k/2 \rfloor} W_2(k,m)\Big((-1)^{k-2m}\gamma_1(k,m) + \gamma_2(k,m)\Big), & x > \mu, \end{cases}
\end{aligned}$$

where

$$W_1(k,m) = \frac{(-1)^{k-m}2^{k-3m-\frac{1}{2}}}{m!(k-2m)!\pi^{\frac{1}{4}}},\ W_2(k,m) = \frac{(-1)^m 2^{k-3m-\frac{1}{2}}}{m!(k-2m)!\pi^{\frac{1}{4}}},$$

and

$$\gamma_1(k,m) = \gamma\left(-m + \frac{k}{2} + \frac{1}{2}, \frac{\left(\frac{\mu}{\sigma}\right)^2}{2}\right),\ \gamma_2(k,m) = \gamma\left(-m + \frac{k}{2} + \frac{1}{2}, \frac{\left(\frac{x-\mu}{\sigma}\right)^2}{2}\right).$$

We now show how it is possible to update estimates of the mean μ and the variance σ^2 with every new sample.

The mean can be calculated by

$$\begin{aligned}
\hat{\mu}_1 &= X_1\\
\hat{\mu}_k &= \frac{1}{k}\left((k-1)\,\hat{\mu}_{k-1} + X_k\right), k \ge 2.
\end{aligned}$$

With [51], the following algorithm can be used to estimate the standard deviation.

$$
\begin{aligned}
M_1 &= X_1, \\
S_1 &= 0, \\
M_k &= M_{k-1} + \frac{x_k - M_{k-1}}{k}, \\
S_k &= S_{k-1} + (x_k - M_{k-1})(x_k - M_k), \\
\hat{\sigma}_k &= \sqrt{\frac{S_k}{k-1}}, k \geq 2.
\end{aligned}
$$

These are online algorithms, which means that the standardizing must not be seen as a preprocessing step but rather as a part of the online algorithm. The need of a preprocessing step would make the online use of the estimator impossible.

7.7 Proofs Hermite

We need the following lemmas for the proofs of Section 7.

Lemma 7.10. *This lemma follows from [52, Lemma 1] which says that in every point x where $f(x)$ is differentiable,*

$$\lim_{N\to\infty}\sum_{k=0}^{N} a_k h_k(x) = f(x)$$

holds. For $f \in L_p, p > 1$, this convergence holds for almost all $x \in \mathbb{R}$. End Lemma

Lemma 7.11. *With [44, Eq. (12)], it holds that*

$$\sum_{k=0}^{N} \mathbb{E}\left[(\hat{a}_k - a_k)^2\right] = O\left(\frac{N^{1/2}}{n}\right)$$

for $\mathbb{E}[|X|^{2/3}] < \infty$. End Lemma

Lemma 7.12. *As can be found in the proof of [44, Theorem 4], it can be shown that*

$$\sum_{k=0}^{N}(\hat{a}_k - a_k)^2 = O(n^{-2r/(2r+1)}\log n) \ \ a.s.$$

for $\mathbb{E}[|X|^s] < \infty$, $s > 8(r+1)/3(2r+1)$. End Lemma

The next lemma follows from [40, Lemma 4] and is used a lot in the sequel.

Lemma 7.13. *It holds that*

$$\int_{-\infty}^{x} |h_k(t)| \,\mathrm{d}t \le 2c_1(k+1)^{-1/4} + 12d_1(k+1)^{1/2}$$

with positive constants c_1 and d_1. End Lemma

Proof. This proof follows the proof of [40, Lemma 4]. We know with [43, Theorem 8.91.3] that

$$\max_{|x|\le a} h_k(x) \le c_a(k+1)^{-1/4},$$
$$\max_{|x|\ge a} x^{\lambda}|h_k(x)| \le d_a(k+1)^{s}$$

for positive constants c_a, d_a depending only on a and $s = \max\left(\frac{\lambda}{2} - \frac{1}{12}, -\frac{1}{4}\right)$, where $\lambda = 1 + b, b > 0$. It follows that

$$\begin{aligned}
\int_{-\infty}^{x} |h_k(t)| \,\mathrm{d}t &\le \int_{-\infty}^{\infty} |h_k(t)| \,\mathrm{d}t \\
&= \int_{-\infty}^{-1} |h_k(t)| \,\mathrm{d}t + \int_{-1}^{1} |h_k(t)| \,\mathrm{d}t + \int_{1}^{\infty} |h_k(t)| \,\mathrm{d}t \\
&= \int_{-1}^{1} |h_k(t)| \,\mathrm{d}t + 2\int_{1}^{\infty} |h_k(t)| \,\mathrm{d}t \\
&\le 2c_1(k+1)^{-1/4} + 2d_1(k+1)^{5/12+b/2}\int_{1}^{\infty} t^{-1-b} \,\mathrm{d}t \\
&= 2c_1(k+1)^{-1/4} + 2\frac{d_1}{b}(k+1)^{5/12+b/2},
\end{aligned}$$

where the last equation follows from $\int_1^\infty t^{-1-b}\,\mathrm{d}t = \frac{1}{b}, b > 0$. To get the desired result, we set $b = \frac{1}{6}$. End Proof

A very similar lemma holds for the $[0,\infty)$ case.

Lemma 7.14. *It holds that*

$$\int_0^x |h_k(t)|\,\mathrm{d}t \le 2xc_x(k+1)^{-1/4}$$

with a positive constant c_x depending on x. End Lemma

Proof. With the inequalities from the lemma above we know that

$$\begin{aligned}\int_0^x |h_k(t)|\,\mathrm{d}t &\le \int_{-x}^x |h_k(t)|\,\mathrm{d}t \\ &\le \int_{-x}^x c_x(k+1)^{-1/4}\,\mathrm{d}t \\ &= 2xc_x(k+1)^{-1/4}\end{aligned}$$

holds. End Proof

Now, we turn our attention to the proofs of this section.

Proof of Lemma 7.1. This proof is similar to the proof of [40, Proposition 2]. It holds with the Cauchy-Schwarz inequality that

$$\begin{aligned}\left|\hat{F}^H_{N,n}(x) - \mathbb{E}\left[\hat{F}^H_{N,n}(x)\right]\right|^2 &= \left|\int_0^x \sum_{k=0}^N \hat{a}_k h_k(t)\,\mathrm{d}t - \int_0^x \sum_{k=0}^N a_k h_k(t)\,\mathrm{d}t\right|^2 \\ &\le \sum_{k=0}^N (\hat{a}_k - a_k)^2 \cdot \sum_{k=0}^N \left|\int_0^x h_k(t)\,\mathrm{d}t\right|^2.\end{aligned}$$

For the second term we know with Lemma 7.14 that

$$\begin{aligned}\sum_{k=0}^N \left|\int_0^x h_k(t)\,\mathrm{d}t\right|^2 &\le \sum_{k=0}^N \left(\int_{-x}^x |h_k(t)|\,\mathrm{d}t\right)^2 \\ &\le \sum_{k=0}^N \left(2xc_x(k+1)^{-1/4}\right)^2 \\ &= O_x\left(\sum_{k=0}^N (k+1)^{-1/2}\right) \\ &= O_x(N).\end{aligned}$$

We get

$$\mathbb{E}\left[\left|\hat{F}^H_{N,n}(x) - \mathbb{E}\left[\hat{F}^H_{N,n}(x)\right]\right|^2\right] \le O_x(N) \sum_{k=0}^N \mathbb{E}\left[(\hat{a}_k - a_k)^2\right] \tag{7.13}$$

$$= O_x\left(\frac{N^{3/2}}{n}\right) \tag{7.14}$$

with Lemma 7.11. End Proof

Proof of Lemma 7.2. This proof follows the proof of [40, Proposition 1]. We know with Lemma 7.10 and $\mathbb{E}[\hat{a}_k] = a_k$ that

$$\begin{aligned}\left|\mathbb{E}\left[\hat{F}^H_{N,n}(x)\right] - F(x)\right| &= \left|\mathbb{E}\left[\int_0^x \sum_{k=0}^N \hat{a}_k h_k(t)\,\mathrm{d}t\right] - \int_0^x \sum_{k=0}^\infty a_k h_k(t))\,\mathrm{d}t\right| \\ &\le \int_0^x \sum_{k=N+1}^\infty |a_k h_k(t)|\,\mathrm{d}t \\ &= \sum_{k=N+1}^\infty |a_k| \int_0^x |h_k(t)|\,\mathrm{d}t.\end{aligned}$$

With Lemma 7.14, it holds that

$$\begin{aligned}\sum_{k=N+1}^\infty |a_k| \int_0^x |h_k(t)|\,\mathrm{d}t &\le 2xc_x \sum_{k=N+1}^\infty |a_k|(k+1)^{-1/4} \\ &\le 2xc_x \sum_{k=N+1}^\infty |b_{k+r}|(k+1)^{-1/4-r/2},\end{aligned}$$

where $a_k^2 \le \frac{b_{k+r}^2}{(k+1)^r}$ follows from [41, Theorem 1] and b_k is the k-th coefficient of the expansion of $\left(x - \frac{d}{\mathrm{d}x}\right)^r f(x)$ which is in L_2 by assumption. Hence, $\left\|\left(x - \frac{d}{\mathrm{d}x}\right)^r f(x)\right\|^2 = \sum_{k=0}^\infty b_k^2$ by Parseval's theorem. We know with the Cauchy-Schwarz inequality that

$$\begin{aligned}\sum_{k=N+1}^\infty |b_{k+r}|(k+1)^{-1/4-r/2} &\le \sqrt{\sum_{k=N+1}^\infty b_{k+r}^2} \cdot \sqrt{\sum_{k=N+1}^\infty (k+1)^{-1/2-r}} \\ &\le \left\|\left(x - \frac{d}{\mathrm{d}x}\right)^r f(x)\right\| \cdot \sqrt{\sum_{k=N+1}^\infty (k+1)^{-1/2-r}}\end{aligned}$$

and it follows that

$$\sum_{k=N+1}^\infty |a_k| \int_{-\infty}^x |h_k(t)|\,\mathrm{d}t \le 2xc_x \cdot \left\|\left(x - \frac{d}{\mathrm{d}x}\right)^r f(x)\right\| \cdot \sqrt{\sum_{k=N+1}^\infty (k+1)^{-1/2-r}}.$$

With [53, 25.11.43] we know with constants $e_i, i = 1, 2, 3$, that

$$\begin{aligned}\sum_{k=N+1}^\infty (k+1)^{-1/2-r} &= \sum_{k=0}^\infty (k+N+2)^{-1/2-r} = \zeta(\frac{1}{2} + r, N+2) \\ &\le \sum_{k=1}^\infty e_1 N^{1/2-r-2k} + e_2 N^{1/2-r} + \frac{1}{2} N^{-1/2-r} \\ &= e_1 \frac{N^{1/2-r}}{N^2 - 1} + e_2 N^{1/2-r} + \frac{1}{2} N^{-1/2-r} \\ &\le e_3 N^{-r+1/2},\end{aligned}$$

where ζ is the Zeta-function and hence

$$\left|\mathbb{E}\left[\hat{F}^H_{N,n}(x)\right] - F(x)\right| \le \sum_{k=N+1}^\infty |a_k| \int_0^x |h_k(t)|\,\mathrm{d}t = O_x\left(N^{-r/2+1/4}\right), \tag{7.15}$$

which completes the proof. End Proof

Proof of Theorem 7.5. We follow [40, Theorem 3]. Note that

$$\left|\hat{F}^H_{N,n}(x) - F(x)\right| \leq \left|\mathbb{E}\left[\hat{F}^H_{N,n}(x)\right] - F(x)\right| + \left|\hat{F}^H_{N,n}(x) - \mathbb{E}\left[\hat{F}^H_{N,n}(x)\right]\right|.$$

We use results from the full Hermite estimator here. They give asymptotic behavior that is worse than what we would get from the half Hermite estimator but are uniform in x. For the first part it holds with the proof of Lemma 7.5 and $N(n) \sim n^{\frac{2}{2r+1}}$ that

$$\left|\mathbb{E}\left[\hat{F}^H_{N,n}(x)\right] - F(x)\right| \leq \int_0^x \sum_{k=N+1}^{\infty} |a_k h_k(t)| \,\mathrm{d}t \leq \int_{-\infty}^x \sum_{k=N+1}^{\infty} |a_k h_k(t)| \,\mathrm{d}t = O\left(N^{-r/2+1}\right) = O\left(n^{-\frac{r-2}{2r+1}}\right).$$

The second term can be written as

$$\left|\hat{F}^H_{N,n}(x) - \mathbb{E}\left[\hat{F}^H_{N,n}(x)\right]\right|^2 \leq \sum_{k=0}^{N} (\hat{a}_k - a_k)^2 \cdot \sum_{k=0}^{N} \left|\int_0^x h_k(t) \,\mathrm{d}t\right|^2 \leq \sum_{k=0}^{N} (\hat{a}_k - a_k)^2 \cdot \sum_{k=0}^{N} \left(\int_{-\infty}^x |h_k(t)| \,\mathrm{d}t\right)^2,$$

where we used part of the proof of Lemma 7.1. It follows that

$$\left|\hat{F}^H_{N,n}(x) - \mathbb{E}\left[\hat{F}^H_{N,n}(x)\right]\right| = O(N)\sqrt{\sum_{k=0}^{N} (\hat{a}_k - a_k)^2} = O(N) \cdot O(n^{-r/(2r+1)} \log n) = O(n^{-\frac{r-2}{2r+1}} \log n) \text{ a.s.},$$

where we used a result from the proof of Lemma 7.6 and Lemma 7.12. The theorem follows directly. End Proof

Proof of Theorem 7.6. This proof takes some ideas from the proofs of [17, Theorem 2]. For fixed N it holds that

$$\begin{aligned}
\hat{F}^H_{N,n}(x) - \mathbb{E}\left[\hat{F}^H_{N,n}(x)\right] &= \int_0^x \sum_{k=0}^{N} \hat{a}_k h_k(t) \,\mathrm{d}t - \int_0^x \sum_{k=0}^{N} a_k h_k(t) \,\mathrm{d}t \\
&= \int_0^x \sum_{k=0}^{N} \left[\frac{1}{n}\sum_{i=1}^{n} h_k(X_i)\right] h_k(t) \,\mathrm{d}t - \int_0^x \sum_{k=0}^{N} a_k h_k(t) \,\mathrm{d}t \\
&= \frac{1}{n}\sum_{i=1}^{n} \left[\int_0^x T_N(X_i, t) \,\mathrm{d}t - \int_0^x \sum_{k=0}^{N} a_k h_k(t) \,\mathrm{d}t\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_{i,N},
\end{aligned}$$

where

$$T_N(x, y) = \sum_{k=0}^{N} h_k(x) h_k(y)$$

and

$$Y_{i,N} = \int_0^x \left[T_N(X_i, t) - \sum_{k=0}^{N} a_k h_k(t)\right] \mathrm{d}t, i \in \{1, ..., n\}.$$

The $Y_{i,N}$ are i.i.d. random variables with mean 0. Define $\gamma_N^2 = \mathbb{E}[Y_{1,N}^2]$. We use the central limit theorem for double arrays (see [33], Section 1.9.3) to show the claim.

Defining

$$A_n = \mathbb{E}\left[\sum_{i=1}^{n} Y_{i,N}\right] = 0 \text{ and } B_n^2 = \mathrm{Var}\left[\sum_{i=1}^{n} Y_{i,N}\right] = n\gamma_N^2,$$

it says that

$$\frac{\sum_{i=1}^{n} Y_{i,N} - A_n}{B_n} \xrightarrow{D} \mathcal{N}(0,1)$$

if and only if the Lindeberg condition

$$\frac{n\mathbb{E}[\mathbb{I}(|Y_{1,N}| > \epsilon B_n) Y_{1,N}^2]}{B_n^2} \to 0 \text{ for } n \to \infty \text{ and all } \epsilon > 0$$

is satisfied. It holds for $n \to \infty$ that

$$\begin{aligned}
\frac{\sum_{i=1}^{n} Y_{i,N} - A_n}{B_n} &\xrightarrow{D} \mathcal{N}(0,1) \\
\Leftrightarrow \frac{\sum_{i=1}^{n} Y_{i,N}}{\sqrt{n}\cdot\gamma_N} &\xrightarrow{D} \mathcal{N}(0,1) \\
\Leftrightarrow \frac{\sqrt{n}}{\gamma_N}\left(\hat{F}_{N,n}^H(x) - \mathbb{E}\left[\hat{F}_{N,n}^H(x)\right]\right) &\xrightarrow{D} \mathcal{N}(0,1) \\
\Leftrightarrow \sqrt{n}\left(\hat{F}_{N,n}^H(x) - \mathbb{E}\left[\hat{F}_{N,n}^H(x)\right]\right) &\xrightarrow{D} \mathcal{N}\left(0,\sigma^2(x)\right),
\end{aligned}$$

which is the claim of Theorem 7.6. The last equivalence holds because of the following. We have to calculate γ_N^2 which is given by

$$\begin{aligned}
\gamma_N^2 &= \mathbb{E}\left[\left(\int_0^x T_N(X_1,t)\,\mathrm{d}t - \int_0^x \sum_{k=0}^{N} a_k h_k(t)\,\mathrm{d}t\right)^2\right] \\
&= \mathbb{E}\left[\left(\int_0^x T_N(X_1,t)\,\mathrm{d}t\right)^2\right] - 2\int_0^x \sum_{k=0}^{N} a_k h_k(t)\,\mathrm{d}t \cdot \mathbb{E}\left[\int_0^x T_N(X_1,t)\,\mathrm{d}t\right] + \left(\int_0^x \sum_{k=0}^{N} a_k h_k(t)\,\mathrm{d}t\right)^2.
\end{aligned} \tag{7.16}$$

The first part is the only part where we do not know the asymptotic behavior. Hence, we now take a closer look at this part. With [54, Eq. (A8)], which only holds on compact sets, we know that

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^x T_N(X_1,t)\,\mathrm{d}t\right)^2\right] &= \lim_{P\to\infty}\int_0^P \left[\int_0^x \frac{\sin(M(r-t))}{\pi(r-t)} + O(N^{-1/2})\,\mathrm{d}t\right]^2 f(r)\,\mathrm{d}r \\
&= \int_0^\infty \left[\int_0^x \frac{\sin(M(r-t))}{\pi(r-t)} + O(N^{-1/2})\,\mathrm{d}t\right]^2 f(r)\,\mathrm{d}r \\
&= \int_0^x \left[\int_0^x \frac{\sin(M(r-t))}{\pi(r-t)} + O(N^{-1/2})\,\mathrm{d}t\right]^2 f(r)\,\mathrm{d}r + \int_x^\infty \left[\int_0^x \frac{\sin(M(r-t))}{\pi(r-t)} + O(N^{-1/2})\,\mathrm{d}t\right]^2 f(r)\,\mathrm{d}r,
\end{aligned} \tag{7.17}$$

where $M = \frac{\sqrt{2n+3}+\sqrt{2n+1}}{2}$. The inner integral can be written as

$$\int_0^x \frac{\sin(M(r-t))}{\pi(r-t)}\,\mathrm{d}t = \int_{M(r-x)}^{Mr} \frac{\sin(l)}{\pi l}\,\mathrm{d}l$$

and with the fact that for $M \to \infty$, we get

$$\int_{Ma}^{Mb} \frac{\sin(l)}{\pi l} dl \to \begin{cases} 1, & a < 0 < b, \\ 0, & 0 < a < b, \\ 0, & a < b < 0, \end{cases}$$

it follows with Eq. (7.17) for $n \to \infty$ (which implies $M \to \infty$) that

$$\mathbb{E}\left[\left(\int_0^x T_N(X_1,t)\,\mathrm{d}t\right)^2\right] \to \int_0^x f(r)\,\mathrm{d}r = F(x). \tag{7.18}$$

In Section 12.2, it is explained in detail why it is possible to move the limit $M \to \infty$ inside the integral. Then, plugging Eq. (7.18) in Eq. (7.16) and using the fact that we know limits of the other parts from Lemma 7.10, it holds for $n \to \infty$ that

$$\gamma_N^2 \to F(x) - 2F(x)^2 + F(x)^2 = \sigma^2(x). \tag{7.19}$$

Now, we have to show that asymptotic normality actually holds. In our case the Lindeberg condition has the form

$$\frac{\mathbb{E}\left[\mathbb{I}(|Y_{1,N}| > \epsilon\sqrt{n}\gamma_N)Y_{1,N}^2\right]}{\gamma_N^2} \to 0 \text{ for } n \to \infty \text{ and all } \epsilon > 0.$$

This is what has to be shown to prove the theorem. Writing the expected value as an integral, we get

$$\int_0^\infty \mathbb{I}\left(\left|\int_0^x \left[T_N(r,t) - \sum_{k=0}^N a_k h_k(t)\right]\mathrm{d}t\right| > \epsilon\sqrt{n}\gamma_N\right)\left(\int_0^x \left[T_N(r,t) - \sum_{k=0}^N a_k h_k(t)\right]\mathrm{d}t\right)^2 f(r)\,\mathrm{d}r.$$

With the arguments from above, the left side of the inequality in the indicator function is bounded by a constant, depending on x, for large n. Using this result, we get for large n that

$$\frac{\mathbb{E}\left[\mathbb{I}(|Y_{1,N}| > \epsilon\sqrt{n}\gamma_N)Y_{1,N}^2\right]}{\gamma_N^2} \le \mathbb{I}(c_x > \epsilon\sqrt{n}\gamma_N)\frac{\mathbb{E}\left[Y_{1,N}^2\right]}{\gamma_N^2} = \mathbb{I}\left(\frac{c_x}{\sqrt{n}\gamma_N} > \epsilon\right) \to 0,$$

where c_x is a constant depending on x, which proves the claim. End Proof

Proof of Lemma 7.5. This proof follows the proof of [40, Proposition 1]. We know with Lemma 7.10, $\mathbb{E}[\hat{a}_k] = a_k$ that

$$\begin{aligned}\left|\mathbb{E}\left[\hat{F}_{N,n}^F(x)\right] - F(x)\right| &= \left|\mathbb{E}\left[\int_{-\infty}^x \sum_{k=0}^N \hat{a}_k h_k(t)\,\mathrm{d}t\right] - \int_{-\infty}^x \sum_{k=0}^\infty a_k h_k(t))\,\mathrm{d}t\right| \\ &\le \int_{-\infty}^x \sum_{k=N+1}^\infty |a_k h_k(t)|\,\mathrm{d}t \\ &= \sum_{k=N+1}^\infty |a_k| \int_{-\infty}^x |h_k(t)|\,\mathrm{d}t.\end{aligned}$$

With Lemma 7.13, it holds that

$$\begin{aligned}\sum_{k=N+1}^\infty |a_k| \int_{-\infty}^x |h_k(t)|\,\mathrm{d}t &\le 2c_1 \sum_{k=N+1}^\infty |a_k|(k+1)^{-1/4} + 12d_1 \sum_{k=N+1}^\infty |a_k|(k+1)^{1/2} \\ &\le 2c_1 \sum_{k=N+1}^\infty |b_{k+r}|(k+1)^{-1/4-r/2} + 12d_1 \sum_{k=N+1}^\infty |b_{k+r}|(k+1)^{1/2-r/2},\end{aligned}$$

where $a_k^2 \leq \frac{b_{k+r}^2}{(k+1)^r}$ follows from [41, Theorem 1] and b_k is the k-th coefficient of the expansion of $\left(x - \frac{d}{dx}\right)^r f(x)$ which is in L_2 by assumption. Hence, $\left\|\left(x - \frac{d}{dx}\right)^r f(x)\right\|^2 = \sum_{k=0}^{\infty} b_k^2$ by Parseval's theorem. We know with the Cauchy-Schwarz inequality that

$$\begin{aligned}\sum_{k=N+1}^{\infty} |b_{k+r}|(k+1)^{-1/4-r/2} &\leq \sqrt{\sum_{k=N+1}^{\infty} b_{k+r}^2} \cdot \sqrt{\sum_{k=N+1}^{\infty} (k+1)^{-1/2-r}} \\ &\leq \left\|\left(x - \frac{d}{dx}\right)^r f(x)\right\| \sqrt{\sum_{k=N+1}^{\infty} (k+1)^{-1/2-r}}\end{aligned}$$

and it follows that

$$\begin{aligned}\sum_{k=N+1}^{\infty} |a_k| \int_{-\infty}^{x} |h_k(t)|\, dt \leq{}& 2c_1 \left\|\left(x - \frac{d}{dx}\right)^r f(x)\right\| \sqrt{\sum_{k=N+1}^{\infty} (k+1)^{-1/2-r}} \\ &+ 12d_1 \left\|\left(x - \frac{d}{dx}\right)^r f(x)\right\| \sqrt{\sum_{k=N+1}^{\infty} (k+1)^{1-r}}.\end{aligned}$$

With [53, 25.11.43] we know with constants $e_i, i = 1, ..., 6$, that

$$\begin{aligned}\sum_{k=N+1}^{\infty} (k+1)^{-1/2-r} &= \sum_{k=0}^{\infty}(k+N+2)^{-1/2-r} = \zeta\left(\frac{1}{2} + r, N+2\right) \\ &\leq \sum_{k=1}^{\infty} e_1 N^{1/2-r-2k} + e_2 N^{1/2-r} + \frac{1}{2} N^{-1/2-r} \\ &= e_1 \frac{N^{1/2-r}}{N^2-1} + e_2 N^{1/2-r} + \frac{1}{2} N^{-1/2-r} \\ &\leq e_3 N^{-r+2},\end{aligned}$$

and

$$\begin{aligned}\sum_{k=N+1}^{\infty} (k+1)^{1-r} &= \sum_{k=0}^{\infty}(k+N+2)^{-1/2-r} = \zeta(r-1, N+2) \\ &\leq \sum_{k=1}^{\infty} e_4 N^{2-r-2k} + e_5 N^{2-r} + \frac{1}{2} N^{1-r} \\ &= e_4 \frac{N^{2-r}}{N^2-1} + e_5 N^{2-r} + \frac{1}{2} N^{1-r} \\ &\leq e_6 N^{-r+2},\end{aligned}$$

where ζ is the Zeta-function and hence

$$\left|\mathbb{E}\left[\hat{F}_{N,n}^F(x)\right] - F(x)\right| \leq \sum_{k=N+1}^{\infty} |a_k| \int_{-\infty}^{x} |h_k(t)|\, dt \leq cN^{-r/2+1} = O\left(N^{-r/2+1}\right), \tag{7.20}$$

which completes the proof. End Proof

Proof of Lemma 7.6. We follow the proof of [40, Proposition 2]. It holds with the Cauchy-Schwarz inequality that

$$\begin{aligned}\left|\hat{F}_{N,n}^F(x) - \mathbb{E}\left[\hat{F}_{N,n}^F(x)\right]\right|^2 &= \left|\int_{-\infty}^{x} \sum_{k=0}^{N} \hat{a}_k h_k(t)\, dt - \int_{-\infty}^{x} \sum_{k=0}^{N} a_k h_k(t)\, dt\right|^2 \\ &\leq \sum_{k=0}^{N} (\hat{a}_k - a_k)^2 \cdot \sum_{k=0}^{N} \left|\int_{-\infty}^{x} h_k(t)\, dt\right|^2.\end{aligned}$$

For the second term we know with Lemma 7.13 that

$$\begin{aligned}\sum_{k=0}^{N}\left|\int_{-\infty}^{x} h_k(t)\,\mathrm{d}t\right|^2 &\leq \sum_{k=0}^{N}\left(2c_1(k+1)^{-1/4}+12d_1(k+1)^{1/2}\right)^2\\ &= O\left(\sum_{k=0}^{N}\left((k+1)^{-1/4}+(k+1)^{1/2}\right)^2\right)\\ &= O\left(\sum_{k=0}^{N}k+1\right)=O\left(N^2\right).\end{aligned}$$

We get

$$\begin{aligned}\mathbb{E}\left[\left|\hat{F}^F_{N,n}(x)-\mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]\right|^2\right] &\leq O\left(N^2\right)\sum_{k=0}^{N}\mathbb{E}\left[(\hat{a}_k-a_k)^2\right]\\ &= O\left(\frac{N^{5/2}}{n}\right) \qquad (7.21)\end{aligned}$$

with Lemma 7.11. End Proof

Proof of Theorem 7.9. We follow [40, Theorem 3]. Note that

$$\left|\hat{F}^F_{N,n}(x)-F(x)\right| \leq \left|\mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]-F(x)\right| + \left|\hat{F}^F_{N,n}(x)-\mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]\right|.$$

For the first term it holds with Lemma 7.5 that

$$\left|\mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]-F(x)\right| = O\left(N^{-r/2+1}\right)=O\left(n^{-\frac{r-2}{2r+1}}\right)=O\left(n^{-\frac{r-2}{2r+1}}\log n\right).$$

The second term can be written as

$$\left|\hat{F}^F_{N,n}(x)-\mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]\right| \leq O(N)\sqrt{\sum_{k=0}^{N}(\hat{a}_k-a_k)^2} = O(N)\cdot O(n^{-r/(2r+1)}\log n)=O(n^{-\frac{(r-2)}{2r+1}}\log n) \text{ a.s.},$$

where we used a result from the proof of Lemma 7.6 and Lemma 7.12. The theorem follows directly. End Proof

Proof of Theorem 7.10. This proof follows the proof of [40, Theorem 4]. As we will use the lemma of Borel-Cantelli later, we first take a look at

$$\begin{aligned}&\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\hat{F}_{N(n)}-F(x)\right|>\epsilon\right)\\ &=\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\hat{F}_{N(n)}-F(x)\right|>\epsilon : N(n)>cn^{\gamma}\right)\mathbb{P}(N(n)>cn^{\gamma})\\ &\quad+\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\hat{F}_{N(n)}-F(x)\right|>\epsilon : N(n)\leq cn^{\gamma}\right)\mathbb{P}(N(n)\leq cn^{\gamma})\end{aligned}$$

with a constant c. With the assumption that $\sum_{n=1}^{\infty}\mathbb{P}\left(\frac{N(n)}{n^{\gamma}}>\epsilon\right)<\infty$, the first part is clearly finite. For the second part it holds with Markov's inequality for $p\in[1,\infty)$ that

$$\mathbb{P}\left(\left|\hat{F}_{N(n)}-F(x)\right|>\epsilon\right) \leq \epsilon^{-p}\mathbb{E}\left[\left|\int_{-\infty}^{x}\sum_{k=0}^{N}(\hat{a}_k-a_k)h_k(t)\,\mathrm{d}t-\int_{-\infty}^{x}\sum_{k=N+1}^{\infty}a_kh_k(t)\,\mathrm{d}t\right|^p\right].$$

Now, with $|f+g|^p \leq 2^{p-1}(|f|^p+|g|^p)$, the Hölder inequality, Lemma 7.13 and the proof of Lemma 7.5, we get

$$\begin{aligned}
&\mathbb{P}\left(\left|\hat{F}_{N(n)}-F(x)\right|>\epsilon\right)\\
&\leq \epsilon^{-p}2^{p-1}\left(\sum_{k=0}^{N}\mathbb{E}[|\hat{a}_k-a_k|^p]\right)\cdot\left(\sum_{k=0}^{N}|\int_{-\infty}^{x}h_k(t)\,\mathrm{d}t|^{\frac{p}{p-1}}\right)^{p-1}+\epsilon^{-p}2^{p-1}\left|\int_{-\infty}^{x}\sum_{k=N+1}^{\infty}a_kh_k(t)\,\mathrm{d}t\right|^p\\
&\leq \epsilon^{-p}2^{p-1}b_1\left(\sum_{k=0}^{N}\mathbb{E}[|\hat{a}_k-a_k|^p]\right)\cdot\left(\sum_{k=0}^{N}(k+1)^{\frac{p}{2(p-1)}}\right)^{p-1}+\epsilon^{-p}2^{p-1}N^{-\frac{rp}{2}+p}
\end{aligned}$$

with positive constants b_1, b_2. With [55, 1. Summary], it holds that

$$\mathbb{E}\left[|\hat{a}_k-a_k|^p\right]=n^{-p}\mathbb{E}\left[|\sum_{i=1}^{n}(h_k(X_i)-a_k)|^p\right]\leq F_pn^{-p/2-1}\sum_{i=1}^{n}\mathbb{E}\left[|h_k(X_i)-a_k|^p\right],$$

where F_p is only depending on p. Then, with $\max_x|h_k(x)|\leq C(k+1)^{-1/12}$,

$$\begin{aligned}
\mathbb{E}\left[|\hat{a}_k-a_k|^p\right]&\leq F_pn^{-p/2-1}\sum_{i=1}^{n}\mathbb{E}\left[|h_k(X_i)-a_k|^p\right]\\
&\leq F_pn^{-p/2-1}\sum_{i=1}^{n}\mathbb{E}\left[\left(|h_k(X_i)|+\int_{-\infty}^{\infty}|h_k(x)|f(x)\,\mathrm{d}x\right)^p\right]\\
&\leq 2F_pn^{-p/2}(k+1)^{-p/12}.
\end{aligned}$$

Using this result, we get

$$\begin{aligned}
&\mathbb{P}\left(\left|\hat{F}_{N(n)}-F(x)\right|>\epsilon\right)\\
&\leq \epsilon^{-p}2^{p-1}b_3''\left(\sum_{k=0}^{N}n^{-p/2-1}\sum_{i=1}^{n}\mathbb{E}[|h_k(X_i)-a_k|^p]\right)\cdot\left(\sum_{k=0}^{N}(k+1)^{\frac{p}{2(p-1)}}\right)^{p-1}+\epsilon^{-p}2^{p-1}N^{-\frac{rp}{2}+p}\\
&\leq \epsilon^{-p}2^{p-1}b_3'n^{-p/2}\sum_{k=0}^{N}(k+1)^{-p/12}(N+1)^{\frac{3}{2}p-1}+\epsilon^{-p}2^{p-1}N^{-\frac{rp}{2}+p}\\
&\leq \epsilon^{-p}2^{p-1}b_3n^{-p/2}N^{-p/12+1}N^{\frac{3}{2}p-1}+\epsilon^{-p}2^{p-1}N^{-\frac{rp}{2}+p}.
\end{aligned}$$

Then, conditioning on $N(n)\leq cn^\gamma$, we get

$$\begin{aligned}
&\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\hat{F}_{N(n)}-F(x)\right|>\epsilon\middle|N(n)\leq cn^\gamma\right)\\
&\leq \epsilon^{-p}2^{p-1}\sum_{n=1}^{\infty}\left[d_1n^{p\left(-\frac{1}{2}+\frac{17}{12}\gamma\right)}+d_2n^{p\gamma\left(1-\frac{r}{2}\right)}\right]
\end{aligned}$$

and hence, for $0<\gamma<\frac{6}{17}$, p can be chosen so that

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\hat{F}_{N(n)}-F(x)\right|>\epsilon : N(n)\leq cn^\gamma\right)\mathbb{P}(N(n)\leq cn^\gamma)<\infty.$$

With the Borel-Cantelli theorem, we know that $\left|\hat{F}_{N(n)}-F(x)\right|>\epsilon$ only holds for finitely many n so that $\left|\hat{F}_{N(n)}-F(x)\right|\to 0$. End Proof

Proof of Theorem 7.11. This proof is similar to the proof of Theorem 7.6. For fixed N it holds that

$$\hat{F}^F_{N,n}(x) - \mathbb{E}\left[\hat{F}^F_{N,n}(x)\right] = \frac{1}{n}\sum_{i=1}^{n} Y_{i,N},$$

where

$$T_N(x,y) = \sum_{k=0}^{N} h_k(x)h_k(y)$$

and

$$Y_{i,N} = \int_{-\infty}^{x} \left[T_N(X_i,t)\,\mathrm{d}t - \sum_{k=0}^{N} a_k h_k(t)\right] \mathrm{d}t, i \in \{1, ..., n\}.$$

The $Y_{i,N}$ are i.i.d. random variables with mean 0. Define $\gamma_N^2 = \mathbb{E}[Y_{1,N}^2]$. We use the central limit theorem for double arrays (see [33], Section 1.9.3) to show the claim.

Defining

$$A_n = \mathbb{E}\left[\sum_{i=1}^{n} Y_{i,N}\right] = 0 \text{ and } B_n^2 = \mathrm{Var}\left[\sum_{i=1}^{n} Y_{i,N}\right] = n\gamma_N^2,$$

it says that

$$\frac{\sum_{i=1}^{n} Y_{i,N} - A_n}{B_n} \xrightarrow{D} \mathcal{N}(0,1)$$

if and only if the Lindeberg condition

$$\frac{n\mathbb{E}[\mathbb{I}(|Y_{1,N}| > \epsilon B_n)Y_{1,N}^2]}{B_n^2} \to 0 \text{ for } n \to \infty \text{ and all } \epsilon > 0$$

is satisfied.

It holds for $n \to \infty$ that

$$\begin{aligned}
\frac{\sum_{i=1}^{n} Y_{i,N} - A_n}{B_n} &\xrightarrow{D} \mathcal{N}(0,1) \\
\Leftrightarrow \frac{\sum_{i=1}^{n} Y_{i,N}}{\sqrt{n}\cdot\gamma_N} &\xrightarrow{D} \mathcal{N}(0,1) \\
\Leftrightarrow \frac{\sqrt{n}}{\gamma_N}\left(\hat{F}^F_{N,n}(x) - \mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]\right) &\xrightarrow{D} \mathcal{N}(0,1) \\
\Leftrightarrow \sqrt{n}\left(\hat{F}^F_{N,n}(x) - \mathbb{E}\left[\hat{F}^F_{N,n}(x)\right]\right) &\xrightarrow{D} \mathcal{N}\left(0,\sigma^2(x)\right),
\end{aligned}$$

which is the claim of Theorem 7.6. The last equivalence holds because of the following. We have to calculate γ_N^2 which is given by

$$\gamma_N^2 = \mathbb{E}\left[\left(\int_{-\infty}^{x} T_N(X_1,t)\,\mathrm{d}t - \int_{-\infty}^{x}\sum_{k=0}^{N} a_k h_k(t)\,\mathrm{d}t\right)^2\right]$$

$$= \mathbb{E}\left[\left(\int_{-\infty}^{x} T_N(X_1,t)\,\mathrm{d}t\right)^2\right] - 2\int_{-\infty}^{x}\sum_{k=0}^{N} a_k h_k(t)\,\mathrm{d}t \cdot \mathbb{E}\left[\int_{-\infty}^{x} T_N(X_1,t)\,\mathrm{d}t\right] + \left(\int_{-\infty}^{x}\sum_{k=0}^{N} a_k h_k(t)\,\mathrm{d}t\right)^2. \tag{7.22}$$

Again, the first part is the only part where we do not know the asymptotic behavior. Hence, we deal with this now. With [54, Eq. (A8)], which only holds on compact sets, we know that

$$\mathbb{E}\left[\left(\int_{-\infty}^{x} T_N(X_1,t)\,\mathrm{d}t\right)^2\right] = \lim_{P\to\infty}\int_{-P}^{P}\left[\int_{-\infty}^{x} \frac{\sin(M(r-t))}{\pi(r-t)} + O(N^{-1/2})\,\mathrm{d}t\right]^2 f(r)\,\mathrm{d}r.$$

With similar arguments as in the proof of Theorem 7.6 we get for $n \to \infty$ (which implies $M \to \infty$) that

$$\mathbb{E}\left[\left(\int_{-\infty}^{x} T_N(X_1,t)\,\mathrm{d}t\right)^2\right] \to \int_{-\infty}^{x} f(r)\,\mathrm{d}r = F(x).$$

To maintain the reading flow, the proof for this has been moved to Section 12.3.

Hence, plugging this in Eq. (7.22) and using the fact that we know the limits of the other parts from Lemma 7.10, it holds for $n \to \infty$ that

$$\gamma_N^2 \to F(x) - 2F(x)^2 + F(x)^2 = \sigma^2(x). \tag{7.23}$$

Now, we have to show that asymptotic normality really holds. In our case the Lindeberg condition has the form

$$\frac{\mathbb{E}\left[\mathbb{I}(|Y_{1,N}| > \epsilon\sqrt{n}\gamma_N)Y_{1,N}^2\right]}{\gamma_N^2} \to 0 \text{ for } n \to \infty \text{ and all } \epsilon > 0.$$

This is what has to be shown to prove the theorem. Writing the expected value as an integral, we get

$$\int_{-\infty}^{\infty} \mathbb{I}\left(\left|\int_{-\infty}^{x} T_N(r,t)\,\mathrm{d}t - \int_{-\infty}^{x}\sum_{k=0}^{N} a_k h_k(t)\,\mathrm{d}t\right| > \epsilon\sqrt{n}\gamma_N\right)\left(\int_{-\infty}^{x} T_N(r,t)\,\mathrm{d}t - \int_{-\infty}^{x}\sum_{k=0}^{N} a_k h_k(t)\,\mathrm{d}t\right)^2 f(r)\,\mathrm{d}r.$$

The left side of the inequality in the indicator function is bounded and with the same arguments as in the proof of Theorem 7.6, we get

$$\frac{\mathbb{E}\left[\mathbb{I}(|Y_{1,N}| > \epsilon\sqrt{n}\gamma_N)Y_{1,N}^2\right]}{\gamma_N^2} \to 0,$$

which proves the claim. End Proof

Proof of Lemma 7.7. We follow the proof of [40, Proposition 3] again. The Gauss-Hermite estimator with fixed N can be represented as

$$T(x,F_n) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}\sum_{k=0}^{N} h_k(t)h_k(y)\,\mathrm{d}y\right]\mathrm{d}F_n(t),$$

where F_n is the EDF because

$$\begin{aligned}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}\sum_{k=0}^{N} h_k(t)h_k(y)\,\mathrm{d}y\right]\mathrm{d}F_n(t) &= \int_{-\infty}^{x}\sum_{k=0}^{N} h_k(y)\int_{-\infty}^{-\infty} h_k(t)\,\mathrm{d}F_n(t)\,\mathrm{d}y\\ &= \int_{-\infty}^{x}\sum_{k=0}^{N} h_k(y)\hat{a}_k\,\mathrm{d}y = \hat{F}_{N,n}^F(x).\end{aligned}$$

With the same arguments it can be shown that

$$T(x,F) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}\sum_{k=0}^{N} h_k(t)h_k(y)\,\mathrm{d}y\right]\mathrm{d}F(t) = F(x).$$

Define

$$d_N(t,y) = \sum_{k=0}^{N} h_k(t)h_k(y).$$

Now, the empirical influence function is

$$\begin{aligned}
\mathrm{IF}(x,x';T,F_n) &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}\sum_{k=0}^{N}h_k(t)h_k(y)\,\mathrm{d}y\right]\mathrm{d}((1-\epsilon)F_n+\epsilon\delta_{x'})(t)\right.\\
&\quad\left.-\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}\sum_{k=0}^{N}h_k(t)h_k(y)\,\mathrm{d}y\right]\mathrm{d}F_n(t)\right]\\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[(1-\epsilon)\int_{-\infty}^{\infty}\int_{-\infty}^{x}d_N(t,y)\,\mathrm{d}y\,\mathrm{d}F_n(t)+\epsilon\int_{-\infty}^{x}d_N(x',y)\,\mathrm{d}y\right.\\
&\quad\left.-\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}d_N(t,y)\,\mathrm{d}y\right]\mathrm{d}F_n(t)\right]\\
&= \int_{-\infty}^{x}d_N(x',y)\,\mathrm{d}y-\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}d_N(t,y)\,\mathrm{d}y\right]\mathrm{d}F_n(t).
\end{aligned}$$

With the same reasoning, it holds that

$$\mathrm{IF}(x,x';T,F) = \int_{-\infty}^{x}d_N(x',y)\,\mathrm{d}y-\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}d_N(t,y)\,\mathrm{d}y\right]\mathrm{d}F(t).$$

We know with Lemma 7.13 that

$$\begin{aligned}
\left|\int_{-\infty}^{x}d_N(t,y)\,\mathrm{d}y\right| &= \left|\int_{-\infty}^{x}\sum_{k=0}^{N}h_k(t)h_k(y)\,\mathrm{d}y\right|\\
&\le \sum_{k=0}^{N}|h_k(t)|\int_{-\infty}^{x}|h_k(y)|\,\mathrm{d}x\\
&\le \sum_{k=0}^{N}u_1(k+1)^{-1/12}(k+1)^{-1/4}+\sum_{k=0}^{N}v_1(k+1)^{-1/12}(k+1)^{1/2},
\end{aligned}$$

where we used the inequality $\max_x |h_k(x)| \le C(k+1)^{-1/12}$ for $C>0$ that can be found in [43, Theorem 8.91.3]. This shows that the gross-error sensitivities are finite and the lemma follows. End Proof

Proof of Lemma 7.8. This proof follows the proof of [40, Proposition 4]. The kernel estimator can be written as

$$T(x,F_n) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}\frac{1}{h}K\left(\frac{t-y}{h}\right)\mathrm{d}y\right]\mathrm{d}F_n(t)$$

and the distribution function as

$$\begin{aligned}
T(x,F) &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}\frac{1}{h}K\left(\frac{t-y}{h}\right)\mathrm{d}y\right]\mathrm{d}F(t)\\
&= \int_{-\infty}^{x}\left[\int_{-\infty}^{\infty}\frac{1}{h}K\left(\frac{t-y}{h}\right)\mathrm{d}y\right]f(t)\,\mathrm{d}t = F(t),
\end{aligned}$$

because the integral over K is one. With similar arguments as in the proof before we get

$$\mathrm{IF}(x,x';T,F_n) = \int_{-\infty}^{x}\frac{1}{h}K\left(\frac{x'-y}{h}\right)\mathrm{d}y-\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}\frac{1}{h}K\left(\frac{t-y}{h}\right)\mathrm{d}y\right]\mathrm{d}F_n(t),$$

$$\mathrm{IF}(x,x';T,F)=\int\limits_{-\infty}^{x}\frac{1}{h}K\left(\frac{x'-y}{h}\right)\mathrm{d}y-\int\limits_{-\infty}^{\infty}\left[\int\limits_{-\infty}^{x}\frac{1}{h}K\left(\frac{t-y}{h}\right)\mathrm{d}y\right]\mathrm{d}F(t).$$

Now, with $\int\limits_{-\infty}^{\infty} K(x)\,\mathrm{d}x=1$, both influence functions are bounded by two so that the kernel estimator is bias-robust.

End Proof

Proof of Lemma 7.9. This proof follows [40, Proposition 5]. The Gram-Charlier distribution estimator Eq. (7.7) has the representation

$$T(x,F_n)=\int\limits_{-\infty}^{\infty}\left[\int\limits_{-\infty}^{x}\sum_{k=0}^{N}\frac{1}{k!}H_{e_k}(t)H_{e_k}(y)\Phi(y)\,\mathrm{d}y\right]\mathrm{d}F_n(t)$$

as

$$\begin{aligned}\int\limits_{-\infty}^{\infty}\left[\int\limits_{-\infty}^{x}\sum_{k=0}^{N}h_k(t)h_k(y)\,\mathrm{d}y\right]\mathrm{d}F_n(t)&=\int\limits_{-\infty}^{x}\sum_{k=0}^{N}\int\limits_{-\infty}^{-\infty}\frac{1}{k!}H_{e_k}(t)H_{e_k}(y)\Phi(y)\,\mathrm{d}F_n(t)\,\mathrm{d}y\\&=\int\limits_{-\infty}^{x}\sum_{k=0}^{N}\hat{c}_kH_{e_k}(y)\Phi(y)\,\mathrm{d}y=\hat{F}_{N,n}^{GC}(x).\end{aligned}$$

Analogously,

$$T(x,F)=\int\limits_{-\infty}^{\infty}\left[\int\limits_{-\infty}^{x}\sum_{k=0}^{N}\frac{1}{k!}H_{e_k}(t)H_{e_k}(y)\Phi(y)\,\mathrm{d}y\right]\mathrm{d}F(t)=F(x).$$

Then it holds with the same arguments as above that

$$\mathrm{IF}(x,x';T,F_n)=\sum_{k=0}^{N}\left[\frac{1}{k!}H_{e_k}(x')\int\limits_{-\infty}^{x}H_{e_k}(y)\Phi(y)\,\mathrm{d}y-\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{x}\frac{1}{k!}H_{e_k}(t)H_{e_k}(y)\Phi(y)\,\mathrm{d}y\,\mathrm{d}F_n(t)\right],$$

and

$$\mathrm{IF}(x,x';T,F)=\sum_{k=0}^{N}\left[\frac{1}{k!}H_{e_k}(x')\int\limits_{-\infty}^{x}H_{e_k}(y)\Phi(y)\,\mathrm{d}y-\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{x}\frac{1}{k!}H_{e_k}(t)H_{e_k}(y)\Phi(y)\,\mathrm{d}y\,\mathrm{d}F(t)\right].$$

The second term is bounded because of [43, Theorem 8.91.3] and

$$\begin{aligned}\int\limits_{-\infty}^{x}|H_{e_k}(y)\Phi(y)|\,\mathrm{d}y&=\int\limits_{-\infty}^{x}2^{-\frac{k}{2}}\left|H_k\left(\frac{y}{\sqrt{2}}\right)\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\right|\mathrm{d}y\\&=\int\limits_{-\infty}^{\frac{x}{\sqrt{2}}}2^{-\frac{k}{2}}\left|H_k(z)\frac{e^{-z^2}}{\sqrt{\pi}}\right|dz\\&\le d_1(k+1)^{-1/12}\sqrt{k!}\pi^{-1/4}\int\limits_{-\infty}^{\infty}e^{-\frac{z^2}{2}}dz=d_2<\infty.\end{aligned}$$

Then,

$$\left|\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{x}\frac{1}{k!}H_{e_k}(t)H_{e_k}(y)\Phi(y)\,\mathrm{d}y\,\mathrm{d}F(t)\right|\le d_2\frac{1}{k!}\int\limits_{-\infty}^{\infty}|H_{e_k}(t)|f(t)\,\mathrm{d}t<\infty$$

and

$$\left| \int_{-\infty}^{\infty} \int_{-\infty}^{x} \frac{1}{k!} H_{e_k}(t) H_{e_k}(y) \Phi(y) \, \mathrm{d}y \, \mathrm{d}F_n(t) \right| \leq d_2 \frac{1}{k!} \int_{-\infty}^{\infty} |H_{e_k}(t)| \, \mathrm{d}F_n(t) = d_2 \frac{1}{k!} \sum_{i=1}^{n} |H_{e_k}(X_i)| < \infty.$$

Now, since $H_{e_k}(x')$ is unbounded, the gross-error sensitivities are not bounded and the proof is completed. End Proof

8 Comparison

In the following, the main results of this thesis are summarized and used to compare the different estimators.

The properties of the estimators are summarized in two tables on pages 98 and 99. The assumptions in the third column of the first table have to be fulfilled for the theoretical results to hold. If there are extra assumptions for one specific result, they are written as a footnote. For more details, please take a look at the respective part in the thesis.

It is important to always make sure that the situation fits to compare different estimators. A comparison between the Bernstein estimator and the Szasz estimator for example only makes sense when the density function on $[0,1]$ can be continued to $[0,\infty)$ so that Assumption 6.1 holds. Of course it is also possible to use the Szasz estimator for distributions where F is continuous on $[0,\infty)$ and f is not. Then, the theoretical results do not hold anymore but convergence is still given. But we know that the Bernstein estimator is always better as it has zero bias and variance for $x=1$, while the Szasz estimator has the continuous derivative

$$\frac{d}{\mathrm{d}x}\hat{F}^S_{m,n}(x) = m\sum_{k=0}^{\infty}\left[F_n\left(\frac{k+1}{m}\right) - F_n\left(\frac{k}{m}\right)\right]e^{-mx}\frac{(mx)^k}{k!}$$

and cannot approximate a non-continuous function that well. This can be seen in Figure 11. It is obvious that the behavior of the Szasz estimator in the point $x=1$ of the Beta$(2,1)$-distribution is worse. This can also be seen later in the simulation in Section 9.

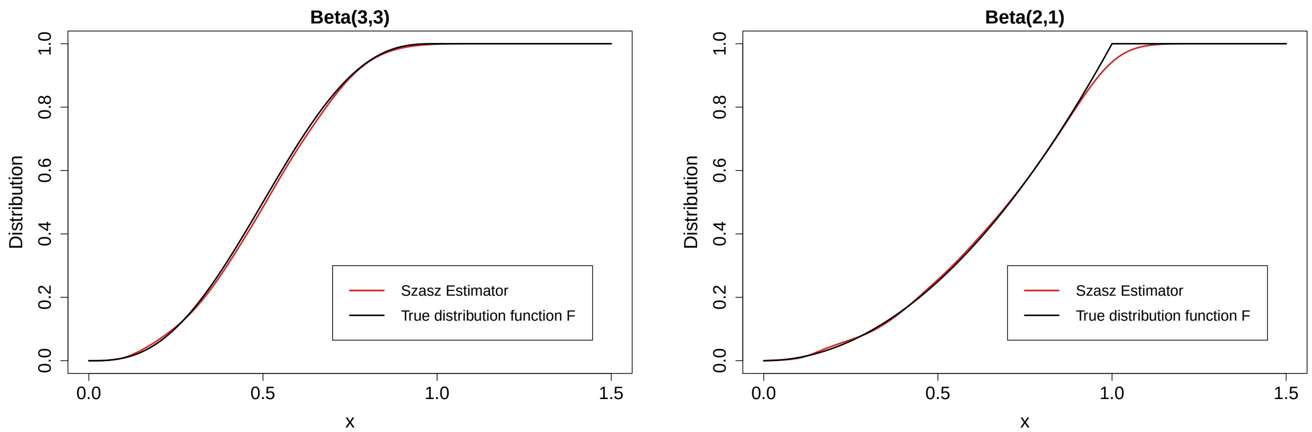

Figure 11: The behavior of the Szasz estimator in the point $x=1$ for $n=500$.

For the Hermite estimators, the properties $f \in L_2$ and $\left(x - \frac{d}{\mathrm{d}x}f\right)^r f \in L_2$ only have to hold on the considered interval (see for example Section 7.3.3). Hence, they can be used for smaller intervals than what they were designed for.

The EDF and the kernel distribution estimator can be used on arbitrary intervals. However, note that the asymptotic results for the kernel estimator hold under the assumption that the density occupies $(-\infty,\infty)$. Hence, if the support is bounded, the results do not hold for the points close to the boundary. For an approach to improve this boundary behavior, see [8] for example.

Some Observations In the following, some important observations regarding the theoretical comparison are listed. It is notable that for the asymptotic order, $h = 1/m$ for the Bernstein estimator is always replaced by h^2 for the Kernel estimator. Also, the results for the Szasz estimator are the same as for the Bernstein estimator with the exception that the orders are often not uniform.

There are some properties that some or all of the estimators have in common. Regarding the deficiency, we found out that the Bernstein estimator, the kernel estimator, and the Szasz estimator all outperform the EDF with respect to MSE and MISE. All of the estimators convergence a.s. uniformly to the true distribution function, and the asymptotic distribution of the scaled difference between estimator and the true value always coincide under different assumptions.

However, there are of course also many differences between the estimators that are addressed now. For the Bernstein estimator and the Szasz estimator, the order of the bias is worse than that of the kernel estimator. For the Szasz estimator, the order is not uniform. The order of the Hermite estimator on the real half line depends on x. This is not the case for the estimator on the real line. On the other hand, the order for the estimator on the real line is worse.

For the variance, the orders of the Bernstein estimator and the Szasz estimator are the same as for the EDF and the kernel estimator but are not uniform. The Hermite estimator on the real line is worse than the estimator for the real half line but uniform. Their orders are both worse than that of the other estimators.

The optimal rate of the MSE is n^{-1} for the first four estimators in the table, two of them uniform and the others not. The rates of the Hermite estimators are worse but for $r \to \infty$, the rates also approach n^{-1}. This is very similar for the optimal rates of the MISE.

	Support	Assumptions	Convergence	Asymptotic distribution: $n^{1/2}(\hat{F}_n(x) - F(x)) \xrightarrow{D} \mathcal{N}(0, \sigma^2(x))$ [a]
EDF (see 2)	Chosen Freely		a.s. uniform	
Kernel (see 3)	Chosen Freely	Density f exists, f' exists and is continuous	a.s. uniform	For $h^{-2}n^{-1/2} \to \infty$
Bernstein (see 5)	$[0, 1]$	F continuous, two continuous and bounded derivatives on $[0, 1]$	a.s. uniform	For $mn^{-1/2} \to \infty$
Szasz (see 6)	$[0, \infty)$	F continuous, two continuous and bounded derivatives on $[0, \infty)$	a.s. uniform	For $mn^{-1/2} \to \infty$
Hermite Half (see 7.3)	$[0, \infty)$	$f \in L_2$	a.s. uniform [b]	For $N^{r/2-1/4}n^{-1/2} \to \infty$ [c]
Hermite Full (see 7.4)	$(-\infty, \infty)$	$f \in L_2$	a.s. uniform [d]	For $N^{r/2-1}n^{-1/2} \to \infty$ [e]

[a] $\hat{F}_n$ stands for all of the estimators, for $x : 0 < F(x) < 1$
[b] For $\left(x - \frac{d}{dx}\right)^r f \in L_2, r \geq 1, \mathbb{E}\left[|X|^s\right] < \infty, s > \frac{8(r+1)}{3(2r+1)}, N \sim n^{\frac{2}{2r+1}}$
[c] For $\left(x - \frac{d}{dx}\right)^r f \in L_2, r \geq 1, \mathbb{E}\left[|X|^{2/3}\right] < \infty$
[d] For $\left(x - \frac{d}{dx}\right)^r f \in L_2, r > 2, \mathbb{E}\left[|X|^s\right] < \infty, s > \frac{8(r+1)}{3(2r+1)}, N \sim n^{\frac{2}{2r+1}}$
[e] For $\left(x - \frac{d}{dx}\right)^r f \in L_2, r > 2$

	Bias	Variance	MSE (all consistent)	MISE (all consistent)
EDF (see 2)	Unbiased	$O(n^{-1})$	$O(n^{-1})$	$O(n^{-1})$
Kernel (see 3)	$o(h^2)$	$O(n^{-1}) + O(h/n)$	$O(n^{-1}) + O(h^4) + O(h/n)$ Optimal: $O(n^{-1})$	$O(n^{-1}) + O(h^4) + O(h/n)$ Optimal: $O(n^{-1})$
Bernstein (see 5)	Zero in $\{0,1\}$ $O(m^{-1}) = O(h)$	Zero in $\{0,1\}$ $O(n^{-1}) + O_x(m^{-1/2}n^{-1})$	Zero in $\{0,1\}$ $O(n^{-1}) + O(m^{-2}) + O_x(m^{-1/2}n^{-1})$ Optimal: $O_x(n^{-1})$	$O(n^{-1}) + O(m^{-2}) + O(m^{-1/2}n^{-1})$ Optimal: $O(n^{-1})$
Szasz (see 6)	Zero in 0 $O_x(m^{-1}) = O_x(h)$	Zero in 0 $O(n^{-1}) + O_x(m^{-1/2}n^{-1})$	Zero in 0 $O(n^{-1}) + O_x(m^{-2}) + O_x(m^{-1/2}n^{-1})$ Optimal: $O_x(n^{-1})$	$O(n^{-1}) + O(m^{-2}) + O(m^{-1/2}n^{-1})$ Optimal: $O(n^{-1})^a$
Hermite Half (see 7.3)	Zero in 0 $O_x\left(N^{-r/2+1/4}\right)^b$	Zero in 0 $O_x(N^{3/2}/n)^c$	Zero in 0 $x\left[O\left(\frac{N^{1/2}}{n}\right) + O\left(N^{-r}\right)\right]$ Optimal: $xO(n^{\frac{-2r}{2r+1}})\ ^b$	$\mu\left[O\left(\frac{N^{1/2}}{n}\right) + O\left(N^{-r}\right)\right]$ Optimal: $\mu O(n^{-\frac{2r}{2r+1}})^d$
Hermite Full (see 7.4)	$O(N^{1-r/2})^e$	$O(N^{5/2}/n)^c$	$O\left(\frac{N^{5/2}}{n}\right) + O\left(N^{-r+2}\right)$ Optimal: $O(n^{-\frac{2(r-2)}{2r+1}})^f$	$O\left(\frac{N^{5/2}}{n}\right) + O\left(N^{-r+2}\right)$ Optimal: $O\left(n^{-\frac{2(r-2)}{2r+1}}\right)^f$

[a] Note that the MISE here is defined differently with weight function e^{-ax}
[b] For $\left(x - \frac{d}{dx}\right)^r f \in L_2, r \geq 1, \mathbb{E}\left[|X|^{2/3}\right] < \infty$
[c] For $\mathbb{E}\left[|X|^{2/3}\right] < \infty$
[d] For $\left(x - \frac{d}{dx}\right)^r f \in L_2, r \geq 1, \mu < \infty$
[e] For $\left(x - \frac{d}{dx}\right)^r f \in L_2, r > 2$
[f] For $\left(x - \frac{d}{dx}\right)^r f \in L_2, r > 2, \mathbb{E}\left[|X|^{2/3}\right] < \infty$

9 Simulation

In this section, the estimators defined in this thesis are compared in a simulation study.

For the kernel distribution estimator, the Gaussian kernel is chosen. Let Φ be the standard normal distribution function. Then, the estimator is of the form

$$F_{h,n}(x) = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{x - X_i}{h}\right). \tag{9.1}$$

The simulation consists of four parts. In the first three parts, the estimators are compared by their MISE on different intervals. To be specific, these three parts are:

1. Comparison on $[0,1]$, $\text{MISE}\left[\hat{F}_n\right] = \mathbb{E}\left[\int_0^1 \left(\hat{F}_n(x) - F(x)\right)^2 \mathrm{d}x\right]$,
2. comparison on $[0,\infty)$, $\text{MISE}\left[\hat{F}_n\right] = \mathbb{E}\left[\int_0^\infty \left(\hat{F}_n(x) - F(x)\right)^2 \cdot f(x)\,\mathrm{d}x\right]$,
3. comparison on $(-\infty,\infty)$, $\text{MISE}\left[\hat{F}_n\right] = \mathbb{E}\left[\int_{-\infty}^\infty \left(\hat{F}_n(x) - F(x)\right)^2 \cdot f(x)\,\mathrm{d}x\right]$,

where $\hat{F}_n$ can be any of the considered estimators. In the fourth part, the asymptotic normality of the estimators is illustrated for one distribution. The details for each part as well as the most important results are explained later.

All of the estimators except for the empirical distribution function (EDF) have a parameter in addition to n. For these estimators, the MISE is calculated for a range of the parameters, which are given in Table 1.

Estimator	Abbr.	Parameters
EDF	F_n	-
Kernel	$F_{h,n}$	$h = i/1000$, $i \in [2, 200]$
Bernstein	$\hat{F}_{m,n}$	$m \in [2, 200]$
Szasz	$\hat{F}^S_{m,n}$	$m \in [2, 200]$
Hermite Half	$\hat{F}^H_{N,n}$	$N \in [2, 60]$
Hermite Full	$\hat{F}^F_{N,n}$	$N \in [2, 60]$

Table 1: The range of the respective parameters.

We obtain a vector of MISE-values for each estimator. Searching for the minimum value in this vector provides the minimal MISE and the respective optimal parameter.

The different sample sizes that are used are $n = 20, 50, 100$, and 500.

Every MISE is calculated by a Monte-Carlo simulation with $M = 10,000$ repetitions. To be specific, let

$$\text{ISE}\left[\hat{F}\right] = \int [\hat{F}(x) - F(x)]^2\,(\cdot f(x))\,\mathrm{d}x \tag{9.2}$$

and with M pseudo-random samples, the estimate of the MISE is calculated by

$$\text{MISE}\left[\hat{F}\right] \simeq \frac{1}{M}\sum_{i=1}^{M} \text{ISE}_i[\hat{F}], \tag{9.3}$$

where ISE_i is the integrated squared error calculated from the ith randomly generated sample.

For the Hermite estimators, the standardization explained in Section 7.6 is used. In this simulation, we do not estimate the parameters μ and σ as we already know the true parameters.

We now explain the different parts and the results.

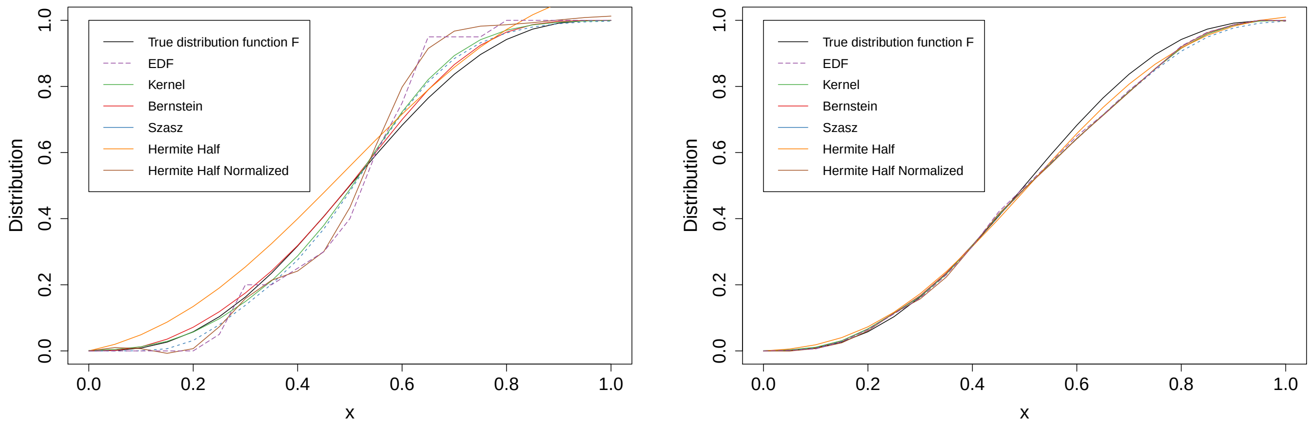

Figure 12: Plot of the considered estimators for $n = 20$ and $n = 500$.

9.1 Comparison on $[0, 1]$

For the comparison on the unit interval, all of the estimators except the Hermite estimator on the real line are compared while estimating a distribution function on $[0, 1]$. This is the largest interval where all of the estimators are defined. For the Hermite estimator on the real half line, both the non-standardized and the standardized estimators are compared.

The distribution used here is the Beta$(3, 3)$-distribution. It fulfills all the assumptions for the different estimators: F, f, f' are continuous and bounded on $[0, 1]$, it can be expanded so that the assumption for the Szasz estimator holds, and $f \in L_2, \left(x - \frac{d}{dx}\right)^r f \in L_2$ holds for the Hermite estimator.

In Figure 12, it is illustrated how the distribution function for $n = 20$ and $n = 500$ is estimated for all of the estimators. It is obvious that with more samples, the estimation quality increases. Note that the illustration only captures the result for one sample and cannot be seen as a general behavior of the estimators.

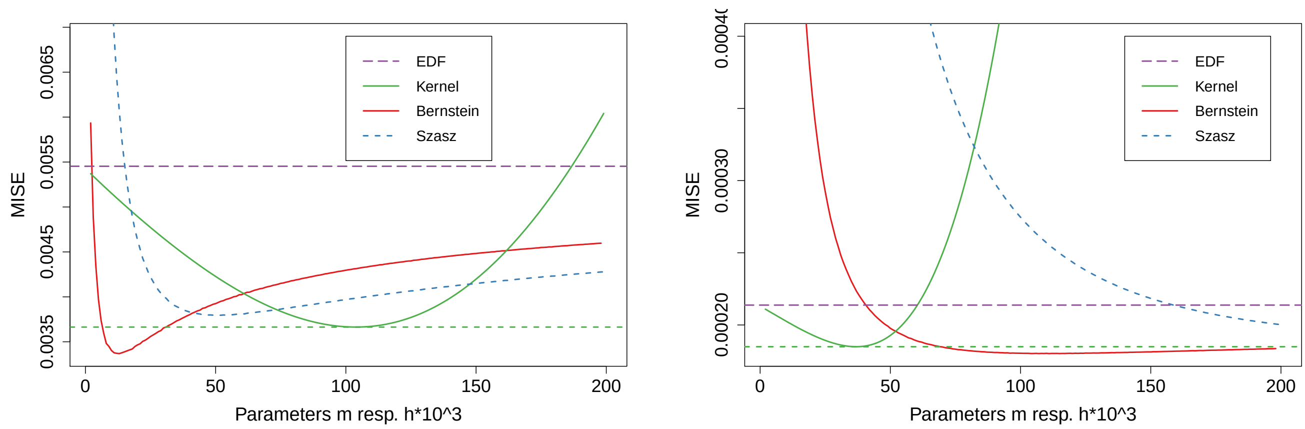

Figure 13: MISE over the respective parameters in $[2, 200]$ for $n = 20$ and $n = 500$.

In Figure 13, the MISE of four of the estimators is plotted over the respective parameters in the interval $[2, 200]$ for $n = 20$ and $n = 500$. The results are very similar for all of the four sample sizes: the Bernstein estimator is always better than the others. This makes sense as it is designed for the unit interval. The EDF has of course the worst MISE and the optimal MISE-values for the Szasz and Kernel estimator are very close.

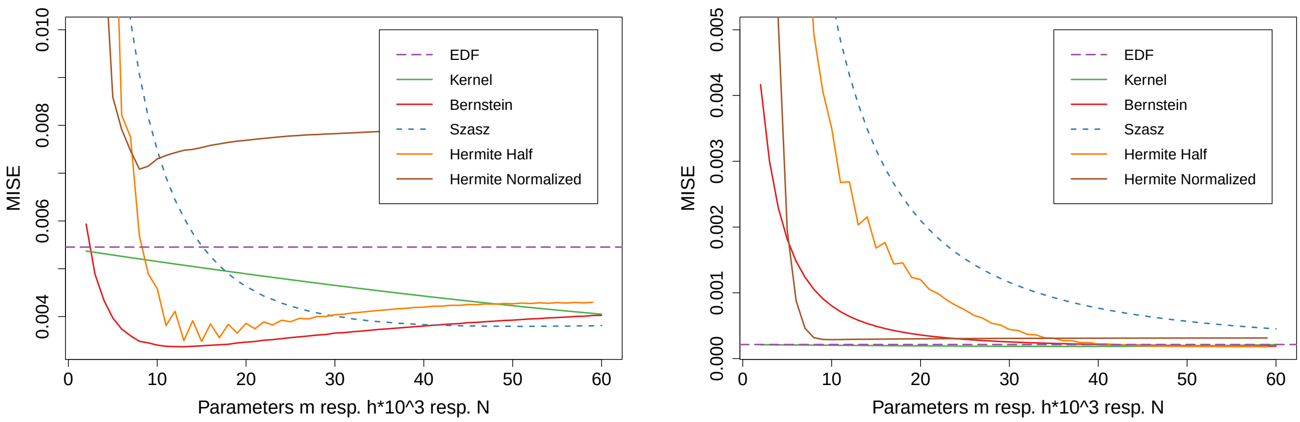

Figure 14: MISE over the respective parameters in $[2, 60]$ for $n = 20$ and $n = 500$.

Figure 14 shows the comparison of the other estimators to the Hermite estimators for $n = 20$ and $n = 500$. It is obvious that in this case, the standardization does not help for $n = 20$. For more samples, the MISE of the normalized estimator decreases a lot quicker, which means that the sum can be truncated earlier.

	n	EDF	Kernel	Bernstein	Szasz	Hermite Half	Hermite Normalized
Beta(3,3)	20	5.45	3.66	3.37	3.79	3.48	7.08
	50	2.15	1.59	1.49	1.64	1.68	2.82
	100	1.07	0.83	0.79	0.85	0.83	1.4
	500	0.21	0.18	0.18	0.2	0.18	0.29

Table 2: The MISE $\cdot 10^{-3}$-values for the interval $[0, 1]$.

Table 2 shows the optimal MISE $\cdot 10^{-3}$-values for all of the estimators, depending on the number of samples. All of the properties that were explained above can also be seen here.

9.2 Comparison on $[0, \infty)$

Of course, the comparison of the estimators on $[0, \infty)$ means that the Bernstein estimator cannot be used anymore. As for the comparison on the unit interval, we omit the Hermite estimator on the real line.

For comparison, the exponential distribution with parameter $\lambda = 2$ is chosen. This distribution fulfills the assumption for the Szasz estimator and the Hermite estimator, see Section 7.3.3.

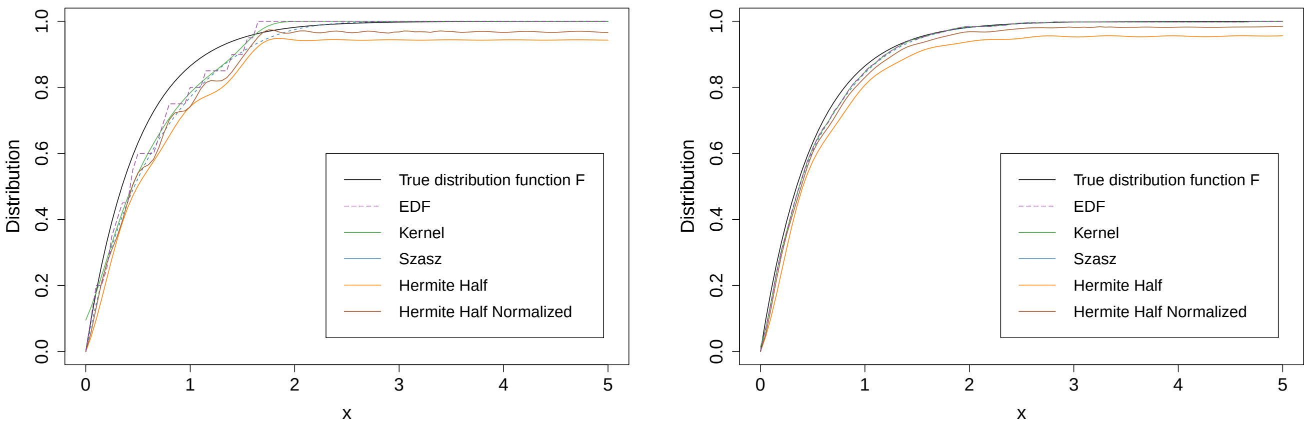

Figure 15: Plot of the considered estimators for $n = 20$ and $n = 500$.

An example of the different estimators can be seen in Figure 15 for $n = 20$ and $n = 500$. It is obvious that the Hermite estimators do not approach one, which is due to the truncation.

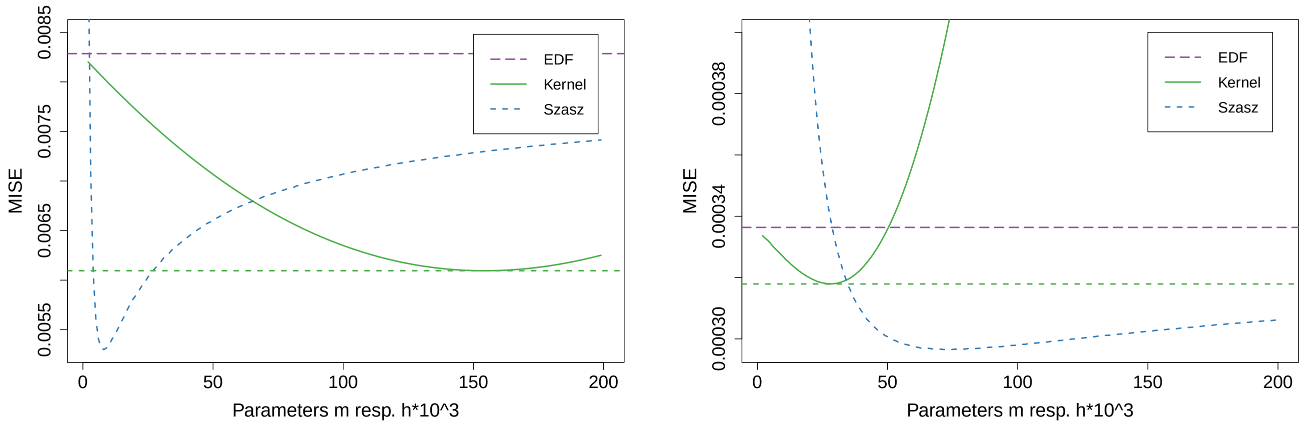

Figure 16: MISE over the respective parameters in $[2, 200]$ for $n = 20$ and $n = 500$.

As in the case of the unit interval, the Szasz estimator designed for the $[0, \infty)$-interval behaves best with respect to MISE. This can be seen in Figure 16. The minimal MISE-value of the Szasz estimator is always lower than that of the other estimators, also for the cases $n = 50$ and $n = 100$ that are not shown here.

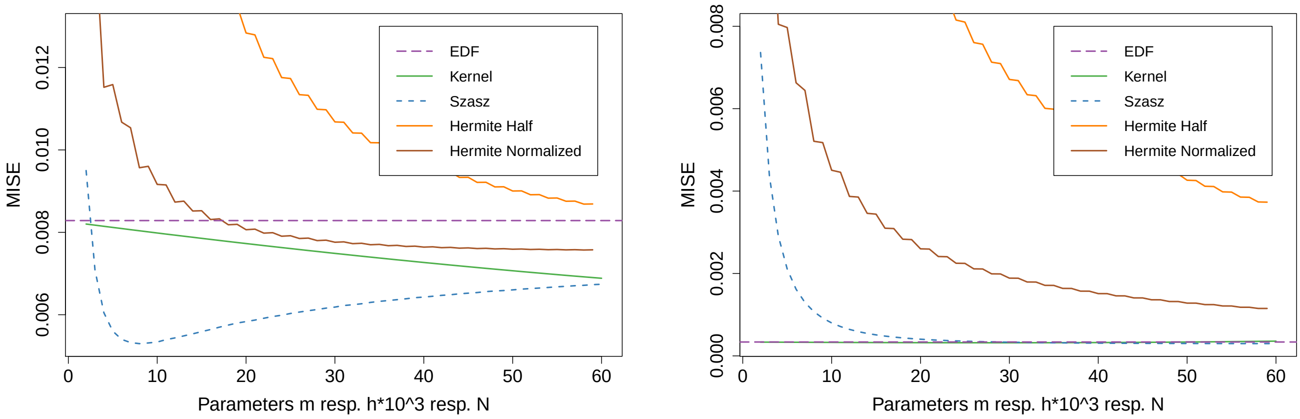

Figure 17: MISE over the respective parameters in $[2, 60]$ for $n = 20$ and $n = 500$.

Figure 17 makes clear that the standardization of the Hermite estimator works better here than in the part for the unit interval, even for small sample sizes. This could be due to the fact that scaling makes a bigger difference in this case than in the unit interval. Note that even though the estimator gets better through standardization, the Hermite estimator is still worse than all of the other estimators (except the EDF) with respect to MISE.

	n	EDF	Kernel	Szasz	Hermite Half	Hermite Normalized
Exponential(2)	20	8.29	6.09	5.3	8.68	7.57
	50	3.3	2.71	2.41	5.61	3.58
	100	1.68	1.47	1.32	4.6	2.26
	500	0.34	0.32	0.3	3.73	1.15

Table 3: The MISE $\cdot 10^{-3}$-values for the interval $[0, \infty)$.

Table 3 shows all the MISE $\cdot 10^{-3}$-numbers of the optimal MISE for the considered estimators. The properties explained above can be found here as well.

9.3 Comparison on $(-\infty, \infty)$

Here, the most important task is to compare the two Hermite estimators Eq. (7.10) and Eq. (7.11) on the real line. They are called Hermite Full estimator and Hermite Full second estimator respectively in the sequel. Both of them are standardized as described in 7.6.

The distribution used here is the Laplace distribution with parameters $\mu = 3, b = 2$.

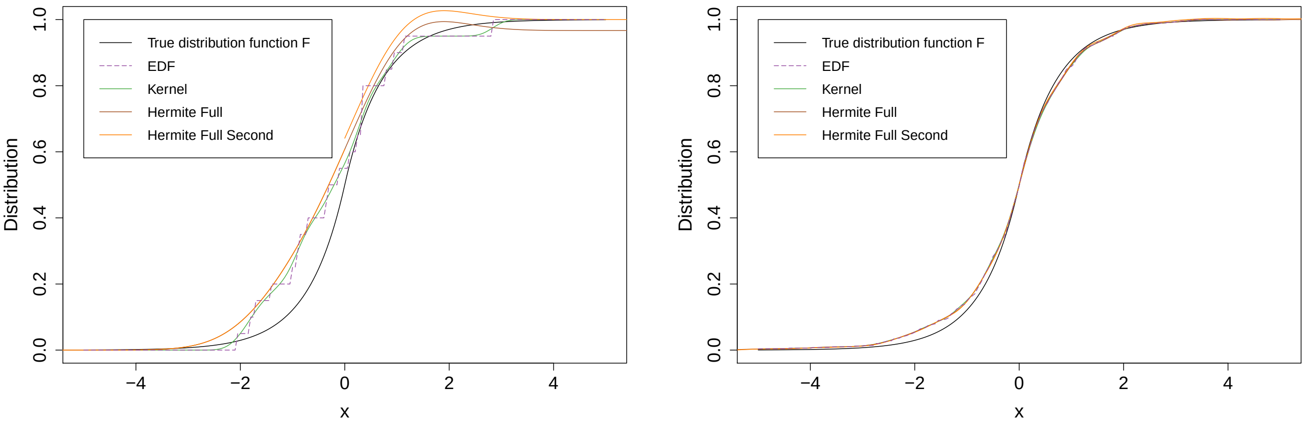

Figure 18: Plot of the considered estimators for $n = 20$ and $n = 500$.

In Figure 18, the considered estimators are plotted again for $n = 20$ and $n = 500$ for one example.

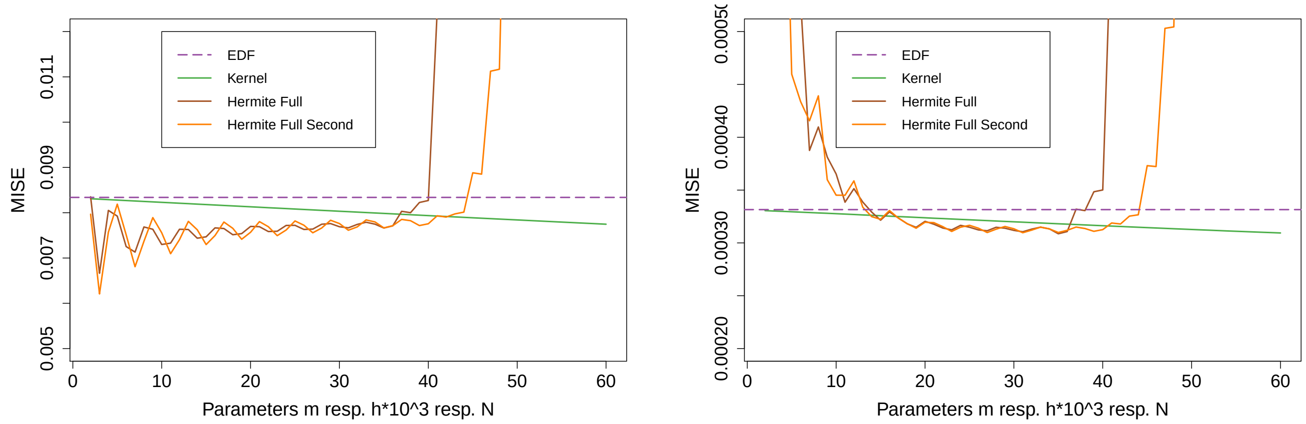

Figure 19: MISE over the respective parameters in $[2, 60]$ for $n = 20$ and $n = 500$.

In Figure 19, it can be seen that the second estimator works better, especially for large parameters, where the MISE curve increases less quickly. Both of them oscillate, particularly for smaller sample sizes, which is not a desirable property. Note, however, that for small parameter values, the Hermite estimator seems to achieve better MISE-values than the kernel estimator.

	n	EDF	Kernel	Hermite Second	Hermite Second
Laplace(3,2)	20	8.34	6.61	6.66	6.21
	50	3.36	2.72	2.95	2.82
	100	1.64	1.35	1.48	1.45
	500	0.33	0.3	0.31	0.31

Table 4: The MISE $\cdot 10^{-3}$-values for the interval $(-\infty, \infty)$.

Table 4 shows the MISE $\cdot 10^{-3}$-values again. As before, the properties explained above can also be found in the table.

9.4 Illustration of the Asymptotic Normality

The goal here is to illustrate the asymptotic normality

$$\sqrt{n}\left(\hat{F}_n(x) - \mathbb{E}\left[\hat{F}_n(x)\right]\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right)$$

of the different estimators, where $\hat{F}_n$ can be any of the estimators. The expression can be rewritten as

$$\hat{F}_n(x) \xrightarrow{D} \mathcal{N}\left(\mathbb{E}\left[\hat{F}_n(x)\right], \frac{\sigma^2(x)}{n}\right).$$

This representation is used in the plots below for a Beta(3, 3)-distribution in the point $x = 0.4$ for $n = 500$. The value is $F(0.4) = 0.32$. In Figure 20, the result can be seen. The red line in the plot shows the distribution function of the normal distribution. Furthermore, the histogram of the value $p = \hat{F}_n(0.4)$ is illustrated. The parameters used for the estimators are derived from the optimal parameters calculated in the simulation.

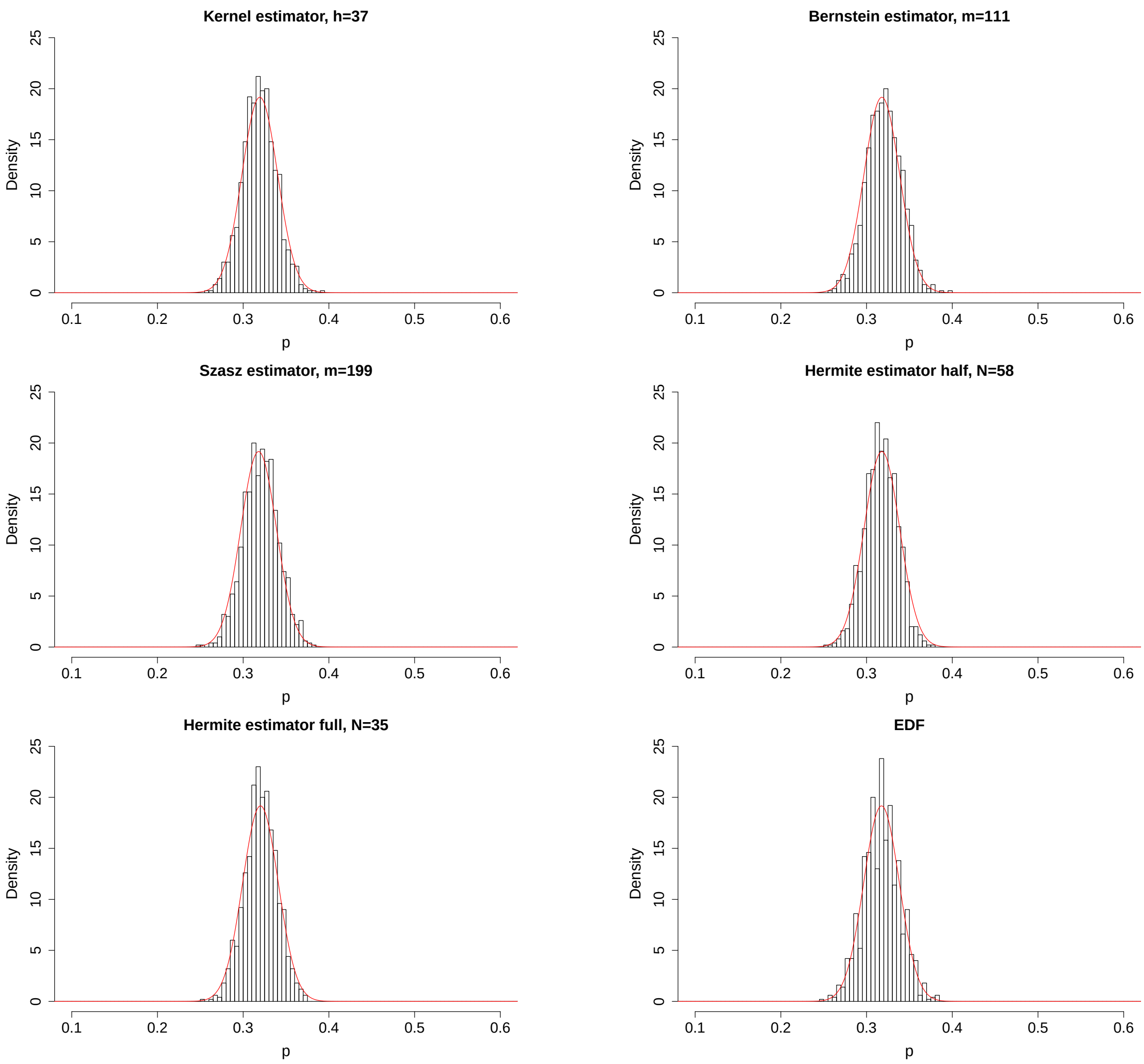

Figure 20: Illustration of the asymptotic normal distribution.

10 Conclusions

In the introduction, it was explained why distribution function estimation is an important topic in statistics. Here, we give a short summary of the findings of this thesis and talk about practical applications as well as further research recommendations.

Findings and contributions In this thesis, many properties of the considered estimators such as the asymptotic behavior, MSE, MISE, and deficieny were explained and derived. It turned out that the estimators often have properties in common. But there are of course many differences as well. This was discussed in Section 8, where the estimators were compared theoretically and it was explained in which cases it makes sense to compare them. In Section 9, the results of a simulation study were presented. In the simulation, all of the estimators were compared on different intervals with respect to MISE. As expected, the estimator designed for a specific interval also had the best behavior there. Furthermore, the asymptotic normality that was proven before for all of the estimators was illustrated.

The most important contribution of this thesis is the Szasz estimator. While the idea is similar to the Bernstein estimator, it allows estimating distribution functions on the real half line. In the theoretical comparison as well as in the simulation study, the Szasz estimator compared very well to the other estimators. Especially on the matching interval $[0, \infty)$, the simulation study showed a clear advantage of the Szasz estimator with respect to the MISE-quality.

Another contribution of this thesis is that the asymptotic normality of the two Hermite distribution estimators was proven. To the best of my knowledge, this is the first proof of this property for the Hermite distribution estimators. Asymptotic normality is an important property that can make the handling of an estimator easier. This is due to the fact that the properties and the behavior of a gaussian variable are well understood.

Practical applications Distribution function estimation can be used everywhere, where data is collected and properties about the data are of interest. This is for example the case in the finance sector. Another example where data is collected is in the medical sector. A data set could give the number of days a treatment takes to heal patients. With distribution function estimation it would be possible to estimate the probability that the treatment takes less than a week.

With the inverse transform sampling that was explained in the introduction, it is even possible to produce more samples than the given ones. Hence, the sample set of the treatment days could be expanded without testing further patients.

The estimators are restricted to different intervals. The Bernstein estimator for example is restricted to $[0, 1]$. It can be applied to all data sets that have an upper and a lower bound. An example is the temperature of water, which always lies between zero and 100 degrees under normal circumstances. The Szasz estimator is restricted $[0, \infty)$. Here, data that can be used has to have a lower bound. This could be the price of an expensive product. It can never be negative but it is not possible to set an upper bound. For estimators on $(-\infty, \infty)$, the data that can be used need not have any bounds at all. This holds for the rate at which the stock market changes. It can rise and fall arbitrarily.

Recommendations for further research As mentioned in Section 3, the kernel estimator shows undesired behavior at the boundary. For the Bernstein estimator, it was established on page 18 that the asymptotic behavior gets better the closer we get to the boundary. Hence, it would be interesting to study the behavior of the Szasz estimator and the Hermite estimator on the half line close to the boundary $x = 0$. It would also be interesting to apply boundary correction techniques in the simulation and compare the MISE results with the non-corrected estimators.

11 Acknowledgements

In the sequel, I want to thank all the people who contributed professionally and personally to the success of this thesis. I really enjoyed the writing process, which was only possible because of their help.

First of all, my special thanks goes to my supervisor PD Dr. Bernhard Klar for his support and helpful feedback on this thesis. While giving me a lot of freedom in the shaping of my work, he provided input in the right places. I always knew how to proceed with the thesis and felt well-supported the whole time.

Furthermore, I want to thank Sina Burkhardt and Tamara Göll for the interesting discussions and the really helpful feedback in the end as well as for making the last semesters so much more fun than they would have been without them.

Lastly, I especially want to thank my family, most of all my dad, and my boyfriend Steven. They always believed in me and supported me the whole time. I am grateful that they were always there for me. Without them I would not be where I am right now.

Ariane Hanebeck

12 Appendix

12.1 Bernstein Basis Polynomials

The Bernstein basis polynomials of degree m are defined by

$$P_{k,m}(x) = \binom{m}{k} x^k (1-x)^{m-k}$$

for $k = 0, ..., m$. We set $P_{k,m} = 0$ for $k < 0$ or $k > m$.

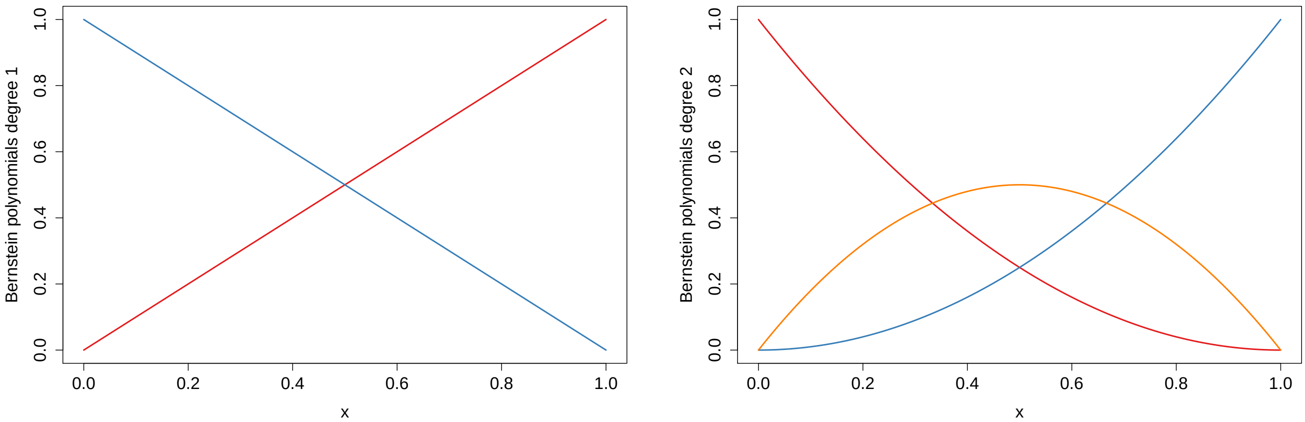

Figure 21: Bernstein basis polynomials of degree one and two.

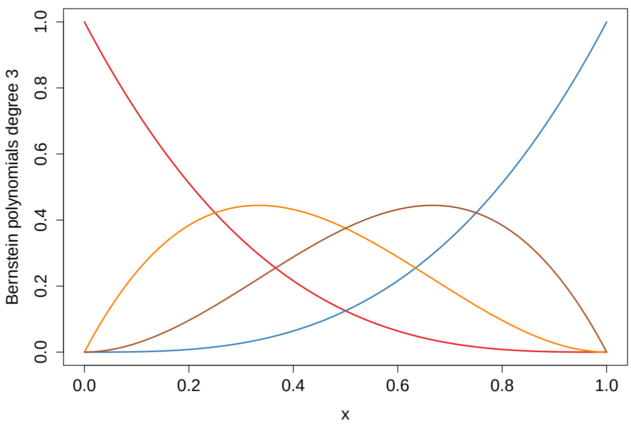

Figure 22: Bernstein basis polynomials of degree three.

The simple cases are the Bernstein basis polynomials of degree 1

$$P_{0,1}(x) = 1 - x$$
$$P_{1,1}(x) = x,$$

degree 2

$$P_{0,2}(x) = (1-x)^2$$
$$P_{1,2}(x) = 2x(1-x)$$
$$P_{2,2}(x) = x^2,$$

and degree 3

$$
\begin{aligned}
P_{0,3}(x) &= (1-x)^3 \\
P_{1,3}(x) &= 3x(1-x)^2 \\
P_{2,3}(x) &= 3x^2(1-x) \\
P_{3,3}(x) &= x^3.
\end{aligned}
$$

The following properties and their proofs can be found in [56].

Recursive Definition

The polynomial $P_{k,m}$ can be written as a convex combination of $P_{k,m-1}$ and $P_{k-1,m-1}$

$$P_{k,m}(x) = (1-x)P_{k,m-1}(x) - xP_{k-1,m-1}(x).$$

With this it can be shown that Bernstein Polynomials are non-negative.

Partition of Unity

The $k+1$ Bernstein polynomials of degree m form a partition of unity, i.e.,

$$\sum_{k=0}^{m} P_{k,m}(x) = 1.$$

Power Basis

A Bernstein polynomial can be written as

$$P_{k,m}(x) = \sum_{i=k}^{m} (-1)^{i-k} \binom{m}{i} \binom{i}{k} x^i$$

from which it follows that

$$x^k = \sum_{i=k-1}^{n-1} \frac{\binom{i}{k}}{\binom{m}{k}} P_{i,m}(x).$$

Derivatives

The derivative of a Bernstein polynomial of degree n can be written as

$$\frac{\partial}{\partial x} P_{k,m}(x) = m(P_{k-1,m-1}(x) - P_{k,m-1}(x))$$

for $k = 0, ..., m$.

12.2 Details for the Proof of Theorem 7.6

We explain here, why it is possible to exchange limit and integral in the proof of Theorem 7.6. We first observe that for $x \neq 0$,

$$-\frac{1}{\pi|x|} - \frac{1}{2} \leq \int_0^x \frac{\sin(l)}{\pi l}\, \mathrm{d}l \leq \frac{1}{\pi|x|} + \frac{1}{2}.$$

It follows that

$$\left(\int_0^x \frac{\sin(l)}{\pi l}\, \mathrm{d}l\right)^2 \leq \left(\frac{1}{\pi|x|} + \frac{1}{2}\right)^2.$$

Hence, for $r \in \{0, x\}$,

$$\left(\int_{M(r-x)}^{Mr} \frac{\sin(l)}{\pi l}\, \mathrm{d}l\right)^2 \leq \left(\frac{1}{\pi|Mx|} + \frac{1}{2}\right)^2$$

and for the rest,

$$\left(\int_{M(r-x)}^{Mr} \frac{\sin(l)}{\pi l}\, \mathrm{d}l\right)^2 = \left(\int_0^{Mr} \frac{\sin(l)}{\pi l}\, \mathrm{d}l - \int_0^{M(r-x)} \frac{\sin(l)}{\pi l}\, \mathrm{d}l\right)^2$$
$$\leq \left(\frac{1}{\pi|Mr|} + \frac{1}{2}\right)^2 + 2\left(\frac{1}{\pi|Mr|} + \frac{1}{2}\right)\left(\frac{1}{\pi|M(r-x)|} + \frac{1}{2}\right) + \left(\frac{1}{\pi|M(r-x)|} + \frac{1}{2}\right)^2.$$

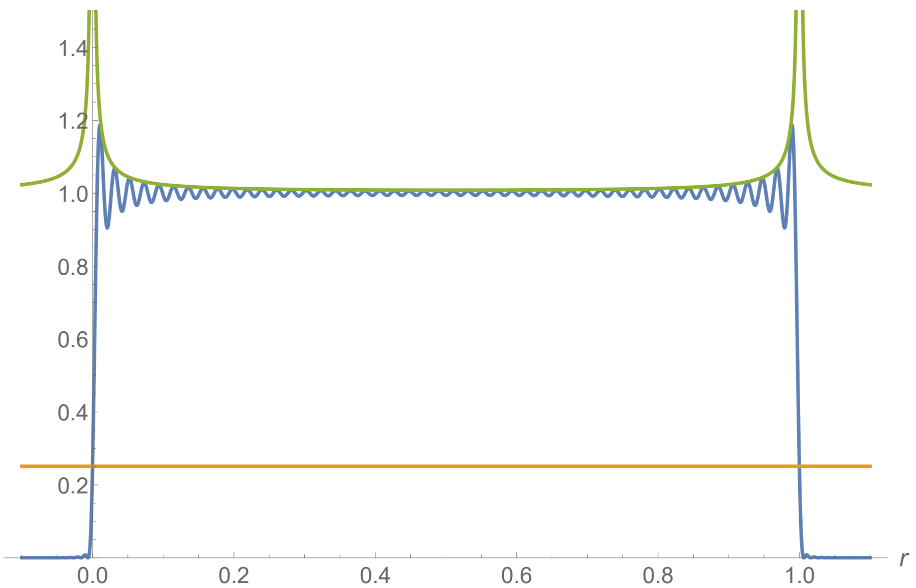

Figure 23: Illustration of the bounds for $M = 300$, $x = 1$.

In Figure 23, the two bounds calculated above are illustrated. The orange line is the bound for $r \in \{0, x\}$ and the green line is the bound for the rest. The only critical parts are close to $r = 0$ and $r = x$, where the function attains its maximum. It is obvious that the maximum value is given by

$$\left(\frac{1}{\pi|Mr_{\max}|} + \frac{1}{2}\right)^2 + 2\left(\frac{1}{\pi|Mr_{\max}|} + \frac{1}{2}\right)\left(\frac{1}{\pi|M(r_{\max}-x)|} + \frac{1}{2}\right) + \left(\frac{1}{\pi|M(r_{\max}-x)|} + \frac{1}{2}\right)^2,$$

where the function attains the maximum value in $r_{\max}$. Now, for $M \geq M_0$, this is bounded by

$$\left(\frac{1}{\pi|M_0 r_{\max}|}+\frac{1}{2}\right)^2 + 2\left(\frac{1}{\pi|M_0 r_{\max}|}+\frac{1}{2}\right)\left(\frac{1}{\pi|M_0(r_{\max}-x)|}+\frac{1}{2}\right) + \left(\frac{1}{\pi|M_0(r_{\max}-x)|}+\frac{1}{2}\right)^2.$$

The part $O(N^{-\frac{1}{2}})$ in Eq. (7.17) is very small for large $M \geq M_0$ and does not change the fact that the function is bounded. We call the bound d_x. This is a function that is integrable because

$$\int_0^\infty d_x f(r) dr = d_x < \infty.$$

With the dominated convergence theorem, it is possible to move the limit over M inside the integral.

12.3 Details for the Proof of Theorem 7.11

We show the fact that was omitted in the proof of Theorem 7.11. The task is to show

$$\mathbb{E}\left[\left(\int_{-\infty}^{x} T_N(X_1,t)\,\mathrm{d}t\right)^2\right] \to \int_{-\infty}^{x} f(r)\,\mathrm{d}r = F(x).$$

With [54, Eq. (A8)], which only holds on compact sets, we know that

$$\mathbb{E}\left[\left(\int_{-\infty}^{x} T_N(X_1,t)\,\mathrm{d}t\right)^2\right] = \lim_{P\to\infty}\int_{-P}^{P}\left[\int_{-\infty}^{x}\frac{\sin(M(r-t))}{\pi(r-t)} + O(N^{-1/2})\,\mathrm{d}t\right]^2 f(r)\,\mathrm{d}r$$
$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}\frac{\sin(M(r-t))}{\pi(r-t)} + O(N^{-1/2})\,\mathrm{d}t\right]^2 f(r)\,\mathrm{d}r.$$

For the inner integral it holds that

$$\int_{-\infty}^{x}\frac{\sin(M(r-t))}{\pi(r-t)}\,\mathrm{d}t = \int_{M(r-x)}^{\infty}\frac{\sin(l)}{\pi l}dl$$

and similar to the proof of Theorem 7.6, it holds that

$$\int_{M(r-x)}^{\infty}\frac{\sin(l)}{\pi l}dl \to \begin{cases}1, & r<x,\\ 0, & r>x.\end{cases}$$

We get that

$$\mathbb{E}\left[\left(\int_{-\infty}^{x} T_N(X_1,t)\,\mathrm{d}t\right)^2\right] = \int_{-\infty}^{x}\Bigg[\underbrace{\int_{-\infty}^{x}\frac{\sin(M(r-t))}{\pi(r-t)} + O(N^{-1/2})\,\mathrm{d}t}_{\to 1}\Bigg]^2 f(r)\,\mathrm{d}r$$
$$+\int_{x}^{\infty}\Bigg[\underbrace{\int_{-\infty}^{x}\frac{\sin(M(r-t))}{\pi(r-t)} + O(N^{-1/2})\,\mathrm{d}t}_{\to 0}\Bigg]^2 f(r)\,\mathrm{d}r \to F(x).$$

It is again possible to exchange the limit and the integral here. The arguments are very similar to the arguments in Section 12.2 with

$$-\frac{1}{\pi|x|} \le \int_{-\infty}^{x}\frac{\sin(l)}{\pi l}\,\mathrm{d}l \le \frac{1}{\pi|x|}.$$

This shows the claim.

13 Notation

- $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i \leq x)$ empirical distribution function based on n observations
- Bernstein estimator
  - $P_{k,m}(x) = \binom{m}{k} x^k (1-x)^{m-k}$ Bernstein basis polynomial
  - $L_m = L_m(x) = \sum_{k=0}^{m} P_{k,m}^2(x)$
  - $R_{j,m} = m^{-j} \sum\sum_{0\leq k<l\leq m} (k-mx)^j P_{k,m}(x) P_{l,m}(x)$
  - $f_{h,n}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)$ Kernel density estimator with kernel K based on n observations with bandwidth h
  - $F_{h,n}(x) = \frac{1}{nh}\sum_{i=1}^{n} \mathbb{K}\left(\frac{x-X_i}{h}\right)$ Kernel distribution estimator with $\mathbb{K}(t) = \int_{-\infty}^{t} K(u)du$ based on n observations with bandwidth h
  - $\hat{F}_{m,n}(x) = \sum_{k=0}^{m} F_n\left(\frac{k}{m}\right) P_{k,m}(x)$ Bernstein estimator with m Bernstein polynomials based on n observations
  - $B_m(F;x) = B_m(x) = \sum_{k=0}^{m} F\left(\frac{k}{m}\right) P_{k,m}(x)$ Bernstein polynomial of order m of F
  - $i_L(n,x) = \min\left\{k \in \mathbb{N} : \text{MSE}[F_k(x)] \leq \text{MSE}\left[\hat{F}_{m,n}(x)\right]\right\}$
  - $i_G(n) = \min\left\{k \in \mathbb{N} : \text{MISE}[F_k] \leq \text{MISE}\left[\hat{F}_{m,n}\right]\right\}$
  - $b(x) = \frac{x(1-x)f'(x)}{2}$
  - $\sigma^2(x) = F(x)(1-F(x))$
  - $V(x) = f(x)\left[\frac{2x(1-x)}{\pi}\right]^{1/2}$
  - $C_1 = \int_0^1 \sigma^2 \,\mathrm{d}x$
  - $C_2 = \int_0^1 V(x) \,\mathrm{d}x$
  - $C_3 = \int_0^1 b^2(x) \,\mathrm{d}x$
  - Notation used in Section 5.7
    - $B_{h,m}(x) = \sum_{k=0}^{m} F_{h,n}\left(\frac{k}{m}\right) P_{k,m}(x)$
    - $\text{Bias}^*\left[\hat{F}_{m,n}(x)\right] = B_{h,m}(x) - F_{h,n}(x)$
    - $\text{Var}^*\left[\hat{F}_{m,n}(x)\right] = \frac{1}{n}\left[\sum_{k=0}^{m} F_n\left(\frac{k}{m}\right) P_{k,m}^2(x) + 2\sum\sum_{0\leq k<l\leq m} F_n\left(\frac{k}{m}\right) P_{k,m}(x)P_{l,m}(x) - \hat{F}_{m,n}^2(x)\right]$

- $\text{MSE}^* \left[\hat{F}_{m,n}(x)\right] = \text{Var}^* \left[\hat{F}_{m,n}(x)\right] + \text{Bias}^* \left[\hat{F}_{m,n}(x)\right]^2$
- $\widehat{m} = \lceil \text{argmin}_{k \in I_n} \text{MSE}^* \left[\hat{F}_{m,n}(x)\right] \rceil$ Locally optimal m as calculated in Eq. (5.17)
- $\widehat{\widehat{m}} = \lceil \text{argmin}_{k \in I_n} \text{MISE}^* \left[\hat{F}_{m,n}\right] \rceil$ Globally optimal m as calculated in Eq. (5.18)

- Szasz estimator
  - $V_{k,m}(x) = e^{-mx} \dfrac{(mx)^k}{k!}$
  - $L_m^S(x) = \sum_{k=0}^{\infty} V_{k,m}^2(x)$
  - $R_{j,m}^S(x) = m^{-j} \sum\sum_{0 \le k < l \le \infty} (k - mx)^j V_{k,m}(x) V_{l,m}(x)$
  - $S_m(F;x) = \sum_{k=0}^{\infty} F\left(\dfrac{k}{m}\right) e^{-mx} \dfrac{(mx)^k}{k!} = \sum_{k=0}^{\infty} F\left(\dfrac{k}{m}\right) V_{k,m}(x)$
  - $\hat{f}_{m,n}^S(x) = \dfrac{m}{n} \sum_{m=0}^{\infty} B_{k,m}^{(n)} e^{-mx} \dfrac{(mx)^k}{k!}$
  - $\hat{F}_{m,n}^S(x) = \sum_{k=0}^{\infty} F_n\left(\dfrac{k}{m}\right) e^{-mx} \dfrac{(mx)^k}{k!}$
  - $b^S(x) = \dfrac{xf'(x)}{2}$
  - $V^S(x) = f(x) \left[\dfrac{x}{\pi}\right]^{1/2}$
  - $C_1^S = \int_0^{\infty} \sigma^2 e^{-ax} f(x)\,\mathrm{d}x$
  - $C_2^S = \int_0^{\infty} V^S(x) e^{-ax} f(x)\,\mathrm{d}x$
  - $C_3^S = \int_0^{\infty} (b^S(x))^2 e^{-ax} f(x)\,\mathrm{d}x$
  - $i_L^S(n,x) = \min\left\{k \in \mathbb{N} : \text{MSE}[F_k(x)] \le \text{MSE}\left[\hat{F}_{m,n}^S(x)\right]\right\}$
  - $i_G^S(n,x) = \min\left\{k \in \mathbb{N} : \text{MISE}[F_k(x)] \le \text{MISE}\left[\hat{F}_{m,n}^S(x)\right]\right\}$

- Hermite estimator
  - $H_k(x) = (-1)^k e^{x^2} \dfrac{d^k}{\mathrm{d}x^k} e^{-x^2} = k! \sum_{m=0}^{\lfloor k/2 \rfloor} \dfrac{(-1)^m}{m!(k-2m)!} (2x)^{k-2m}$ Hermite polynomial
  - $h_k = (2^k k! \sqrt{\pi})^{-1/2} e^{\frac{-x^2}{2}} H_k(x)$ normalized Hermite function
  - $H_{e_k}(x) = 2^{-\frac{k}{2}} H_k\left(\frac{x}{\sqrt{2}}\right)$ Chebyshev-Hermite polynomials
  - $Z(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
  - $\alpha_k = \dfrac{\sqrt{\pi}}{2^{k-1} k!}$

- $a_k = \int\limits_{-\infty}^{\infty} f(x) h_k(x)\,\mathrm{d}x$
- $f_N(x) = \sum\limits_{k=0}^{N} a_k h_k(x) = \sum\limits_{k=0}^{N} \sqrt{\alpha_k} \cdot a_k H_k(x) Z(x)$
- $\hat{a}_k = \frac{1}{n} \sum\limits_{i=1}^{n} h_k(X_i)$
- $\hat{f}_{N,n}(x) = \sum\limits_{k=0}^{N} \hat{a}_k h_k(x) = \sum\limits_{k=0}^{N} \sqrt{\alpha_k} \cdot \hat{a}_k H_k(x) Z(x)$ Gauss-Hermite density estimator
- $\hat{F}^{H}_{N,n}(x) = \int\limits_{0}^{x} \hat{f}_{N,n}(t)\,\mathrm{d}t$ Gauss-Hermite distribution estimator on the real half line
- $\hat{F}^{F}_{N,n}(x) = \int\limits_{-\infty}^{x} \hat{f}_{N,n}(t)\,\mathrm{d}t$ Gauss-Hermite distribution estimator on the real line
- $\mathrm{IF}(x, x'; T, F) = \lim\limits_{\epsilon \to 0} \frac{T(x, (1-\epsilon)F + \epsilon\delta_{x'}) - T(x, F)}{\epsilon}$ influence function evaluated at a point x and distribution F

References

[1] R. Lockhart, “The Basics of Nonparametric Models.”

[2] M. Rosenblatt, “Remarks on Some Nonparametric Estimates of a Density Function,” *The Annals of Mathematical Statistics*, vol. 27, no. 3, pp. 832–837, Sep. 1956.

[3] E. Parzen, “On Estimation of a Probability Density Function and Mode,” *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.

[4] A. Gramacki, *Nonparametric Kernel Density Estimation and Its Computational Aspects*, ser. Studies in Big Data. Springer International Publishing, 2018.

[5] H. Yamato, “Uniform Convergence of an Estimator of a Distribution Function,” *Bull. Math. Statist.*, vol. 15, no. 3, pp. 69–78, Mar. 1973.

[6] C. Kim, S. Kim, M. Park, and H. Lee, “A Bias Reducing Technique in Kernel Distribution Function Estimation,” *Computational Statistics*, vol. 21, no. 3, pp. 589–601, Dec. 2006.

[7] C. Shalizi, “Estimating Distributions and Densities,” Nov. 2009.

[8] S. Zhang, Z. Li, and Z. Zhang, “Estimating a Distribution Function at the Boundary,” *Austrian Journal of Statistics*, vol. 49, no. 1, pp. 1–23, Feb. 2020.

[9] G. S. Watson and M. R. Leadbetter, “Hazard Analysis II,” 1964.

[10] A. Azzalini, “A Note on the Estimation of a Distribution Function and Quantiles by a Kernel Method,” *Biometrika*, vol. 68, no. 1, pp. 326–328, Apr. 1981.

[11] M. C. Jones, “The Performance of Kernel Density Functions in Kernel Distribution Function Estimation,” *Statistics & Probability Letters*, vol. 9, no. 2, pp. 129–132, Feb. 1990.

[12] S. Bernstein, “Démonstration Du Théorème De Weierstrass Fondée Sur La Calcul Des Probabilités,” p. 3, 1912.

[13] G. J. Babu, A. J. Canty, and Y. P. Chaubey, “Application of Bernstein Polynomials for Smooth Estimation of a Distribution and Density Function,” *Journal of Statistical Planning and Inference*, vol. 105, no. 2, pp. 377–392, Jul. 2002.

[14] V. S. Videnskii, “Papers of L.V. Kantorovich on Bernstein Polynomials,” *Vestnik St. Petersburg University: Mathematics*, vol. 46, no. 2, pp. 85–88, Apr. 2013.

[15] R. A. Vitale, “A Bernstein Polynomial Approach to Density Function Estimation,” 1975.

[16] A. Leblanc, “Personal Communication,” Feb. 2020.

[17] A. Leblanc, “On Estimating Distribution Functions Using Bernstein Polynomials,” *Annals of the Institute of Statistical Mathematics*, vol. 64, no. 5, pp. 919–943, Oct. 2012.

[18] A. Leblanc, “Chung–Smirnov Property for Bernstein Estimators of Distribution Functions,” *Journal of Nonparametric Statistics*, vol. 21, no. 2, pp. 133–142, Feb. 2009.

[19] G. G. Lorentz, *Bernstein Polynomials*, 2nd ed. New York, N.Y: Chelsea Pub. Co, 1986.

[20] A. Leblanc, "On the Boundary Properties of Bernstein Polynomial Estimators of Density and Distribution Functions," *Journal of Statistical Planning and Inference*, vol. 142, no. 10, pp. 2762–2778, Oct. 2012.

[21] N. L. Hjort and S. G. Walker, "A Note on Kernel Density Estimators with Optimal Bandwidths," *Statistics & Probability Letters*, vol. 54, no. 2, pp. 153–159, Sep. 2001.

[22] A. Leblanc, "A Bias-Reduced Approach to Density Estimation Using Bernstein Polynomials," *Journal of Nonparametric Statistics*, vol. 22, no. 4, pp. 459–475, May 2010.

[23] J. L. Hodges and E. L. Lehmann, "Deficiency," *The Annals of Mathematical Statistics*, vol. 41, no. 3, pp. 783–801, Jun. 1970.

[24] M. Falk, "Relative Efficiency and Deficiency of Kernel Type Estimators of Smooth Distribution Functions," *Statistica Neerlandica*, vol. 37, no. 2, pp. 73–83, 1983.

[25] S. Dutta, "Distribution Function Estimation Via Bernstein Polynomial of Random Degree," *Metrika*, vol. 79, no. 3, pp. 239–263, Apr. 2016.

[26] S. Dutta, "Local Smoothing for Kernel Distribution Function Estimation," *Communications in Statistics - Simulation and Computation*, vol. 44, no. 4, pp. 878–891, 2013.

[27] N. Altman and C. Léger, "Bandwidth Selection for Kernel Distribution Function Estimation," *Journal of Statistical Planning and Inference*, vol. 46, no. 2, pp. 195–214, Aug. 1995.

[28] F. Ouimet, "Complete Monotonicity of Multinomial Probabilities and Its Application to Bernstein Estimators on the Simplex," *Journal of Mathematical Analysis and Applications*, vol. 466, no. 2, pp. 1609–1617, Oct. 2018.

[29] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed., ser. A Wiley Publication in Mathematical Statistics. New York: Wiley, 1965.

[30] N. Cressie, "A Finely Tuned Continuity Correction," *Annals of the Institute of Statistical Mathematics*, vol. 30, no. 3, pp. 435–442, Dec. 1978.

[31] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, 9th ed., ser. Dover Books on Mathematics. New York, NY: Dover Publ, 1964, oCLC: 935935300.

[32] P. Billingsley, *Probability and Measure*, 3rd ed., ser. Wiley Series in Probability and Mathematical Statistics. New York: Wiley, 1995.

[33] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, 1980, oCLC: 959994695.

[34] D. H. Greene and D. E. Knuth, *Mathematics for the Analysis of Algorithms.* Springer Science & Business Media, May 2009.

[35] F. Merlevède, M. Peligrad, and E. Rio, *Bernstein Inequality and Moderate Deviations Under Strong Mixing Conditions.* Institute of Mathematical Statistics, 2009.

[36] O. Szasz, "Generalization of S. Bernstein's Polynomials to the Infinite Interval," *Journal of Research of the National Bureau of Standards*, vol. 45, Sep. 1950.

[37] W. Gawronski and U. Stadtmueller, "On Density Estimation by Means of Poisson's Distribution," *Scandinavian Journal of Statistics*, vol. 7, Jan. 1980.

[38] P. Hall, "A Unified Approach to the Correction of Normal Approximations," *SIAM Journal on Applied Mathematics*, vol. 43, no. 5, pp. 1187–1193, Oct. 1983.

[39] M. Stephanou, M. Varughese, and I. Macdonald, "Sequential Quantiles Via Hermite Series Density Estimation," *Electronic Journal of Statistics*, vol. 11, no. 1, pp. 570–607, 2017.

[40] Stephanou, "On the Properties of Hermite Series Based Distribution Function Estimators," 2020.

[41] G. G. Walter, "Properties of Hermite Series Estimation of Probability Density," *The Annals of Statistics*, vol. 5, no. 6, pp. 1258–1264, 1977.

[42] A. Jmaei, Y. Slaoui, and W. Dellagi, "Recursive Distribution Estimator Defined by Stochastic Approximation Method Using Bernstein Polynomials," *Journal of Nonparametric Statistics*, p. 15.

[43] G. Szegö, *Orthogonal Polynomials.* American Mathematical Society, 1959.

[44] W. Greblicki and M. Pawlak, "Hermite Series Estimates of a Probability Density and Its Derivatives," *Journal of Multivariate Analysis*, vol. 15, no. 2, pp. 174–182, Oct. 1984.

[45] S. Blinnikov and R. Moessner, "Expansions for Nearly Gaussian Distributions," *Astronomy and Astrophysics Supplement Series*, vol. 130, no. 1, pp. 193–205, May 1998.

[46] J. Puuronen and A. Hyvärinen, "Hermite Polynomials and Measures of Non-Gaussianity," in *Artificial Neural Networks and Machine Learning – ICANN 2011*, ser. Lecture Notes in Computer Science, T. Honkela, W. Duch, M. Girolami, and S. Kaski, Eds. Berlin, Heidelberg: Springer, 2011, pp. 205–212.

[47] H. F. Davis, *Fourier Series and Orthogonal Functions / Harry F. Davis*, 1963.

[48] F. R. Hampel, "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, 1974.

[49] F. R. Hampel, *Robust Statistics: The Approach Based on Influence Functions*, 1st ed. New York: Wiley-Interscience, Mar. 2005.

[50] M. Stephanou, "Personal Communication," Feb. 2020.

[51] B. P. Welford, "Note on a Method for Calculating Corrected Sums of Squares and Products," *Technometrics*, vol. 4, no. 3, pp. 419–420, Aug. 1962.

[52] W. Greblicki and M. Pawlak, "Pointwise Consistency of the Hermite Series Density Estimate," *Statistics & Probability Letters*, vol. 3, no. 2, pp. 65–69, 1985.

[53] "DLMF: NIST Digital Library of Mathematical Functions," https://dlmf.nist.gov/.

[54] E. Liebscher, "Hermite Series Estimators for Probability Densities," *Metrika*, vol. 37, no. 1, pp. 321–343, Dec. 1990.

[55] S. W. Dharmadhikari and K. Jogdeo, "Bounds on Moments of Certain Random Variables," *The Annals of Mathematical Statistics*, vol. 40, no. 4, pp. 1506–1509, Aug. 1969.

[56] K. I. Joy, "Bernstein Polynomials," *On-Line Geometric Modeling Notes.*

14 Erklärung

Ich versichere wahrheitsgemäß, die Arbeit selbstständig verfasst, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde, sowie die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.

Ort, den Datum