Smooth distribution function estimation for lifetime distributions using Szasz–Mirakyan operators

Ariane Hanebeck[1] · **Bernhard Klar**[2]

Abstract
In this paper, we introduce a new smooth estimator for continuous distribution func-tions on the positive real half-line using Szasz–Mirakyan operators, similar to Bern-stein's approximation theorem. We show that the proposed estimator outperforms the empirical distribution function in terms of asymptotic (integrated) mean-squared error and generally compares favorably with other competitors in theoretical com-parisons. Also, we conduct the simulations to demonstrate the finite sample perfor-mance of the proposed estimator.

Keywords Distribution function estimation · Nonparametric · Szasz–Mirakyan operator · Hermite estimator · Mean squared error

1 Introduction

This paper considers the nonparametric smooth estimation of continuous distribution functions on the positive real half line. Arguably, such distributions are the most important univariate probability models, occurring in diverse fields such as life sci-ences, engineering, actuarial sciences or finance, under various names such as life, lifetime, loss or survival distributions. The well-known compendium of (Johnson et al. 1994) treats in its first volume solely distributions on the positive half line with the exception of the normal and the Cauchy distribution. In the two volumes (Johnson et al. 1994, 1995) as well as in the compendiums about life and loss distri-butions of (Marshall and Olkin 2007) and (Hogg and Klugman 1984), respectively,

✉ Ariane Hanebeck
Ariane.Hanebeck@tum.de

Bernhard Klar
bernhard.klar@kit.edu

[1] Institute of Applied Mathematical Statistics, Technical University of Munich, Boltzmannstraße 3, 85748 Garching b. München, Germany

[2] Institute of Stochastics, Karlsruhe Institute of Technology, Englerstr. 2, 76131 Karlsruhe, Germany

an abundance of parametric models for the distribution of nonnegative random variables and pertaining estimation methods can be found. However, there is a paucity of nonparametric estimation methods especially tailored to this situation. It is the aim of this paper to close this gap by introducing a new nonparametric estimator for distribution functions on $[0, \infty)$ using Szasz–Mirakyan operators.

Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed (i.i.d.) random variables having an underlying unknown distribution function F and associated density function f. In this paper, all the considered distribution function estimators are of nonparametric type. The best-known distribution function estimators with well-established properties are the empirical distribution function (EDF) and the kernel estimator. The EDF is the simplest way to estimate the underlying distribution function, given a finite random sample $X_1, \dots, X_n, n \in \mathbb{N}$. It 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. This estimator is obviously not continuous. The kernel distribution function estimator

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

is continuous, where $\mathbb{K}(t) = \int_{-\infty}^{t} K(u)\mathrm{d}u$ is a cumulative kernel function of a kernel $K : \mathbb{R} \to \mathbb{R}$, which has to fulfill specific properties (see, e.g., (Gramacki 2018)). The parameter $h > 0$ is called the bandwidth. This estimator was first introduced by (Yamato 1973). The corresponding kernel density estimator $f_{h,n}$ was first introduced by (Rosenblatt 1956) and (Parzen 1962). An important task in kernel density estimation is the choice of the bandwidth. In (Duin 1976) and (Rudemo 1982), this topic was addressed. (Slaoui 2014) presented a method to automatically select the parameter with the help of the stochastic approximation algorithm by (Mokkadem et al. 2009). Different methods to choose the bandwidth in the case of the distribution function are given in (Altman and Léger 1995), (Bowman et al. 1998), (Polansky and Bake 2000), and (Tenreiro 2006).

The two estimators can estimate distribution functions on any arbitrary real interval. The Bernstein estimator, on the other hand, is designed for functions on [0, 1]. (Babu et al. 2002) and (Leblanc 2012) studied the Bernstein 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. In (Jmaei et al. 2017), a recursive estimator using Bernstein polynomials was introduced. (Helali and Slaoui 2020) used Lagrange polynomials with Tchebychev–Gauss points, instead of Bernstein polynomials. A further estimator is the Hermite distribution function estimator on the real half line, see Sect. 4 and (Stephanou et al.

2017). The Hermite density estimator was first defined by (Schwartz 1967). More information on the different estimators can be found in the cited literature and in (Hanebeck 2020). For ease of reference, many properties of the estimators are listed in Sect. 4.

In this paper, we consider the Szasz estimator, as an alternative estimator of the distribution function on $[0, \infty)$. The kernel estimator can also estimate functions on $[0, \infty)$ but is not specifically designed for this interval. To get satisfactory results, special boundary corrections in the point zero are necessary (see (Zhang et al. 2020)), which is not the case for the Szasz estimator. The Hermite estimator on the real half line is designed for $[0, \infty)$, but theoretical results and simulations later show that the Szasz estimator performs better on the positive real line.

The paper is organized as follows. In Sect. 2, the approach and most important properties of the proposed estimator are explained. Then, in Sect. 3, we derive asymptotic properties of the estimator. In Sect. 4, the properties are compared with other estimators in a theoretical comparison and then in a simulation study in Sect. 5. Section 6 concludes the paper. Most of the proofs are similar to (Leblanc 2012), except for the use of Poisson probabilities (Szasz–Mirakyan operators for the semi-infinite interval $[0, \infty)$), instead of binomial probabilities (Bernstein operators for the interval [0, 1]). There are, however, differences and extensions that are mentioned after the respective results if necessary. All proofs can be found in the authors' arXiv paper (Hanebeck and Klar 2020), with the same title "arXiv:2005.09994".

Throughout the paper, the notation $f = o(g)$ means that $\lim |f/g| = 0$ as $m, n \to \infty$. A subscript (for example $f = o_x(g)$) indicates which parameters the convergence rate can depend on. Furthermore, the notation $f = O(g)$ means that $\limsup |f/g| < C$ for $m, n \to \infty$ and some $C \in (0, \infty)$. A subscript in this case means that C could depend on the corresponding parameter.

2 The Szasz Distribution Function Estimator

The idea of the estimator presented in this paper is similar to the Bernstein approach. The main difference is that instead of the Bernstein basis polynomials, we use Poisson probabilities. Hence, in the former case, we consider $\operatorname{supp}(f) = [0, 1]$, while the latter case assumes $\operatorname{supp}(f) = [0, \infty)$. We make use of the following claim that can be found in (Szasz 1950).

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

$$S_m(u;x) = \sum_{k=0}^{\infty} u\left(\frac{k}{m}\right) V_{k,m}(x) \to u(x) \tag{1}$$

uniformly for $x \in (0, \infty)$, where $V_{k,m}(x) = e^{-mx} \frac{(mx)^k}{k!}$ for $k, m \in \mathbb{N}$.

The operator $S_m(u;x)$ is called the Szasz–Mirakyan operator of the function u at the point x. One can the extend above claim 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 $S_m(u;x)\to u(x)$ uniformly for $x\in[0,\infty)$. In particular, given a continuous distribution function F on $[0,\infty)$, Eq. 1 remains valid, uniformly for $x\in[0,\infty)$. Then, a possible estimator of F on $[0,\infty)$ is

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

replacing the unknown distribution function F in the operator $S_m(F;x)$ by the EDF F_n. We call $\hat{F}^S_{m,n}$ the Szasz estimator. The sum is infinite but can be written as a finite sum as shown in the next subsection.

In the remainder of this paper, we make the following general assumption:

Assumption 1 The distribution function F is continuous. The first and second derivatives f and f' of F are continuous and bounded on $[0,\infty)$.

2.1 Basic Properties of the Szasz Estimator

The behavior of the Szasz estimator $\hat{F}^S_{m,n}(x)$ at $x=0$ is appropriate, since we get $\hat{F}^S_{m,n}(0)=0=F(0)=S_m(F;0)$. This means that bias and variance at the point $x=0$ are zero.

In the sequel, we use the gamma function $\Gamma(z)=\int_0^\infty x^{z-1}e^{-x}dx$, as well as the upper and lower incomplete gamma functions, defined by

$$\Gamma(z,s)=\int_s^\infty x^{z-1}e^{-x}dx,\ \text{and}\ \gamma(z,s)=\int_0^s x^{z-1}e^{-x}dx,$$

respectively. Note that $\lim_{x\to\infty}\hat{F}^S_{m,n}(x)=1=\lim_{x\to\infty}F(x)=\lim_{x\to\infty}S_m(F;x)$, since

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

where the random variable Y has a Poisson distribution with expected value mx ($Y\sim$ Po(mx) for short). Since the above representation only contains a finite number of summands, it can be used to easily simulate the estimator.

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

Theorem 1 *The function* $\hat{F}^S_{m,n}(x)$ *is non-decreasing in* x *on* $[0, \infty)$.

Proof This proof is similar to the proof for the Bernstein estimator that can be found in (Babu et al. 2002), but now applied to Poisson probabilities instead of Bernstein polynomials. Let

$$g_n(0) = 0, g_n\left(\frac{k}{m}\right) = F_n\left(\frac{k}{m}\right) - F_n\left(\frac{k-1}{m}\right), \; k = 1, 2, \ldots,$$

and

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

Then,

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

The claim follows as $g_n\left(\frac{k}{m}\right)$ is nonnegative for at least one k and $U_k(m, x)$ is increasing. □

The next theorem shows that $\hat{F}^S_{m,n}(x)$ is uniformly strongly consistent. Its proof follows the proof of Theorem 2.1 in (Babu et al. 2002).

Theorem 2 *If* F *is a continuous probability distribution function on* $[0, \infty)$, *then*

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

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

3 Asymptotic properties of the Szasz estimator

3.1 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 $x = 0$. The following lemma is similar to (Lorentz 1986, Sect. 1.6.1).

Lemma 1 *We have, for $x \in (0, \infty)$ that*

$$\mathbb{E}[F^S_{m,n}(x)] = S_m(F;x) = F(x) + m^{-1}b^S(x) + o_x(m^{-1}),$$

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

The following theorem establishes asymptotic expressions for the bias and the variance of the Szasz estimator $\hat{F}^S_{m,n}$ as $m, n \to \infty$. The statement is similar to Theorem 1 in Leblanc (2012) but applied to the Szasz estimator. The bias follows directly from Lemma 1. For the proof of the variance, ideas of (Ouimet 2020) and (Leblanc 2012) have to be combined.

Theorem 3 *For each $x \in (0, \infty)$,*

$$\text{Bias}\left[\hat{F}^S_{m,n}(x)\right] = m^{-1}b^S(x) + o_x(m^{-1}),$$

and

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

where $b^S(x)$ is defined in Lemma 1 and

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

3.2 Asymptotic normality

Here, we turn our attention to the asymptotic behavior of the Szasz estimator $\hat{F}^S_{m,n}(x)$, similar to Theorem 2 in (Leblanc 2012).

Theorem 4 *Let $x \in (0, \infty)$, such that $0 < F(x) < 1$. Then, for $m, n \to \infty$,*

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

where $\sigma^2(x)$ is defined in Theorem 3.

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 1, it is easy to see that

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

This leads directly to the following corollary, which is similar to Corollary 2 in (Leblanc 2012).

Corollary 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*

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

where $\sigma^2(x)$ and $b^S(x)$ are defined in Lemma 1 and Theorem 3.

3.3 Asymptotically optimal *m* with respect to mean-squared error

For the estimator $\hat{F}^S_{m,n}$, it is interesting to calculate the mean-squared error (MSE)

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

and the asymptotically optimal m with respect to MSE. The MSE at $x = 0$ is zero. For $(0, \infty)$, the asymptotic MSE and the optimal m can easily be obtained from Theorem 3, i.e.,

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

for $x \in (0, \infty)$. Using Eq. 2 and assuming $f(x) \neq 0$ and $f'(x) \neq 0$, the asymptotically optimal choice of m for estimating $F(x)$ with respect to MSE is

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

Therefore, the optimal MSE can be written as

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

for $x \in (0, \infty)$, where $\sigma^2(x)$, $b^S(x)$, and $V^S(x)$ are defined in Lemma 1 and Theorem 3.

3.4 Asymptotically optimal *m* with respect to mean-integrated squared error

We now focus on the mean-integrated squared error (MISE). As we deal with an infinite integral, we use a nonnegative weight function ω. Here, the weight function is chosen as $\omega(x) = e^{-ax}f(x)$. Following (Altman and Léger 1995), the MISE is then defined by

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

Technically, $\mathrm{MISE}\left[\hat{F}^S_{m,n}\right]$ cannot be calculated by integrating the expression of $\mathrm{MSE}\left[\hat{F}^S_{m,n}\right]$ obtained in Eq. 2 as the asymptotic expressions depend on x. The next theorem gives the asymptotic MISE of the Szasz operator and is similar to Theorem 3 in (Leblanc 2012). One big difference is the extension e^{-ax} to the weight function here.

Theorem 5 *We have*

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

with

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

where $\sigma^2(x)$, $b^S(x)$, and $V^S(x)$ are defined in Lemma 1 and Theorem 3.

Very similar to Corollary 4 in (Leblanc 2012), the next corollary gives the asymptotically optimal m for estimating F with respect to MISE.

Corollary 2 *The asymptotically optimal m for estimating F with respect to MISE is*

$$m_{opt} = n^{2/3}\left[\frac{4C^S_3}{C^S_2}\right]^{2/3},$$

which leads to the optimal MISE

$$\mathrm{MISE}\left[\hat{F}^S_{m_{opt},n}\right] = n^{-1}C^S_1 - \frac{3}{4}n^{-4/3}\left[\frac{(C^S_2)^4}{4C^S_3}\right]^{1/3} + o(n^{-4/3}). \tag{4}$$

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 term (including the minus sign ahead of it) in Eqs. 3 and 4 is always negative so that the Szasz estimator seems to outperform the EDF. This is proven in the following.

3.5 Asymptotic deficiency of the empirical distribution function

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

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

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 4 in (Leblanc 2012). A result in a similar form for kernel estimators can be found in (Falk 1983).

Theorem 6 *Let $x \in (0,\infty)$ and $m,n \to \infty$. If $mn^{-1/2} \to \infty$, then,*

$$i^S_L(n,x) = \lceil n[1 + o_x(1)]\rceil \text{ and } i^S_G(n) = \lceil n[1 + o(1)]\rceil.$$

In addition, the following statements are true.

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

$$i^S_L(n,x) - n = m^{-1/2}n\left[V^S(x)/\sigma^2(x) + o_x(1)\right], \text{ and}$$
$$i^S_G(n) - n = m^{-1/2}n[C^S_2/C^S_1 + o(1)].$$

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

$$i^S_L(n,x) - n = n^{2/3}\left[c^{-1/2}V^S(x)/\sigma^2(x) - c^{-2}(b^S(x))^2/\sigma^2(x) + o_x(1)\right], \text{ and}$$
$$i^S_G(n) - n = n^{2/3}[c^{-1/2}C^S_2/C^S_1 - c^{-2}C^S_3/C^S_1 + o(1)].$$

Here, $V^S(x)$, $\sigma^2(x)$, and $b^S(x)$ are defined in Lemma 1 and Theorem 3, and C^S_1, C^S_2, and C^S_3 are defined in Theorem 5.

Parts a) and b) of this theorem show 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 also has to increase to infinity so that the EDF outperforms the Szasz estimator. Hence, the EDF is asymptotically deficient to the Szasz estimator.

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 Szasz estimator outperforms the EDF as much as possible.

Lemma 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 2.*

4 Theoretical comparison

In the following, the properties that were derived in this paper for the Szasz estimator are compared to the different estimators defined in the introduction. The comparison can be found in Tables 1,2, 3, and 4. 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. More details can be found in (Hanebeck 2020).

For the EDF, the properties mainly follow from famous theorems. The uniform, almost sure convergence follows from the Glivenko–Cantelli theorem, while the asymptotic normality can be proven with the central limit theorem. The MSE can be found in (Lockhart 2013), and the other properties are easy to calculate. For the kernel estimator, the asymptotic normality can be found in Watson and (Leadbetter 1964) and (Zhang et al. 2020), while bias and variance can be found in (Kim et al. 2006). The optimal MSE and MISE can be found in Zhang et al. (2020). The properties for the Bernstein estimator mainly follow from (Leblanc 2012), where some results are using ideas from (Babu et al. 2002). The ideas and most of the proofs for the Hermite estimator can be found in (Stephanou et al. 2017).

The next result on the asymptotic normality of the Hermite estimator (Stephanou et al. 2017) is new. We first quickly introduce the Hermite estimator. It makes use of the so-called Hermite polynomials $H_k(x) = (-1)^k e^{x^2} \frac{d^k}{\mathrm{d}x^k} e^{-x^2}$. These polynomials are orthogonal under e^{-x^2}. The Hermite distribution function estimator on the real half line is defined by

$$\hat{F}^H_{N,n}(x) = \int_0^x \hat{f}_{N,n}(t)\mathrm{d}t.$$

Here, $\hat{f}_{N,n}(x) = \sum_{k=0}^N \hat{a}_k h_k(x) = \sum_{k=0}^N \sqrt{\alpha_k} \cdot \hat{a}_k H_k(x) Z(x)$ with $\hat{a}_k = \frac{1}{n} \sum_{i=1}^n h_k(X_i)$.

Theorem 7 *For* $x \in (0, \infty)$ *with* $0 < F(x) < 1$ *and if f is differentiable in x, we obtain*

$$\sqrt{n}\left(\hat{F}^H_{N,n}(x) - \mathbb{E}\left[\hat{F}^H_{N,n}(x)\right]\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right),$$

as $n \to \infty$, *where* $\sigma^2(x) = F(x)(1 - F(x))$.

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 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

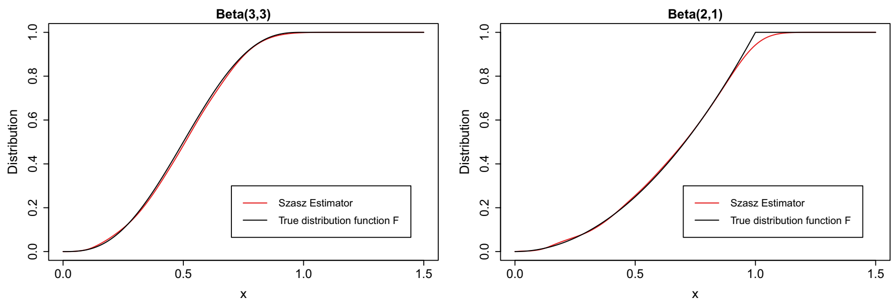

Fig. 1 The behavior of the Szasz estimator at $x = 1$ for $n = 500$

Table 1 Support of the estimators and assumptions

	Support	Assumptions
EDF	Chosen Freely	
Kernel	Chosen Freely	Density f exists, f' exists and is continuous
Bernstein	$[0, 1]$	F continuous, two continuous and bounded derivatives on $[0, 1]$
Szasz	$[0, \infty)$	F continuous, two continuous and bounded derivatives on $[0, \infty)$
Hermite Half	$[0, \infty)$	$f \in L_2$

Table 2 Convergence behavior and asymptotic distribution of the estimators

	Convergence	Asymptotic distribution: $n^{1/2}(\hat{F}_n(x) - F(x)) \xrightarrow{D} \mathcal{N}\big(0, \sigma^2(x)\big)^1$
EDF	a.s. uniform	
Kernel	a.s. uniform	For $h^{-2}n^{-1/2} \to \infty$
Bernstein	a.s. uniform	For $mn^{-1/2} \to \infty$
Szasz	a.s. uniform	For $mn^{-1/2} \to \infty$
Hermite Half	a.s. uniform[2]	For $N^{r/2-1/4}n^{-1/2} \to \infty^3$

[1] $\hat{F}_n$ stands for all of the estimators, for $x : 0 < F(x) < 1$

[2] For $\left(x - \frac{d}{dx}\right)^r f \in L_2, r \geq 1, \mathbb{E}[|X|^s] < \infty, s > \frac{8(r+1)}{3(2r+1)}, N \sim n^{\frac{2}{2r+1}}$

[3] For $\left(x - \frac{d}{dx}\right)^r f \in L_2, r \geq 1, \mathbb{E}\left[|X|^{2/3}\right] < \infty$

Table 3 Bias and variance of the estimators

	Bias	Variance
EDF	Unbiased	$O(n^{-1})$
Kernel	$o(h^2)$	$O(n^{-1})+O(h/n)$
Bernstein	Zero at $\{0,1\}$, $O(m^{-1})=O(h)$	Zero at $\{0,1\}$, $O(n^{-1})+O_x(m^{-1/2}n^{-1})$
Szasz	Zero at 0 , $O_x(m^{-1})=O_x(h)$	Zero at 0, $O(n^{-1})+O_x(m^{-1/2}n^{-1})$
Hermite Half	Zero at 0, $O_x\left(N^{-r/2+1/4}\right)^4$	Zero at 0, $O_x(N^{3/2}/n)^5$

[4]For $\left(x-\frac{d}{dx}\right)^r f\in L_2, r\geq 1, \mathbb{E}\left[|X|^{2/3}\right]<\infty$
[5]For $\mathbb{E}\left[|X|^{2/3}\right]<\infty$

Table 4 MSE and MISE of the estimators

MSE (all consistent)	MISE (all consistent)
$O(n^{-1})$	$O(n^{-1})$
$O(n^{-1})+O\left(h^4\right)+O(h/n)$, Optimal: $O(n^{-1})$	$O(n^{-1})+O(h^4)+O(h/n)$, Optimal: $O(n^{-1})$
Zero at $\{0,1\}$, $O(n^{-1})+O(m^{-2})+O_x(m^{-1/2}n^{-1})$, Optimal: $O_x(n^{-1})$	$O(n^{-1})+O(m^{-2})+O(m^{-1/2}n^{-1})$, Optimal: $O(n^{-1})$
Zero at 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: $O(n^{-1})$
Zero at 0 $x\left[O\left(\frac{N^{1/2}}{n}\right)+O(N^{-r})\right]$, Optimal: $xO(n^{\frac{-2r}{2r+1}})^4$	$\mu\left[O\left(\frac{N^{1/2}}{n}\right)+O(N^{-r})\right]$, Optimal: $\mu O(n^{-\frac{2r}{2r+1}})$

[4]For $\left(x-\frac{d}{dx}\right)^r f\in L_2, r\geq 1, \mathbb{E}\left[|X|^{2/3}\right]<\infty$
[6]Note that the MISE here is defined differently with weight function e^{-ax}
[7] For $\left(x-\frac{d}{dx}\right)^r f\in L_2, r\geq 1, \mu=\int_0^\infty xf(x)dx<\infty$

$$\frac{d}{dx}\hat{F}^S_{m,n}(x)=m\sum_{k=0}^{\infty}\left[F_n\left(\frac{k+1}{m}\right)-F_n\left(\frac{k}{m}\right)\right]V_{k,m}(x)$$

and cannot approximate a non-continuous function that well. This can be seen in Fig. 1. It is obvious that the behavior of the Szasz estimator at $x=1$ of the Beta(2, 1) -distribution is worse.

For the Hermite estimator, the properties $f\in L_2$ and $\left(x-\frac{d}{dx}f\right)^r f\in L_2$ only have to hold on the considered interval. Hence, they can be used for smaller intervals than what they were designed for.

The EDF and the kernel distribution function 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 $(-\infty,\infty)$. Unless the support is $(-\infty,\infty)$, the results do not hold for the points close to the boundary. For an approach to improve this boundary behavior, see (Zhang et al. 2020) for example.

4.1 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, 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 and the Hermite estimator, the order is not uniform. 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 order of the Hermite estimator is 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 rate of the Hermite estimator is worse but for $r \to \infty$, the rate approaches n^{-1}. This is very similar for the optimal rates of the MISE.

5 Simulation

In this section, the different estimators are compared in a simulation study with respect to the MISE. For the kernel distribution function estimator, the Gaussian kernel is chosen, i.e. $F_{h,n}(x) = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{x-X_i}{h}\right)$, where Φ is the standard normal distribution function.

The simulation consists of two parts. In the first part, the estimators are compared by their MISE on $[0, \infty)$ with respect to

$$\text{MISE}\left[\hat{F}_n\right] = \mathbb{E}\left[\int_0^\infty \left(\hat{F}_n(x) - F(x)\right)^2 \cdot f(x)\mathrm{d}x\right],$$

Table 5 The range of the respective parameters

Estimator	Abbr.	Parameters
EDF	F_n	–
Kernel	$F_{h,n}$	$h = i/1000, i \in [2, 200]$
Szasz	$\hat{F}^S_{m,n}$	$m \in [2, 200]$
Hermite Half	$\hat{F}^H_{N,n}$	$N \in [2, 60]$

where $\hat{F}_n$ can be any of the considered estimators. In the second part, the asymptotic normality of the estimators is illustrated for one distribution.

All of the estimators except for the 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 5. 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.

Note that a selection of m could be based on m_{opt}, defined in Corollary 2, using ideas from automatic bandwidth selection in kernel density estimation. Rule-of-thumb selectors replace the unknown density and distribution function with a reference distribution, for example the exponential distribution in our case. For plug-in selectors, the unknown quantities are estimated using pilot values of m. However, the analysis of such proposals is clearly far beyond the scope of this work.

Every MISE is calculated by a Monte Carlo simulation with $M = 10\,000$ repetitions. To be specific, let

$$\text{ISE}\left[\hat{F}_n\right] = \int_0^\infty \left[\hat{F}_n(x) - F(x)\right]^2 \cdot f(x)\mathrm{d}x,$$

and with M pseudo-random samples, the averaged ISE is calculated by

$$\overline{ISE}\left[\hat{F}_n\right] = \frac{1}{M}\sum_{i=1}^{M} \text{ISE}_i[\hat{F}_n] \simeq MISE\left[\hat{F}_n\right],$$

where ISE_i is the integrated squared error calculated from the ith randomly generated sample. For the Hermite estimator, the standardization explained in (Hanebeck 2020) is used. In this simulation, we do not estimate the mean μ and the standard deviation σ as we already know the true parameters.

5.1 Comparison of the estimators

For comparison, the exponential distribution with parameter $\lambda = 2$ is chosen as well as three different Weibull mixture distributions. The bi- and trimodal mixtures that are considered are:

Weibull 1: $0.5 \cdot Weibull(1,1) + 0.5 \cdot Weibull(4,4)$

Weibull 2: $0.5 \cdot Weibull(3/2,3/2) + 0.5 \cdot Weibull(5,5)$

Weibull 3: $0.35 \cdot Weibull(3/2,3/2) + 0.35 \cdot Weibull(4.5,4.5) + 0.3 \cdot Weibull(8,8)$.

For the exponential distribution, the different sample sizes that are used are $n = 20, 50, 100$, and 500. For the Weibull distributions, only $n = 50$ and $n = 200$ are considered. An example of the different estimators for the exponential distribution can be seen in Fig. 2 for $n = 20$ and $n = 500$. It is obvious that the Hermite estimators do not approach one, which is due to the truncation. Table 6 shows that the Szasz estimator has a best performance.

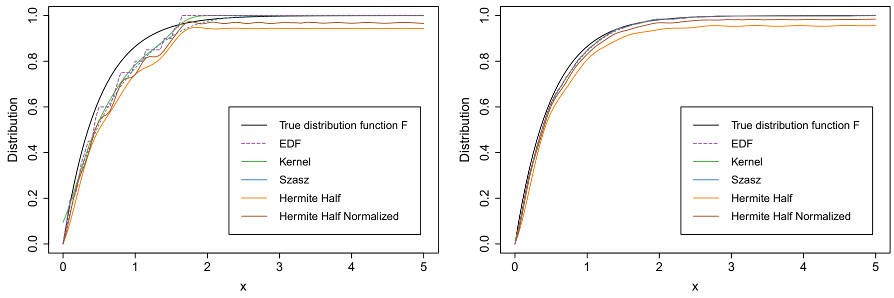

Fig. 2 Plot of the considered estimators for $n = 20$ and $n = 500$

Table 6 The averaged ISE values, multiplied by 10^{-3}

	n	EDF	Kernel	Szasz	Hermite Half	Hermite Norm.
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
Weibull 1	50	3.32	2.92	2.55	3.26	3.45
	200	0.83	0.76	0.71	0.99	1.33
Weibull 2	50	3.32	2.96	2.59	3.08	2.76
	200	0.83	0.75	0.72	0.79	0.79
Weibull 3	50	3.36	3.11	2.55	3.26	2.91
	200	0.83	0.77	0.73	0.81	0.8

5.2 Illustration of the asymptotic normality

The goal here is to illustrate the asymptotic normality

$$\sqrt{n}\left(\hat{F}_n(x) - F(x)\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2(x)\right)$$

of the different estimators, where $\hat{F}_n$ can be any of the estimators. The expression can be rewritten as

$$\hat{F}_n(x) \sim \mathcal{AN}\left(F(x), \frac{\sigma^2(x)}{n}\right).$$

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 Fig. 3, the red line 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.

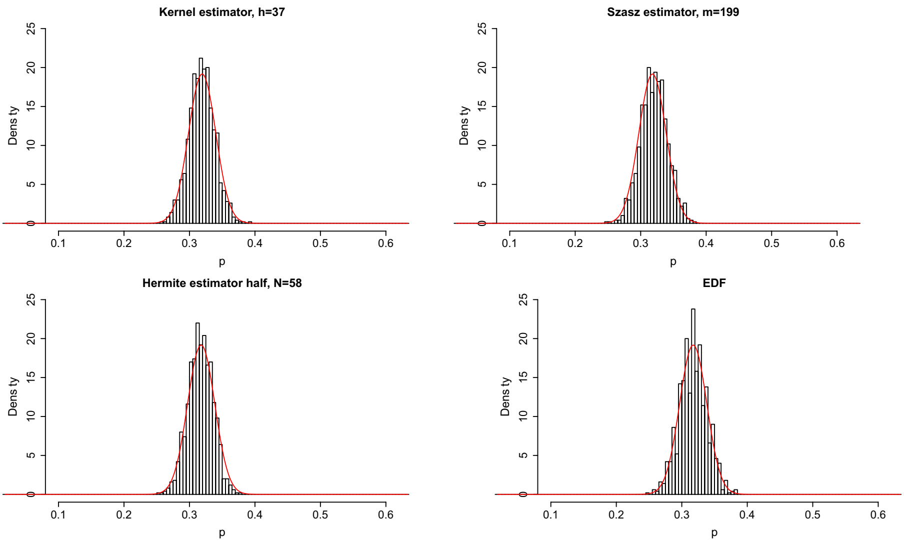

Fig. 3 Illustration of the asymptotic normal distribution

6 Conclusions

In this paper, we have introduced an estimator for distribution functions on $[0, \infty)$ based on Szasz–Mirakyan operators. We have studied asymptotic properties of the Szasz estimator, and conducted the simulations to demonstrate its finite sample performance.

Appendix

The following theorem can be found in Ouimet (2020). He pointed out a mistake in the paper of Leblanc (2012) which also has an impact on this paper. The asymptotic behavior of $R^S_{1,m}$ in Lemma 3 has been corrected compared to Lemma 3 in Hanebeck and Klar 2020, arXiv v.1.

Theorem 8 *We define*

$$V_{k,m}(x) = \frac{(mx)^k}{k!}e^{-mx}, \ \phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \ \textit{and} \ \delta_k = \frac{k - mx}{\sqrt{mx}}.$$

Pick any $\eta \in (0, 1)$. *Then, we have uniformly for* $k \in \mathbb{N}_0$ *with* $\left|\frac{\delta_k}{\sqrt{mx}}\right| \leq \eta$ *that*

$$\frac{V_{k,m}(x)}{\frac{1}{\sqrt{mx}}\phi(\delta_k)} = 1 + m^{-1/2}\frac{1}{\sqrt{x}}\left(\frac{1}{6}\delta_k^3 - \frac{1}{2}\delta_k\right)$$
$$+ m^{-1}\frac{1}{x}\left(\frac{1}{72}\delta_k^6 - \frac{1}{6}\delta_k^4 + \frac{3}{8}\delta_k^2 - \frac{1}{12}\right) + O_{x,\eta}\left(\frac{|1+\delta_k|^9}{m^{3/2}}\right)$$

as $n \to \infty$.

We now present various properties of $V_{k,m}$ that are needed for the proofs. The following lemma and its proof are similar to Lemma 2 and Lemma 3 in Leblanc (2012). As mentioned before, parts (e) and (h) take the suggestions in Ouimet (2020) into account. The proofs for these parts are adjusted accordingly.

Lemma 3 *Define*

$$L_m^S(x) = \sum_{k=0}^{\infty} V_{k,m}^2(x),$$
$$R_{j,m}^S(x) = m^{-j}\sum\sum_{0\le k<l\le\infty}(k-mx)^j V_{k,m}(x)V_{l,m}(x) \text{ for } j \in \{0,1,2\},$$

and

$$\tilde{R}_{1,m}^S(x) = m^{1/2}\sum_{k,l=0}^{\infty}\left(\frac{k\wedge l}{m} - x\right)V_{k,m}(x)V_{l,m}(x),$$

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.*

(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) $\tilde{R}_{1,m}^S(x) = -\sqrt{\frac{x}{\pi}} + o_x(1)$ for $x \in (0,\infty)$ *and* $R_{1,m}^S(x) = m^{-1/2}\left[-\sqrt{\frac{x}{4\pi}} + o_x(1)\right]$,
(f) $m^{1/2}\int_0^{\infty} L_m^S(x)e^{-ax}\mathrm{d}x = \frac{1}{2\sqrt{a}} + o(1)$ *for* $a \in (0,\infty)$,
(g) $m^{1/2}\int_0^{\infty} xL_m^S(x)e^{-ax}\mathrm{d}x = \frac{1}{4a^{3/2}} + o(1)$ *for* $a \in (0,\infty)$,

(h) *For any continuous and bounded function g on* $[0,\infty)$, $m^{1/2}\int_0^\infty g(x)R^S_{1,m}(x)e^{-ax}\mathrm{d}x = -\int_0^\infty g(x)\frac{\sqrt{x}}{\sqrt{4\pi}}e^{-ax}\mathrm{d}x + o(1)$ *for* $a \in (0,\infty)$ *and* $\int_0^\infty g(x)\tilde{R}^S_{1,m}(x)e^{-ax}\mathrm{d}x = -\int_0^\infty g(x)\frac{\sqrt{x}}{\sqrt{\pi}}e^{-ax}\mathrm{d}x + o(1)$.

Acknowledgements The authors are grateful to two reviewers and the editors for their helpful remarks and comments on an earlier version of this manuscript. They are also sincerely grateful to Frédéric Ouimet for pointing out an error in a previous version of Lemma 3, for helpful discussions and for sharing his preprint Ouimet (2020).

References

Altman, N., Léger, C. (1995). Bandwidth selection for Kernel distribution function estimation. *Journal of Statistical Planning and Inference*, *46*(2), 195–214.

Babu, G. J., Canty, A. J., Chaubey, Y. P. (2002). Application of Bernstein polynomials for smooth estimation of a distribution and density function. *Journal of Statistical Planning and Inference*, *105*(2), 377–392.

Bowman, A., Hall, P., Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika*, *85*(4), 799–808.

Duin, R. (1976). On the choice of smoothing parameters for Parzen estimators of probability density functions. *IEEE Transactions on Computers* C-25(11):1175–1179.

Falk, M. (1983). Relative efficiency and deficiency of Kernel type estimators of smooth distribution functions. *Statistica Neerlandica*, *37*(2), 73–83.

Gramacki, A. (2018). *Nonparametric Kernel density estimation and its computational aspects studies in big data*. New York: Springer International Publishing.

Hanebeck, A. (2020). Nonparametric distribution function estimation. Master's thesis, Karlsruher Institut für Technologie (KIT).

Hanebeck, A., Klar, B. (2020). Smooth distribution function estimation for lifetime distributions using Szasz-Mirakyan operators. arXiv:200509994 [math, stat] 2005.09994.

Helali, S., Slaoui, Y. (2020). Estimation of a distribution function using Lagrange polynomials with Tchebychev-Gauss points. *Statistics and Its Interface*, *13*(3), 399–410.

Hogg, R. V., Klugman, S. A. (1984). *Loss distributions*. New York: Wiley-Interscience.

Jmaei, A., Slaoui, Y., Dellagi, W. (2017). Recursive distribution estimator defined by stochastic approximation method using bernstein polynomials. *Journal of Nonparametric Statistics*, *29*, 792–805.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). *Continuous univariate distributions* 2nd ed., Vol. 1. New York: Wiley-Interscience.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1995). *Continuous univariate distributions* 2nd ed., Vol. 2. New York: Wiley-Interscience.

Kim, C., Kim, S., Park, M., Lee, H. (2006). A bias reducing technique in kernel distribution function estimation. *Computational Statistics*, *21*(3), 589–601.

Leblanc, A. (2012). On estimating distribution functions using bernstein polynomials. *Annals of the Institute of Statistical Mathematics*, *64*(5), 919–943.

Lockhart, R. (2013). The basics of nonparametric models. http://people.stat.sfu.ca/~lockhart/richard/830/13_3/lectures/nonparametric_basics/.

Lorentz, G. G. (1986). *Bernstein polynomials* 2nd ed. Co, New York, N.Y.: Chelsea Pub.

Marshall, A. W., Olkin, I. (2007). *Life distributions*: *Structure of nonparametric, semiparametric, and parametric families*. Springer Series in Statistics, New York: Springer.

Mokkadem, A., Pelletier, M., Slaoui, Y. (2009). The Stochastic approximation method for the estimation of a multivariate probability density. arXiv:08072960 [math, stat] 0807.2960.

Ouimet, F. (2020). A local limit theorem for the Poisson distribution and its application to the Le Cam distance between Poisson and Gaussian experiments and asymptotic properties of szasz estimators. arXiv:201005146 [math, stat] 2010.05146.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, *33*(3), 1065–1076.

Polansky, A. M., Baker, E. R. (2000). Multistage plug- in bandwidth selection for Kernel distribution function estimates. *Journal of Statistical Computation and Simulation*, *65*(1–4), 63–80.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, *27*(3), 832–837.

Rudemo, M. (1982). Empirical choice of histograms and Kernel density estimators. *Scandinavian Journal of Statistics*, *9*(2), 65–78.

Schwartz, S. C. (1967). Estimation of probability density by an orthogonal series. *The Annals of Mathematical Statistics*, *38*(4), 1261–1265.

Slaoui, Y. (2014). Bandwidth selection for recursive Kernel density estimators defined by stochastic approximation method. https://www.hindawi.com/journals/jps/2014/739640/.

Stephanou, M., Varughese, M., Macdonald, I. (2017). Sequential quantiles via Hermite series density estimation. *Electronic Journal of Statistics*, *11*(1), 570–607.

Szasz, O. (1950). Generalization of S. Bernstein's polynomials to the infinite interval. *J ournal of Research of the National Bureau of Standards*, *45*, 239–245.

Tenreiro, C. (2006). Asymptotic behaviour of multistage plug-in bandwidth selections for Kernel distribution function estimators. *Journal of Nonparametric Statistics*, *18*(1), 101–116.

Watson, G. S., Leadbetter, M. R. (1964). Hazard analysis ii. *Sankhya The Indian Journal of Statistics Series A (1961–2002)*, *26*(1), 101–116.

Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. *Bulletin of Mathematical Statistics*, *15*(3), 69–78.

Zhang, S., Li, Z., Zhang, Z. (2020). Estimating a distribution function at the boundary. *Austrian Journal of Statistics*, *49*(1), 1–23.