Nonparametric estimation of unrestricted distributions and their jumps

Kairat Mynbaev

New Economic School
Satbayev University
Almaty 050013, Kazakhstan
and
Institute of Mathematics and Mathematical Modeling
Almaty 050000, Kazakhstan
email: kairat_mynbayev@yahoo.com

Carlos Martins-Filho

Department of Economics
University of Colorado
Boulder, CO 80309-0256, U.S.A.
email: carlos.martins@colorado.edu

Daniel J. Henderson

Department of Economics, Finance and Legal Studies
University of Alabama
Tuscaloosa, AL 35487-0224, U.S.A.
email: djhender@cba.ua.edu

and

School of Mathematical Sciences
Nankai University
Tianjin, 300071, China

May, 2021

Abstract. We consider nonparametric estimation of a distribution function F associated with a random variable X based on a random sample. First, for x a point of continuity of F, we define a class of estimators for $F(x)$ and obtain their rates of convergence. Contrary to the existing literature, we impose no restriction on the existence or smoothness of the derivatives of F. The traditional kernel estimator for $F(x)$ is a member of the class. Second, for x that is either the location of a jump discontinuity or an isolated point of the support, we define a class of estimators for the jump $p_x = F(x) - F(x_-)$ and obtain their rates of convergence. Again, no additional restriction is imposed on F beyond right-continuity. Our results are of significant practical use as there are numerous examples in Economics, Finance and Biomedicine of distributions that have point masses. Our main insight is also applied to obtain new inversion theorems for characteristic functions and explicit estimates for convergence rates. A small simulation study provides some evidence on the finite sample properties of our proposed estimators and contrasts their performance with some existing alternatives. An empirical section illustrates the use of our estimators using data on global elevation and data associated with "P-hacking" in economics journals.

Key words and phrases. Discontinuous distribution functions; estimation of distribution function jumps; inversion theorems; nonparametric distribution estimation.

MSC 2020: Primary 62G05; secondary 62G20.

1 Introduction

A classic problem in statistics is the pointwise estimation of a distribution function F associated with a random variable X using a random sample of observations $\{X_i\}_{i=1}^n$ with $n \in \mathbb{N}$. For any $x \in \mathbb{R}$, a natural estimator for $F(x)$ is the empirical distribution $F_n(x)$, but the fact that $F_n(x)$ is a step function has been the motivation to consider smooth, nonparametric estimators for F when it is assumed to be absolutely continuous (Nadaraya, 1964; Watson and Leadbetter, 1964; Azzalini, 1981; Falk, 1983; Swanepoel and Van Graan, 2005). A much smaller set of papers has studied nonparametric estimators where F is assumed to be merely continuous or even discontinuous at x. Murthy (1965) proposes consistent and asymptotically normal estimators for $F(x)$ when x is a point of continuity of F and for the jump $p_x = F(x) - \lim_{h \downarrow 0} F(x-h)$ when x is a point of discontinuity. Winter (1973) and Yamato (1973) establish the almost sure, uniform convergence of a class of nonparametric estimators for cases where F is continuous but a density does not exist. However, none of these papers provide rates of convergence of the estimators under study and, surprisingly, no rates of convergence are available in the existing literature. Whenever rates of convergence are available, such as in Azzalini (1981), Shirahata and Chu (1992), Altman and Léger (1995), Bowman et al. (1998), Li and Racine (2007, p. 21, Theorem 2.1), Tenreiro (2013), and Cheng (2017), F is assumed to be at least differentiable at x.

In this paper we define general classes of nonparametric estimators for $F(x)$ when x is a point of continuity of F and for p_x when x is a point of discontinuity of F. For estimators in both classes, we provide rates at which their biases and variances decay to zero as $n \to \infty$ without imposing any smoothness restrictions on F. These results are new and fill a surprising gap in the literature on the nonparametric estimation of distribution functions and their jumps. As a direct consequence of our main results, we obtain sufficient conditions for asymptotic normality of estimators in these classes.

The method used in proving our main results relies on the Stieltjes integral and, as a byproduct, we are able to obtain new inversion theorems for characteristic functions together with *convergence rates*, generalizing Lévy's theorem (Lukacs, 1970, Thm. 3.2.1) and Borovkov's theorem (Borovkov, 2009, Thm 7.2.2).

The rest of this paper is organized as follows. In Section 2 we consider the estimation of $F(x)$ at continuity points and in Section 3 we consider the estimation of p_x. The lemmas used in Sections 2 and 3 to prove our main theorems on estimation are applied in Section 4 to obtain new inversion theorems for characteristic functions. Section 5 contains simulation results and Section 6 provides simple examples of empirical uses for our estimators. Section 7 gives a brief conclusion. All proofs are provided in the appendix. In what follows, since F is a distribution function, we take F to be right-continuous. We let L_1 denote the space of functions f on $\mathbb{R}$ with a finite norm $\|f\|_{L_1} = \int_{\mathbb{R}} |f(t)|dt$; C the space of uniformly bounded, continuous functions on $\mathbb{R}$ with norm $\|f\|_C = \sup_{t\in\mathbb{R}} |f(t)|$; χ_S the indicator function of a set S; and $E(X)$ the expectation and $V(X)$ the variance of a random variable X.

2 Estimation at points of continuity of F

We start by defining a general class of estimators for $F(x)$ when x is a point of continuity of F. Given a function U and a random sample $\{X_i\}_{i=1}^n$ from F we consider estimators defined as

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} U\left(\frac{X_i - x}{h}\right) \tag{1}$$

where $h = h_n > 0$ is a bandwidth, or smoothing parameter, satisfying $h \to 0$ as $n \to \infty$. In addition, we place the following restriction on U.

Assumption 1. *Let $U \in C$, $\lim_{x\to-\infty} U(x) = 1$ and $\lim_{x\to\infty} U(x) = 0$.*

If $K : \mathbb{R} \to \mathbb{R}$ is integrable with $\int_{-\infty}^{\infty} K(t)dt = 1$ and we set $U(x) = \int_x^{\infty} K(t)dt$, then it follows that U satisfies Assumption 1. In addition, if K is an even function, then

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} L\left(\frac{x - X_i}{h}\right), \tag{2}$$

where $L(x) = \int_{-\infty}^{x} K(t)dt$. This is the smoothed kernel distribution estimator, first proposed by Nadaraya (1964) and extensively studied in the literature. See, more recently, Berg and Politis (2009), Giné and Nickl (2009), Chacón and Rodríguez-Casal (2010), and Cheng (2017).

Letting $\phi_U(N) = \max\left\{\sup_{z<-N} |U(z) - 1|, \sup_{z>N} |U(z)|\right\}$ for $N > 0$, Assumption 1 implies that

$\phi_U(N) \to 0$ as $N \to \infty$. Also, letting $\omega(x,\varepsilon) = \sup_{0<h\leq\varepsilon}(F(x+h) - F(x-h))$, if x is a point of continuity of F, then $\omega(x,\varepsilon) \to 0$ as $\varepsilon \to 0$.

Lemma 1. *If U satisfies Assumption 1 and $\delta \in (0,1)$, then for all $h > 0$*

$$\left|\int_{\mathbb{R}} U\left(\frac{y-x}{h}\right) dF(y) - F(x)\right| \leq \phi_U(h^{\delta-1}) + \omega(x, h^\delta)(1 + \|U\|_C).$$

The standard procedure used in the existing literature (Azzalini, 1981; Altman and Léger, 1995; Bowman et al., 1998; Li and Racine, 2007; Cheng, 2017) is to obtain $\int_{\mathbb{R}} U\left(\frac{y-x}{h}\right) dF(y)$ using integration by parts and then try to obtain convergence rates by imposing conditions on the derivatives of F. Using Stieltjes integration, we are able to bypass the imposing additional smoothing conditions on F. Lemma 1 is the basis for the following theorem.

Theorem 1. *Suppose that U satisfies Assumption 1 and let x be a continuity point of F. Then*

(i) $\left|E(\hat{F}(x)) - F(x)\right| \leq \phi_U(h^{\delta-1}) + \omega(x, h^\delta)(1 + \|U\|_C)$ *and*

(ii) if $F(x) \in (0,1)$ *then* $nV(\hat{F}(x)) = F(x)(1 - F(x)) + o(1)$.

We note that Murthy (1965) and Yamato (1973) established asymptotic unbiasedness for the estimator $\hat{F}(x)$ in Equation (2) when x is a point of continuity of F, but provided no rates of convergence, as we do in part (i) of Theorem 1.[1] If F is assumed to be continuous at every $x \in \mathbb{R}$, then the following corollaries give a uniform convergence rate for the bias and almost sure uniform convergence of $\hat{F}$ to F.

Corollary 1. *If U satisfies Assumption 1 and $F \in C$, then*

$$\sup_{x\in\mathbb{R}} \omega(x,\varepsilon) \to 0 \text{ as } \varepsilon \to 0 \tag{3}$$

and $\sup_{x\in\mathbb{R}}\left|E(\hat{F}(x)) - F(x)\right| \leq \phi_U(h^{\delta-1}) + \sup_{x\in\mathbb{R}} \omega(x, h^\delta)(1 + \|U\|_C) \to 0$ *as* $h \to 0$.

Corollary 2. *If U satisfies Assumption 1 and $F \in C$, then* $\sup_{x\in\mathbb{R}}\left|\hat{F}(x) - F(x)\right| \to 0$ *almost surely.*

---

[1] Murthy (1965) focuses on the estimation of the reliability function $R(x) = 1 - F(x)$ using $\hat{R}(x) = \frac{1}{n}\sum_{i=1}^n L\left(\frac{X_i - x}{h}\right)$, but his results apply directly to $\hat{F}(x)$ in Equation (2). Yamato (1973) establishes asymptotic unbiasedness for a class of estimators defined by $\tilde{F}(x) = \frac{1}{n}\sum_{i=1}^n W_n(x - X_i)$, where $W_n(x) \to \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases}$. Letting $W_n(x) = L\left(\frac{x}{h}\right)$ shows that $\hat{F}(x)$ in Equation (2) is in his class.

The following theorem gives the asymptotic distribution of $\hat{F}(x)$ under suitable normalization and centering.

Theorem 2. *Suppose that U satisfies Assumption 1 and that x is a point of continuity of F such that $F(x) \in (0,1)$. For $\delta \in (0,1)$, if $h \to 0$, $\sqrt{n}\,\phi_U(h^{\delta-1}) \to 0$, and $\sqrt{n}\,\omega(x,h^\delta) \to 0$ as $n \to \infty$, then*

$$\sqrt{n}(\hat{F}(x) - F(x)) \xrightarrow{d} N(0, (1-F(x))F(x)).$$

Remark 1. The requirements that $\sqrt{n}\,\phi_U(h^{\delta-1}) \to 0$ and $\sqrt{n}\,\omega(x,h^\delta) \to 0$ can always be satisfied by an appropriate choice of the sequence $\{h_n\}$. If $U(x) = \int_x^\infty K(t)dt$ and K has a compact support (as we choose in our simulations), then $\sqrt{n}\,\phi_U(h^{\delta-1}) = 0$ for any $\delta \in (0,1)$ and all $h \in (0, h(\delta))$. If K has a compact support, the choice of the bandwidth is dictated solely by the condition that $\sqrt{n}\,\omega(x,h^\delta) \to 0$. If, in addition to continuity at x, F satisfies a local Lipschitz condition of order $\alpha > 0$, then for some positive constant M_x and $u \in (0, h^\delta]$, $F(x+u) - F(x-u) < 2M_x u^\alpha \le 2M_x h^{\alpha\delta}$ and $\sqrt{n}\,\omega(x,h^\delta) \le 2M_x\sqrt{n}\,h^{\alpha\delta} \to 0$ if $nh^{2\alpha\delta} = o(1)$. If, as usual, $h \propto n^{-\gamma}$ for $\gamma > 0$, then $nh^{2\alpha\delta} = o(1)$ if $\gamma > 1/(2\alpha\delta)$.

Now, we turn to the estimation of the probability $F(x,y) \equiv F(y) - F(x)$ for $x < y$. From Theorem 1 we can derive the rate of convergence of $\hat{F}(y) - \hat{F}(x)$ to $F(x,y)$. However, the variance of $\hat{F}(y) - \hat{F}(x)$ cannot be easily obtained from Theorem 1. To do so, we make the following assumption.

Assumption 2. *Let $G = G([a,b])$ be a bounded continuous function of intervals $[a,b] \subset \mathbb{R}$ or, more precisely, of an ordered pair (a,b), where $a < b$, with $\lim_{b \to -\infty} G([a,b]) = 0$, $\lim_{a \to -\infty,\ b \to \infty} G([a,b]) = 1$, and $\lim_{a \to \infty} G([a,b]) = 0$.*

Define $\phi_G(N) = \max\left\{\sup_{b<-N} |G([a,b])|, \sup_{a>N} |G([a,b])|, \sup_{a<-N,\ b>N} |G([a,b]) - 1|\right\}$ and note that Assumption 2 implies that $\phi_G(N) \to 0$ as $N \to \infty$. If a kernel K satisfies $\int_{\mathbb{R}} K(t)dt = 1$, then $G([a,b]) = \int_a^b K(t)dt$ satisfies Assumption 2. Additionally, if U_1 and U_2 satisfy Assumption 1, then $G([a,b]) = U_1(a) - U_2(b)$, with $-\infty < a < b < \infty$, satisfies Assumption 2. We estimate $F(x,y)$ by

$$\hat{F}(x,y) = \frac{1}{n}\sum_{i=1}^{n} G\left(\left[\frac{X_i - y}{h}, \frac{X_i - x}{h}\right]\right).$$

Lemma 2. *If G satisfies Assumption 2 and $\delta \in (0,1)$, then for all $h > 0$,*

$$\left| \int_{\mathbb{R}} G\left(\left[\frac{z-y}{h}, \frac{z-x}{h}\right]\right) dF(z) - F(x,y) \right| \leq \phi_G(h^{\delta-1}) + (1 + \|G\|_C) \left[\omega(x,h^\delta) + \omega(y,h^\delta)\right].$$

Theorem 3. *Suppose that G satisfies Assumption 2, $x < y$ are continuity points of F, and $\delta \in (0,1)$. Then*

(i) $\left| E\hat{F}(x,y) - F(x,y) \right| \leq \phi_G(h^{\delta-1}) + (1 + \|G\|_C) \left[\omega(x,h^\delta) + \omega(y,h^\delta)\right]$ *for $h > 0$ and*

(ii) if $F(x,y) \in (0,1)$, then $V(\hat{F}(x,y)) = \frac{1}{n}\left[F(x,y)(1 - F(x,y)) + o(1)\right]$.

The following corollary on the uniform rate of convergence is promptly obtained.

Corollary 3. *If G satisfies Assumption 2 and $F \in C$, then*

$$\sup_{x,y\in\mathbb{R}} \left| E\hat{F}(x,y) - F(x,y) \right| \leq \phi_G(h^{\delta-1}) + 2(1 + \|G\|_C) \sup_{x\in\mathbb{R}} \omega(x,h^\delta) \to 0 \text{ as } h \to 0.$$

3 Estimation of jumps of a distribution function

Let the jump of F at a point x be given by $p_x = F(x) - F(x_-)$, where $F(x_-) \equiv \lim_{\varepsilon\downarrow 0} F(x-\varepsilon)$. We estimate p_x by

$$\hat{p}_x = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{X_i - x}{h}\right), \tag{4}$$

where the function W satisfies the following assumption.

Assumption 3. *The function W is bounded and continuous and satisfies $W(0) = 1$ and $\lim_{|x|\to\infty} W(x) = 0$.*

We let $\omega_W(\varepsilon) = \sup_{|x|\leq\varepsilon} |W(x) - 1|$, $\phi_W(N) = \sup_{|x|\geq N} |W(x)|$, and $\delta_F(\varepsilon) = \int_{|z-x|<\varepsilon} dF(z) - p_x \geq 0$. Assumption 3 implies that $\omega_W(\varepsilon) \to 0$ as $\varepsilon \to 0$ and $\phi_W(N) \to 0$ as $N \to \infty$. Since $\bigcap_{h>0}(x-h, x+h) = \{x\}$, by the continuity of probability, $\lim_{\varepsilon\to 0} \delta_F(\varepsilon) = 0$.

Denote by $(\mathcal{F}f)$ the exponential Fourier transform of f. Thus, for a distribution function F, $(\mathcal{F}F)(t) = \int_{\mathbb{R}} e^{itz} dF(z)$ and, for the kernel K, $(\mathcal{F}K)(t) = \int_{\mathbb{R}} e^{itz} K(z)dz$. If K is a kernel, its exponential Fourier transform $(\mathcal{F}K)$ satisfies Assumption 3.

Lemma 3. *(i) Let $\varepsilon_1 \in (0,1)$ and $\varepsilon_2 > \varepsilon_1$. For any $\lambda \in (0,1)$,*

$$\left| \int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x \right| \leq \omega_W(\lambda^{\varepsilon_2-\varepsilon_1})\left[p_x + \delta_F(\lambda^{1-\varepsilon_1})\right] + (1 + \|W\|_C)\, \delta_F(\lambda^{1-\varepsilon_1}) + \phi_W(\lambda^{-\varepsilon_1}).$$

(ii) If x is an isolated point of the support of F, that is, $\int_{|z-x|<h} dF(z) = p_x$ for all small $h > 0$, then for all small $\lambda > 0$,

$$\left| \int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x \right| \leq \phi_W(\lambda^{-\varepsilon_1}).$$

In particular, when W has a compact support, $\int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) = p_x$ for all small λ.

Remark 2. Theorem 3.2.3 of Lukacs (1970) shows that the jumps of F can be found from the characteristic function of F according to

$$p_x = \lim_{\lambda \to 0} \int_{\mathbb{R}} e^{-itx} (\mathcal{F}F)(t) \lambda K(\lambda t) dt$$

for $K = \chi_{[-1/2,1/2]}$. Mynbaev (2012) extended this result to symmetric $K \in L_1$. Neither work provided an estimate of the rate of convergence. Noting that

$$\int_{\mathbb{R}} e^{-itx} (\mathcal{F}F)(t) \lambda K(\lambda t) dt = \int_{\mathbb{R}} (\mathcal{F}K)\left(\frac{z-x}{\lambda}\right) dF(z), \tag{5}$$

it follows immediately that Lemma 3 improves Mynbaev's result by lifting the symmetry condition and providing a rate of convergence.

Theorem 4. *Suppose that W satisfies Assumption 3.*

(i) Let $\varepsilon_1 \in (0,1)$ and $\varepsilon_2 > \varepsilon_1$. Then for any $h \in (0,1)$,

$$|E(\hat{p}_x) - p_x| \leq \omega_W(h^{\varepsilon_2 - \varepsilon_1}) \left[p_x + \delta_F(h^{1-\varepsilon_1})\right] + (1 + \|W\|_C)\, \delta_F(h^{1-\varepsilon_1}) + \phi_W(h^{-\varepsilon_1}).$$

(ii) If x is an isolated point of the support of F, then for all small h,

$$|E(\hat{p}_x) - p_x| \leq \phi_W(h^{-\varepsilon_1}).$$

In particular, when W has a compact support, $\hat{p}_x$ is unbiased for all small h.

(iii) If $p_x \in (0,1)$ then $V(\hat{p}_x) = \frac{1}{n}\left[p_x(1-p_x) + o(1)\right]$ as $h \to 0$.

Remark 3. Now suppose that the moments $\int_{\mathbb{R}} t^j K(t) dt$ are zero for $j = 1, ..., m-1$ and the derivative $(\mathcal{F}K)^{(m)}$ is bounded. Then, since $(\mathcal{F}K)^{(j)}(0) = 0$ for $j = 1, ..., m-1$, by Taylor's theorem,

$$(\mathcal{F}K)\left(\frac{z-x}{\lambda}\right) - (\mathcal{F}K)(0) \quad = \quad (\mathcal{F}K)^{(m)}(\theta)\left(\frac{z-x}{\lambda}\right)^m \frac{1}{m!}.$$

Using this equation, taking $W = (\mathcal{F}K)$ and choosing h and h_1 as in the proof of Lemma 3,

$$\begin{aligned}\left|\int_{|z-x|<h_1} (\mathcal{F}K)\left(\frac{z-x}{\lambda}\right) dF(z) - p_x\right| &\le \frac{\sup\left|(\mathcal{F}K)^{(m)}\right|}{m!}\left(\frac{h_1}{\lambda}\right)^m \left|\int_{|z-x|<h_1} dF(z)\right| \\ &+ \left|\int_{|z-x|<h_1} dF(z) - p_x\right| \\ &\le c\lambda^{m(\varepsilon_2-\varepsilon_1)}\left[p_x + \delta_F(\lambda^{1-\varepsilon_1})\right] + \delta_F(\lambda^{1-\varepsilon_1}).\end{aligned}$$

The resulting bound,

$$|E(\hat{p}_x) - p_x| \le c\lambda^{m(\varepsilon_2-\varepsilon_1)}\left[p_x + \delta_F(\lambda^{1-\varepsilon_1})\right] + \left(1 + \|K\|_{L_1}\right)\delta_F(\lambda^{1-\varepsilon_1}) + \sup_{|z|\ge\lambda^{-\varepsilon_1}} |(\mathcal{F}K)(z)|,$$

shows that imposing higher-order kernel conditions, without restricting the class of distribution functions, has a limited effect on the convergence rate because of the term $\delta_F(\lambda^{1-\varepsilon_1})$. The following theorem establishes the asymptotic normality of $\hat{p}_x$ under suitably centering and normalization.

Theorem 5. *Suppose that W satisfies Assumption 3 and that F has a jump p_x at x, where $p_x \in (0,1)$. For $\varepsilon_1 \in (0,1)$ and $\varepsilon_2 > \varepsilon_1$, if $h \to 0$, $\sqrt{n}\,\phi_W(h^{-\varepsilon_1}) \to 0$, $\sqrt{n}\,\omega_W(h^{\varepsilon_2-\varepsilon_1}) \to 0$, and $\sqrt{n}\delta_F(h^{1-\varepsilon_1}) \to 0$ as $n \to \infty$, then*

$$\sqrt{n}(\hat{p}_x - p_x) \xrightarrow{d} N(0, (1-p_x)p_x).$$

Remark 4. The requirements that $\sqrt{n}\,\phi_W(h^{-\varepsilon_1}) \to 0$ and $\sqrt{n}\,\omega_W(h^{\varepsilon_2-\varepsilon_1}) \to 0$ as $n \to \infty$ always hold with appropriately chosen $\{h_n\}$. If $W(t) = (\mathcal{F}K)(t)$ such that W has a compact support (e.g., the W and K used in Subsection 5.2), then $\sqrt{n}\,\phi_W(h^{-\varepsilon_1}) = 0$ for all small h. The condition that $\sqrt{n}\delta_F(h^{1-\varepsilon_1}) \to 0$ can always be satisfied if h_n tends to 0 sufficiently quickly. However, to obtain a tractable dependence of the bandwidth on n, note that

$$\delta_F(h^{1-\varepsilon_1}) = F(x + h^{1-\varepsilon_1}) - \left(F(x - h^{1-\varepsilon_1}) + p_x\right).$$

If a Lipschitz condition of order α is imposed on the difference on the right hand side of the equality, then $\sqrt{n}\delta_F(h^{1-\varepsilon_1}) < 2M_x\sqrt{n}h^{(1-\varepsilon_1)\alpha}$. As in Remark 1, if $h \propto n^{-\gamma}$, it suffices to choose $\gamma > 1/(2(1-\varepsilon_1)\alpha)$ to satisfy the condition.

Theorem 6. *If K is a kernel and F_1 is purely discrete with a jump p_j at ξ_j, then $\sum_j p_j^2$ can be recovered from the characteristic function of F_1, as $\sum_j p_j^2 = \lim_{\lambda \to 0} \int_{\mathbb{R}} |(\mathcal{F}F_1)(t)|^2 \lambda K(\lambda t) dt$.*

This result improves upon Theorem 3.3.4 of Lukacs (1970) and Mynbaev (2012), where the rate of convergence was not established and the assumptions on K are more restrictive. Both works miss the requirement that F_1 should be purely discrete.

4 Inversion theorems

Let the Fourier transform of the distribution F be written as $\psi(s) = (\mathcal{F}F)(s) = \int_{\mathbb{R}} e^{ist} dF(t)$. When ψ is integrable, we denote by $\mathcal{F}^{-1}$ the inverse Fourier transform and write $(\mathcal{F}^{-1}\psi)(t) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ist} \psi(s) ds$. In this case, F has a bounded density f and $\mathcal{F}^{-1}\psi = f$ (Shiryaev, 1996, p. 283). When ψ is not integrable, there are three tasks of interest: (1) recovering $F(x)$ when x is a continuity point of F; (2) recovering $F(x, y) = F(y) - F(x)$, where $x < y$ are both points of continuity; and (3) recovering the jumps of F. The derivation in Gil-Pelaez (1951) accomplishes the first task and gives

$$F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \frac{e^{itx}\psi(-t) - e^{-itx}\psi(t)}{it} dt.$$

For the other two tasks, we provide new inversion theorems whose main advantages are explicit convergence rates. These results, through Lemmas 2 and 3, reveal the common structure of inversion theorems and their link with distribution function estimation. Furthermore, our proofs do not rely on probabilistic methods, so our results hold for any function F of bounded variation.

The heuristics for task (2) are as follows. First,

$$F(y) - F(x) = \int_{\mathbb{R}} \chi_{[x,y]}(t) dF(t). \tag{6}$$

Letting $g_{x,y}(t) = \chi_{[x,y]}(-t)$, we write

$$U(s) \equiv (g_{x,y} * F)(s) = \int g_{x,y}(s - t) dF(t),$$

the convolution of $g_{x,y}$ and F. Then, Equation (6) becomes

$$F(y) - F(x) = \int_{\mathbb{R}} g_{x,y}(-t) dF(t) = (g_{x,y} * F)(0) = U(0). \tag{7}$$

Since $(y-x)^{-1}U(s)$ is a density (Borovkov, 2009, p. 62), U is integrable. The Fourier transform is known to map a convolution to a product as

$$(\mathcal{F}U)(t) = (\mathcal{F}g_{x,y})(t)\psi(t). \tag{8}$$

If the product on the right-hand side of Equation (8) is integrable, then by the Fourier inversion theorem, Equations (7) and (8) yield that $F(y) - F(x) = (\mathcal{F}^{-1}\mathcal{F}U)(0) = \mathcal{F}^{-1}\left[(\ \mathcal{F}g_{x,y})\psi\right](0)$. When the right side of Equation (8) is not integrable, we regularize it by multiplying by a "cutoff" function $H(h\cdot)$, where $h > 0$ is a parameter and $\cdot$ is a placeholder for its argument. The function H is chosen to satisfy

$$F(y) - F(x) = \lim_{h\to 0} \mathcal{F}^{-1}\left[H(h\cdot)(\mathcal{F}g_{x,y})\psi\right](0). \tag{9}$$

Theorem 7. *Let x and y be points of continuity of F. (i) Suppose that H is integrable with $(\mathcal{F}H)(v) = \int_{\mathbb{R}} e^{isv}H(s)ds$ and such that the function*

$$G([a,b]) = \frac{1}{2\pi}\int_a^b (\mathcal{F}H)(v)dv$$

satisfies Assumption 2. Then for $\delta \in (0,1)$ and all $h > 0$,

$$\left|\frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{-ixt} - e^{-iyt}}{it} H(ht)\psi(t)dt - F(x,y)\right| \leq \phi_G(h^{\delta-1}) + (1 + \|G\|_C)\left[\omega(x,h^\delta) + \omega(y,h^\delta)\right].$$

(ii) In the case where $H = \chi_{[-1,1]}$, Assumption 2 is satisfied and Lévy's theorem (Lukacs, 1970, Thm. 3.2.1),

$$F(x,y) = \lim_{h\to 0}\frac{1}{2\pi}\int_{-1/h}^{1/h} \frac{e^{-ixt} - e^{-iyt}}{it}\psi(t)dt,$$

follows with an estimate of the convergence rate.

(iii) In the case where $H(t) = e^{-t^2}$, Assumption 2 is satisfied and Borovkov's theorem (Borovkov, 2009, Thm 7.2.2),

$$F(x,y) = \lim_{h\to 0}\frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{-ixt} - e^{-iyt}}{it} e^{-h^2t^2}\psi(t)dt,$$

follows with an estimate of the convergence.

The following theorem addresses task (3), where interest is in recovering a jump p_x of F at x.

Theorem 8. *Let K be a kernel and set $W = \mathcal{F}K$. (i) Let $\varepsilon_1 \in (0,1)$ and $\varepsilon_2 > \varepsilon_1$. For any $\lambda \in (0,1)$,*

$$\begin{aligned}\left|\int_{\mathbb{R}} e^{-itx}\psi(t)\lambda K(\lambda t)dt - p_x\right| &\leq \omega_W(\lambda^{\varepsilon_2-\varepsilon_1})\left[p_x + \delta_F(\lambda^{1-\varepsilon_1})\right] \\ &\quad + (1+\|W\|_C)\,\delta_F(\lambda^{1-\varepsilon_1}) + \phi_W(\lambda^{-\varepsilon_1}).\end{aligned}$$

(ii) If x is an isolated point of the support of F, then for all small $\lambda > 0$

$$\left|\int_{\mathbb{R}} e^{-itx}\psi(t)\lambda K(\lambda t)dt - p_x\right| \leq \phi_W(\lambda^{-\varepsilon_1}).$$

In particular, if K is an entire function of exponential type, then W has a compact support by the Paley–Wiener theorem (Iosida, 1980, p.162) and $\int_{\mathbb{R}} e^{-itx}\psi(t)\lambda K(\lambda t)dt = p_x$ for all small λ.

Next, we briefly describe modifications required for the case of functions of bounded variation.[2] Let a function F be defined and finite on the interval $[a,b]$ with $-\infty < a < b < \infty$. The total variation of F on $[a,b]$ is defined as

$$V_a^b(f) = \sup \sum_{i=1}^{n} |F(x_i) - F(x_{i-1})|$$

where the supremum is taken over all partitions $x_0 = a < x_1 < \cdots < x_n = b$. If F is defined on the real line and $\sup_{a<b} V_a^b(F) < \infty$, then F is said to be of finite variation on $\mathbb{R}$ and the number $V_{-\infty}^{\infty}(F) = \sup_{a<b} V_a^b(F)$ is called the total variation of F. Using Theorem 1 of Natanson (1955) where appropriate, it is easy to obtain the following generalization of our basic Lemmas 1, 2, and 3.

Lemma 4. *(i) If U satisfies Assumption 1 and $\delta \in (0,1)$, then for all $h > 0$*

$$\left|\int_{\mathbb{R}} U\left(\frac{y-x}{h}\right) dF(y) - F(x)\right| \leq \phi_U(h^{\delta-1})V_{-\infty}^{\infty}(F) + V_{x-h^\delta}^{x+h^\delta}(F)(1+\|U\|_C).$$

(ii) If G satisfies Assumption 2 and $\delta \in (0,1)$, then for all $h > 0$,

$$\left|\int_{\mathbb{R}} G\left(\left[\frac{z-y}{h}, \frac{z-x}{h}\right]\right) dF(z) - F(x,y)\right| \leq \phi_G(h^{\delta-1})V_{-\infty}^{\infty}(F) + (1+\|G\|_C)\left[V_{x-h^\delta}^{x+h^\delta}(F) + V_{y-h^\delta}^{y+h^\delta}(F)\right].$$

(iii) Let $\varepsilon_1 \in (0,1)$ and $\varepsilon_2 > \varepsilon_1$. For any $\lambda \in (0,1)$

$$\left|\int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x\right| \leq \omega_W(\lambda^{\varepsilon_2-\varepsilon_1})V_{x-\lambda^{1-\varepsilon_1}}^{x+\lambda^{1-\varepsilon_1}}(F) + (1+\|W\|_C)\left[V_{x-\lambda^{1-\varepsilon_1}}^{x+\lambda^{1-\varepsilon_1}}(F) - p_x\right] + \phi_W(\lambda^{-\varepsilon_1})V_{-\infty}^{\infty}(F).$$

[2]See chapter 8 in Natanson (1955), especially the appendix.

(iv) If x is an isolated point of the support of F, that is, $\int_{|z-x|<h} dF(z) = p_x$ for all small $h > 0$, then for all small $\lambda > 0$,

$$\left| \int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x \right| \leq \phi_W(\lambda^{-\varepsilon_1}) V_{-\infty}^{\infty}(F).$$

In particular, when W has a compact support, $\int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) = p_x$ for all small λ.

One application of this result is to $F(x) = \int_{-\infty}^{x} f(t)dt$ where $f \in L_1$. This F is absolutely continuous and of bounded variation. An analogue of Theorem 7 based on Lemma 4 shows that $F(x,y) = \int_x^y f(t)dt$ can be recovered from ψ and that f can be found as

$$f(y) = \frac{d}{dy} \lim_{h \to 0} \mathcal{F}^{-1} \left[H(h\cdot)(\mathcal{F}g_{x,y})\psi \right](0).$$

5 Simulations

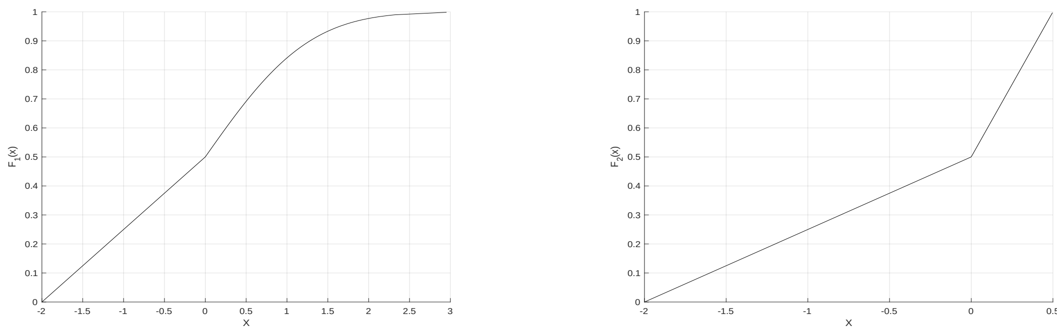
In this section we conduct a small simulation study to shed some light on the finite sample performance of the class of distribution estimators defined by Equation (1), where the distribution is continuous at x but not differentiable, and the class of estimators for the jump p_x at x defined by Equation (4).

5.1 Estimation of F at a point of continuity x

We consider the estimator $\hat{F}(x)$ for $F(x)$ at $x = 0$ with $U(x) = \int_x^{\infty} K(t)dt$ and $K(t) = \frac{3}{4}(1-t^2)$ for $|t| \leq 1$, called the Epanechnikov kernel. Data are generated from the following distributions, where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt$:

$$F_1(x) = \begin{cases} 0, \text{ if } x < -2, \\ \frac{1}{4}(x+2), \text{ if } -2 \leq x < 0, \\ \Phi(x), \text{ if } x \geq 0 \end{cases} \quad \text{and } F_2(x) = \begin{cases} 0, \text{ if } x < -2, \\ \frac{1}{4}(x+2), \text{ if } -2 \leq x < 0, \\ (x+\frac{1}{2}), \text{ if } 0 \leq x \leq \frac{1}{2}, \\ 1, \text{ if } x > \frac{1}{2}. \end{cases}$$

The distribution F_1 pieces together the distribution associated with a uniform density on $[-2,2]$ to the left of $x = 0$ and a standard Gaussian distribution to the right of $x = 0$. While F_1 is continuous at $x = 0$, it is not differentiable, with $\lim_{x \uparrow 0} F_1^{(1)}(x) = 0.25$ and $\lim_{x \downarrow 0} F_1^{(1)}(x) = \frac{1}{\sqrt{2\pi}}$. The distribution F_2 pieces together the distribution associated with a uniform density on $[-2,2]$ to the left of $x = 0$ and the distribution associated with a uniform density on $[-1/2, 1/2]$ to the right of $x = 0$. Again, F_2 is continuous but not differentiable at $x = 0$ with $\lim_{x \uparrow 0} F_2^{(1)}(x) = 0.25$ and $\lim_{x \downarrow 0} F_2^{(1)}(x) = 1$. The "kink" in F_2 at $x = 0$ is more pronounced than that of F_1. Figure 1 shows the graphs of both distributions.



(a) F_1 (b) F_2

Figure 1: The distributions F_1 and F_2.

We consider two versions of $\hat{F}(0)$. The first is calculated with a bandwidth $h \propto Cn^{-1/3}$ with $C > 0$, which is the optimal rate of decay under the assumption that F is twice differentiable at $x = 0$ (Bowman et al., 1998; Li and Racine, 2007). Given that both F_1 and F_2 are not differentiable at $x = 0$, this decay rate is not suitable for estimating $F_1(0)$ or $F_2(0)$. The second is calculated with a bandwidth $h \propto Cn^{-1/2-\epsilon}$ for $\epsilon > 0$, conforming to the requirements in Remark 1. In particular, we use $h \propto Cn^{-1/1.99}$.[3]

We consider sample sizes $n = 100, 500, 2500$ and, for each distribution, we draw $10,000$ samples. Simulated bias, variance, and mean squared error (MSE) at $x = 0$ are reported in Table 1. For comparison, we also report results for the empirical distribution F_n.

We make the following general observations. First, as expected, bias, variance and MSE decrease with

[3]Both F_1 and F_2 are Lipschitz of order $\alpha = 1$, thus it suffices to take $2\delta = 1.99$ in Remark 1. Since in both cases it is not possible to calculate the constant C, in these simulations, we set $C = 1$.

Table 1: Bias, variance$\times 10^3$, and MSE$\times 10^3$ for $\hat{F}$ with $h \propto n^{-1/3}$ and $h \propto n^{-1/1.99}$ and for F_n, at $x = 0$ and $n = 100, 500, 2500$.

	$h \propto n^{-1/3}$	$h \propto n^{-1/1.99}$	F_n
F_1			
$n = 100$			
Bias	-0.0097	-0.0046	-0.0005
Variance$\times 10^3$	0.1650	0.0734	2.5170
MSE$\times 10^3$	0.2600	0.0945	2.5172
$n = 500$			
Bias	-0.0058	-0.0020	-0.0000
Variance$\times 10^3$	0.0201	0.0066	0.4995
MSE$\times 10^3$	0.0533	0.0104	0.4995
$n = 2500$			
Bias	-0.0034	-0.0009	0.0002
Variance$\times 10^3$	0.0023	0.0006	0.0974
MSE$\times 10^3$	0.0140	0.0013	0.0974
F_2			
$n = 100$			
Bias	-0.0502	-0.0224	0.0003
Variance$\times 10^3$	0.3005	0.1416	2.4675
MSE$\times 10^3$	2.8242	0.6438	2.4675
$n = 500$			
Bias	-0.0296	-0.0098	-0.0004
Variance$\times 10^3$	0.0363	0.0126	0.5067
MSE$\times 10^3$	0.9137	0.1096	0.5069.
$n = 2500$			
Bias	-0.0173	-0.0043	-0.0000
Variance$\times 10^3$	0.0043	0.0011	0.0996
MSE$\times 10^3$	0.3023	0.0199	0.0996

sample sizes for all estimators. Second, for both distributions, using the rate of decay for the bandwidth suggested by our theoretical results produces smaller bias, variance and MSE, relative to the case where the estimator is calculated using $h \propto n^{-1/3}$. The gains can be significant, e.g., in the case of F_1 for $n = 2500$, the MSE of $\hat{F}$ calculated using $h \propto n^{-1/3}$ is ten times larger than that of $\hat{F}$ using $h \propto n^{-1/1.99}$. In the case of F_2, the MSE is more than 15 times larger. Irrespective of the rate of decay of h, $\hat{F}$ has better performance than F_n in terms of MSE but, as expected, F_n has a smaller and negligible bias. Third, bias, variance and MSE for $\hat{F}$ using $h \propto n^{-1/3}$ or $h \propto n^{-1/1.99}$ is larger for F_2, where the kink at $x = 0$ is more pronounced than in F_1. As expected, the performance of F_n is not impacted by the magnitude of the kink.

All finite sample results are in line with our expectations and confirm, in an experimental setting, the asymptotic results we have derived.

5.2 Estimation of the jump p_x

We consider two estimators for the jump p_x where $x = 0$, denoted by p_0. The first is an element of the class defined by Equation (4) with $W(t) = (1 - t^2)_+^3$, which is the exponential Fourier transform of the kernel $K(t) = \frac{48\cos(t)}{\pi x^4}\left(1 - \frac{15}{x^2}\right) - \frac{144\sin(t)}{\pi t^5}\left(2 - \frac{5}{t^2}\right)$. The second is the estimator proposed in Equation (3.10) of Murthy (1965),

$$\tilde{p}_0 = \frac{1}{n}\sum_{i=1}^{n} G\left(X_i/h\right) = \frac{1}{n}\sum_{i=1}^{n} 2\left(\int_{-\infty}^{X_i/h} K(t)dt - \chi_{\{X_i/h>0\}}\right)$$

where $G(x) = 2\left(\int_{-\infty}^{x} K(t)dt - \chi_{\{x>0\}}\right)$ and K is the standard Gaussian kernel. Now, $\tilde{p}_0$ does not belong to the class of jump estimators studied in this work because G is not continuous at every point in $\mathbb{R}$. Murthy (1965) showed that $\tilde{p}_0$ is asymptotically unbiased but did not obtain the rate of decay of this bias.

We consider data generated from the following distributions, where $p_0 = 0.2, 0.1, 0.01$ and $\mu_R = -\Phi^{-1}(0.5+p_0)$:

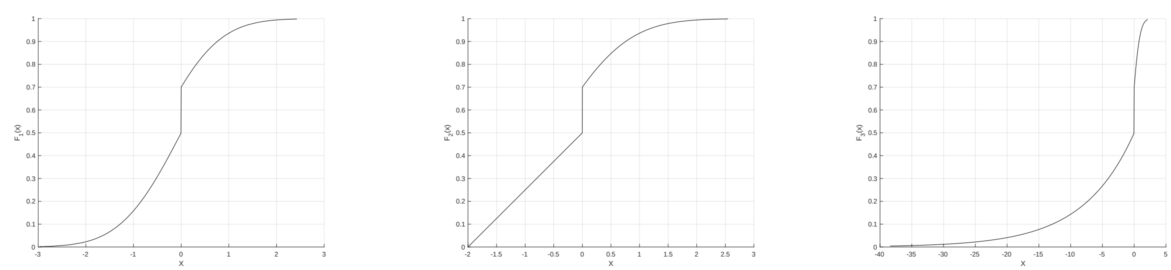
$$F_1(x) = \begin{cases} \Phi(x), \text{ if } -\infty < x < 0 \\ \frac{1}{2} + p_0, \text{ if } x = 0 \\ \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(t-\mu_R)^2\right) dt, \text{ if } x > 0; \end{cases}$$

$$F_2(x) = \begin{cases} 0, \text{ if } x < -2 \\ \frac{1}{4}(x+2), \text{ if } -2 \le x < 0 \\ \frac{1}{2} + p_0, \text{ if } x = 0 \\ \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(t-\mu_R)^2\right) dt, \text{ if } x > 0; \end{cases}$$

and

$$F_3(x) = \begin{cases} \frac{1}{2} \exp(x/8), \text{ if } -\infty < x < 0 \\ \frac{1}{2} + p_0, \text{ if } x = 0 \\ \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(t-\mu_R)^2\right) dt, \text{ if } x > 0. \end{cases}$$

For all three distributions, there is a jump at $x = 0$ of size p_0. For F_1, $\lim_{x\uparrow 0} F_1^{(1)}(x) = \lim_{x\downarrow 0} F_1^{(1)}(x)$, whereas for F_2 and F_3, these two limits are different, with the largest discrepancy occurring in the case of F_3. Figure 2 shows the graphs of each distribution with a jump $p_0 = 0.2$.



(a) F_1 (b) F_2 (c) F_3

Figure 2: The distributions F_1, F_2, and F_3 with $p_0 = 2$

As in the previous subsection, we consider sample sizes $n = 100, 500, 2500$ and, for each distribution, we draw $10,000$ samples. Both $\hat{p}_0$ and $\tilde{p}_0$ are calculated with a bandwidth $h \propto Cn^{-1/1.99}$ with $C = 1$. Simulated bias, variance and MSE at $x = 0$ are reported in Table 2.

The following general observations can be made. First, for both estimators, bias, variance and MSE decrease with sample size. Bias has no apparent relationship with jump size, remaining approximately the

Table 2: Bias, variance×10^3, and MSE×10^3 for $\hat{p}_0$ and $\tilde{p}_0$, where $n = 100, 500, 2500$ and $p_0 = 0.2, 0.1, 0.01$.

	Bias		Variance×10^3		MSE×10^3	
	$\hat{p}_0$	$\tilde{p}_0$	$\hat{p}_0$	$\tilde{p}_0$	$\hat{p}_0$	$\tilde{p}_0$
$n = 100$, $p_0 = 0.2$						
F_1	0.0329	0.0051	1.7441	1.9434	2.8267	1.9697
F_2	0.0249	-0.0074	1.7046	1.8926	2.3245	1.9471
F_3	0.0176	-0.0208	1.5976	1.7974	1.9066	2.2305
$n = 500$, $p_0 = 0.2$						
F_1	0.0146	0.0023	0.3224	0.3356	0.5362	0.3407
F_2	0.0116	-0.0030	0.3253	0.3409	0.4592	0.3499
F_3	0.0076	-0.0096	0.3182	0.3376	0.3764	0.4295
$n = 2500$, $p_0 = 0.2$						
F_1	0.0063	0.0007	0.0650	0.0665	0.1042	0.0670
F_2	0.0051	-0.0014	0.0648	0.0673	0.0909	0.0692
F_3	0.0034	-0.0042	0.0647	0.0658	0.0760	0.0838
$n = 100$, $p_0 = 0.1$						
F_1	0.0340	0.0008	1.0919	1.2755	2.2461	1.2761
F_2	0.0281	-0.0095	1.0381	1.1884	1.8279	1.2778
F_3	0.0195	-0.0242	1.0250	1.1489	1.4041	1.7333
$n = 500$, $p_0 = 0.1$						
F_1	0.0150	0.0005	0.1918	0.2095	0.4161	0.2097
F_2	0.0123	-0.0044	0.1974	0.2085	0.3496	0.2276
F_3	0.0087	-0.0107	0.1831	0.1980	0.2591	0.3123
$n = 2500$, $p_0 = 0.1$						
F_1	0.0067	0.0002	0.0378	0.0388	0.0822	0.0389
F_2	0.0053	-0.0021	0.0368	0.0377	0.0644	0.0422
F_3	0.0038	-0.0047	0.0371	0.0378	0.0518	0.0599
$n = 100$, $p_0 = 0.01$						
F_1	0.0347	0.0001	0.3363	0.4444	1.5387	0.4444
F_2	0.0284	-0.0112	0.2994	0.3948	1.1065	0.5206
F_3	0.0201	-0.0254	0.2421	0.2958	0.6479	0.9417
$n = 500$, $p_0 = 0.01$						
F_1	0.0154	0.0001	0.0419	0.0516	0.2777	0.0516
F_2	0.0125	-0.0049	0.0375	0.0456	0.1939	0.0698
F_3	0.0090	-0.0112	0.0321	0.0367	0.1123	0.1620
$n = 2500$, $p_0 = 0.01$						
F_1	0.0067	-0.0000	0.0059	0.0067	0.0513	0.0067
F_2	0.0055	-0.0022	0.0055	0.0061	0.0358	0.0108
F_3	0.0039	-0.0050	0.0050	0.0055	0.0201	0.0301

same for $p_0 = 0.2, 0.1, 0.01$ for both estimators. The same is not true for variance or MSE, which clearly decay as the jump size decreases from 0.2 to 0.01. Another clear pattern that emerges is that, in the case of F_1 where $\lim_{x\uparrow 0} F_1^{(1)}(x) = \lim_{x\downarrow 0} F_1^{(1)}(x)$ or when the difference between these two limits is small, as in the case of F_2, $\tilde{p}_0$ clearly outperforms $\hat{p}_0$ in terms of MSE. However, in the case of F_3, where the difference between $\lim_{x\uparrow 0} F_3^{(1)}(x)$ and $\lim_{x\downarrow 0} F_3^{(1)}(x)$ is much larger, $\hat{p}_0$ clearly outperforms $\tilde{p}_0$ in terms of MSE. Although not reported in Table 2, both estimators perform much worse if $h \propto Cn^{-1/3}$ with $C = 1$ is used in their calculation. Overall, the observed experimental finite sample performance confirms our theoretical results.

6 Empirical illustrations

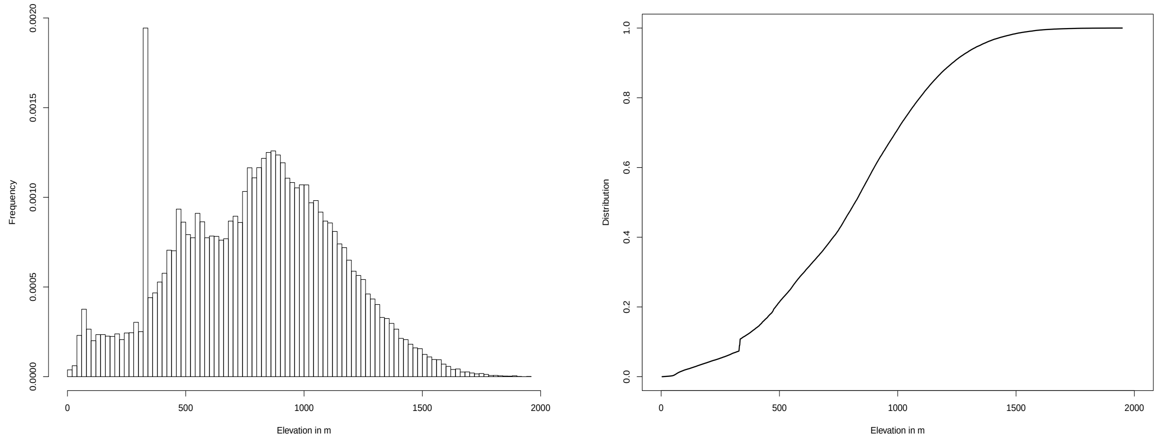
6.1 Global elevation

We look at global elevation (i.e., elevation above sea level) data from Zabotin et al. (2014). This paper examines the long-range propagation of heater-produced signals from the European Incoherent Scatter (EISCAT) heating facility at the Tromsø Observatory in the Scandinavian Mountains.

Table 3: Descriptive statistics for the empirical illustrations: elevation, in meters, from Zabotin et al. (2014) and z-statistics from Brodeur et al. (2016).

	Mean	SD	Max	Min	Mode	n
Elevation						
Full	800.0954	340.1575	1989.0000	4.0000	333.0000	2,159,615
Random	799.7601	338.6744	1949.0000	4.0000	333.0000	99,983
z-statistics						
Raw	2.4353	2.0568	9.9977	3.9961e-05	1.0000	44,952
Eye-catcher	2.3636	1.9934	9.9977	3.9961e-05	1.0000	29,565

The full data set, available from NASA[4], measures global digital elevation and is quite large, covering roughly 99% of the Earth's landmass. Zabotin et al. (2014) extracted a relatively small subsample of the total data (specifically, 68°N to 69°N and 18°E to 19.5°E), resulting in 2,159,615 observations. We took a random subsample of 99,983 observations of this subset. Table 3 gives descriptive statistics for both samples



(a) Histogram (b) Distribution

Figure 3: Histogram and distribution of our random sample of the global elevation data from Zabotin et al. (2014).

and shows that our random sample is representative of the full sample used in Zabotin et al. (2014).

Panel (a) of Figure 3 plots the histogram of our random sample of the elevation data using 100 equally spaced bins. This is analogous to Figure 5 in Zabotin et al. (2014). In both histograms, we see a sharp spike for the bin that includes the mode elevation of 333 meters, given the presence of lakes in the area under consideration. This suggests that there may be a jump in the distribution at that value. Panel (b) of Figure 3 visually confirms this phenomenon. Using our estimator of the distribution, we see a jump in the estimated distribution at $x = 333$.[5]

To formally test for this jump, we use Theorem 5 to test the null hypothesis that there is no jump at 333 meters (i.e., $H_0 : p_{333} = 0$). The estimated value of p_{333} is $\widehat{p}_{333} = 0.0175$, with a relatively tight 95% confidence interval of [0.0167, 0.0183].[6] Thus, we reject the null hypothesis of the absence of a jump at $x = 333$.

[4]https://asterweb.jpl.nasa.gov/gdem.asp

[5]As in the simulations described in Subsection 5.1, our distribution estimator was implemented using an Epanechnikov kernel and a bandwidth $h \propto Ch^{-1/1.99}$ with $C = 1$. We experimented with $C = 0.5, 2, 10$ with little numerical difference in the estimated distribution.

[6]The estimate $\hat{p}_{333}$ was obtained as described in Subsection 5.2.

6.2 P-hacking in top economics journals

As students in introductory statistics, we were required to memorize common critical values from the standard normal distribution in order to perform inference. We were also taught that there was nothing special about these values, but nonetheless they play an important role in the publication of academic papers. As null results are often seen as less important from a publication standpoint, publication bias arises, specifically towards tests rejecting null hypotheses.

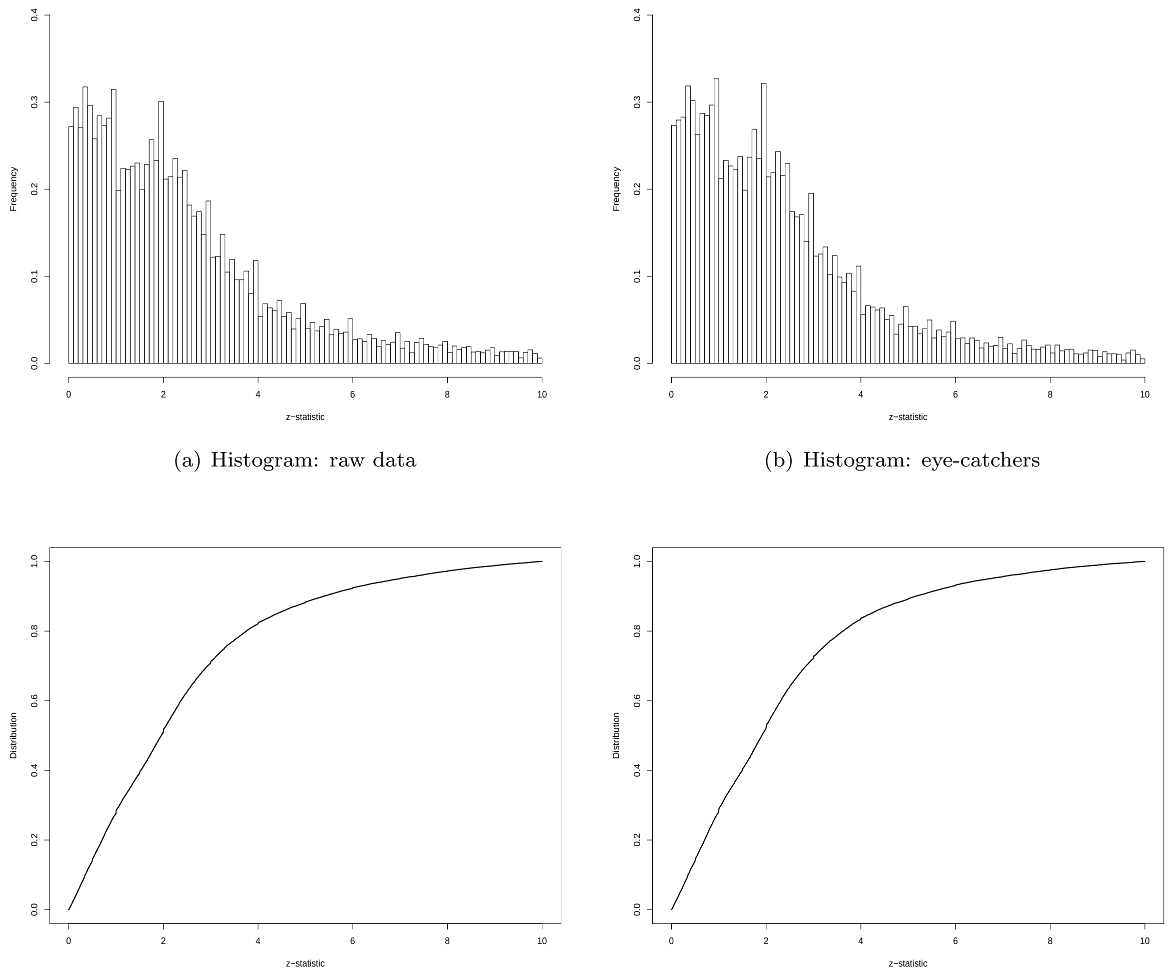
This phenomenon has been known for some time Sterling (1959) and has been studied in great detail in Brodeur et al. (2016). The authors of this paper look at three of the top journals in economics (*American Economic Review*, *Journal of Political Economy* and *Quarterly Journal of Economics*) and examine roughly 50,000 tests published in those journals over the 2005-2011 period.[7] They find visual evidence of bimodality in the estimated density of test statistics. This trough suggests an under-representation of insignificant tests just below the 5% level of significance.

In order for their kernel estimated densities to be valid, there should be no jumps in the distribution of test statistics. In the "Discontinuities" section of their online appendix, the authors attempted to identify potential discontinuities at the standard critical values of 1.65, 1.96 and 2.57. They found very little evidence of discontinuities other than for their "eye-catcher" sample at the 10% significance level, where eye-catcher is defined as a paper using asterisks to define statistical significance at a given confidence level e.g., $*$, $**$, and $***$ for 10%, 5% and 1%, respectively.

Here, we formally test for these discontinuities, specifically, at the three values used to create Figure 1A (pp. 9) and Figure 6A (pp. 27) of Brodeur et al. (2016). Our two replicated histograms with 100 bins can be found in the top row of Figure 4. Descriptive statistics corresponding to those histograms are given in Table 3. We purposely chose the "raw distribution" of the test statistics as Brodeur et al. (2016) believed that these raw distributions were more likely to have "potential discontinuities" (pp. 10).

The bottom row of Figure 4 presents our estimated distribution functions. The estimated jumps in both

[7] The full data set is available at https://www.aeaweb.org/articles?id=10.1257/app.20150044



(a) Histogram: raw data

(b) Histogram: eye-catchers

(c) Distribution: raw data

(d) Distribution: eye-catchers

Figure 4: Histograms and distributions for raw and eye-catcher z-statistics from Brodeur et al. (2016).

distributions at 1.65, 1.96, and 2.57 were zero to at least two decimal places and none were significantly different from zero. Hence, we find no evidence of jumps in either distribution at these levels.

However, looking closely at the estimated distribution, there appear to be small jumps at $x = 1$, 2, 3, 4. The first is the most common, as the mode for each distribution is shown to be equal to one in Table 3. Brodeur et al. (2016) notes that "the way authors *report* the value of their tests and the way we *reconstruct* the underlying statistics" lead to tests being "expressed as ratios of small integers" which get "overrepresented because of the low precision used by authors." The largest estimated jump in the distribution occurred at $x = 1$, where $\widehat{p}_1 = 0.0110$ with a 95% confidence bound of $[0.0100, 0.0119]$ for the raw data and $\widehat{p}_1 = 0.0120$ with a 95% confidence bound of $[0.0107, 0.0132]$ for the eye-catcher data.

In short, we see no evidence of discontinuities in the distribution of test statistics at the standard critical values, but we suggest using the "de-rounded" distribution of test statistics in Panel B of the aforementioned figures if using these data in practice.

7 Conclusion

We show that the use of the Stieltjes integral can easily provide rates of decay for biases of certain classes of estimators of distribution functions at their points of continuity and at their jumps without imposing any restrictions on the distribution. These results are useful in providing additional asymptotic characterizations for the estimators in these classes and in establishing new Fourier inversion theorems with convergence rates. Future work on exact criteria for bandwidth selection under our assumptions is needed to facilitate implementation of the estimators in empirical work.

Appendix

Lemma 1. *Proof.* Since U is bounded and continuous and F is of bounded variation, $\int_{\mathbb{R}} U\left(\frac{y-x}{h}\right) dF(y)$ exists (Natanson, 1955, p. 238). For $\varepsilon > 0$,

$$\left|\int_{\mathbb{R}} U\left(\frac{y-x}{h}\right) dF(y) - \int_{-\infty}^{x} dF(y)\right| \leq \left|\int_{y<x-\varepsilon} \left[U\left(\frac{y-x}{h}\right) - 1\right] dF(y)\right| + \left|\int_{|x-y|\leq\varepsilon} U\left(\frac{y-x}{h}\right) dF(y)\right| + \left|\int_{y>x+\varepsilon} U\left(\frac{y-x}{h}\right) dF(y)\right| + \int_{x-\varepsilon}^{x} dF(y).$$

We obtain the bounds

$$\left|\int_{y<x-\varepsilon} \left[U\left(\frac{y-x}{h}\right) - 1\right] dF(y)\right| \leq \sup_{z<-\varepsilon/h} |U(z) - 1| \int_{y<x-\varepsilon} dF(y) \leq \phi_U(\varepsilon/h)F(x)$$
$$\left|\int_{|x-y|\leq\varepsilon} U\left(\frac{y-x}{h}\right) dF(y)\right| \leq \|U\|_C \left[F(x+\varepsilon) - F(x-\varepsilon)\right] \leq \|U\|_C\, \omega(x,\varepsilon),$$
$$\left|\int_{y>x+\varepsilon} U\left(\frac{y-x}{h}\right) dF(y)\right| \leq \sup_{z>\varepsilon/h} |U(z)| \int_{y>x+\varepsilon} dF(y) \leq \phi_U(\varepsilon/h)(1 - F(x))$$

and $\int_{x-\varepsilon}^{x} dF(y) \leq \omega(x,\varepsilon)$. Collecting the above inequalities,

$$\left|\int_{\mathbb{R}} U\left(\frac{y-x}{h}\right) dF(y) - F(x)\right| \leq \phi_U(\varepsilon/h) + (1 + \|U\|_C)\, \omega(x,\varepsilon).$$

To finish the proof, we set $\varepsilon = h^\delta$. □

Theorem 1. *Proof.* (i) follows directly from Lemma 1 since $\left|E\left(\hat{F}(x)\right) - F(x)\right| = \left|\int_{\mathbb{R}} U\left(\frac{t-x}{h}\right) dF(t) - F(x)\right|$.
(ii) We have that $V(\hat{F}(x)) = \frac{1}{n}\left\{E\left(U^2\left(\frac{X-x}{h}\right)\right) - \left[E\left(U\left(\frac{X-x}{h}\right)\right)\right]^2\right\}$. From part (i) we know that $E\left(U\left(\frac{X-x}{h}\right)\right) = F(x) + o(1)$ as $h \to 0$. U^2 satisfies Assumption 1, so by Lemma 1,

$$\left|E\left(U^2\left(\frac{X-x}{h}\right)\right) - F(x)\right| \leq \phi_{U^2}(h^{\delta-1}) + \omega(x,h^\delta)(1 + \left\|U^2\right\|_C) \to 0 \text{ as } h \to 0.$$

With $F(x) \in (0,1)$ the statement follows. □

Corollary 1. *Proof.* Since $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$, the variation of F on $(-\infty, -N]$ and $[N, \infty)$ will be small for large $N > 0$. On $[-N, N]$, F is uniformly continuous and its variation on the segment $[x-\varepsilon, x+\varepsilon]$ will be small for small ε and all $x \in [-N, N]$. Thus, (3) is true and the statement follows from Theorem 1 (i). □

Corollary 2. *Proof.* Let $F_n(x) = \frac{1}{n}\sum_{i=1}^n e(x-X_i)$, where $e(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases}$ is the empirical distribution function. Then $\sup_{x\in\mathbb{R}} |\hat{F}(x)-F(x)| \leq \sup_{x\in\mathbb{R}} |\hat{F}(x)-F_n(x)|+\sup_{x\in\mathbb{R}} |F_n(x)-F(x)|$. By the Glivenko-Cantelli theorem, $\sup_{x\in\mathbb{R}} |F_n(x)-F(x)| \to 0$ almost surely, so to complete the proof we need only show that $\sup_{x\in\mathbb{R}} |\hat{F}(x) - F_n(x)| \to 0$ almost surely. Since $\hat{F}(x) = \int_{\mathbb{R}} U\left(\frac{y-x}{h}\right) dF_n(y)$ and $F_n(x) = \int_{\mathbb{R}} e(x-y)dF_n(y)$, $|\hat{F}(x) - F_n(x)| \leq \int_{\mathbb{R}} \left|U\left(\frac{y-x}{h}\right) - e(x-y)\right| dF_n(y)$. Then for $\varepsilon > 0$,

$$\begin{aligned}\int_{\mathbb{R}} \left|U\left(\frac{y-x}{h}\right) - e(x-y)\right| dF_n(y) &= \int_{y<x-\varepsilon} \left|U\left(\frac{y-x}{h}\right) - 1\right| dF_n(y) \\ &+ \int_{|x-y|\leq\varepsilon} \left|U\left(\frac{y-x}{h}\right) - e(x-y)\right| dF_n(y) \\ &+ \int_{y>x+\varepsilon} \left|U\left(\frac{y-x}{h}\right)\right| dF_n(y).\end{aligned}$$

From the proof of Lemma 1, since F_n is a distribution,

$$\begin{aligned}|\hat{F}(x) - F_n(x)| &\leq \phi_U(\varepsilon/h) + (1 + \|U\|_C)(F_n(x+\varepsilon) - F_n(x-\varepsilon)) \\ &\leq \phi_U(\varepsilon/h) + (1 + \|U\|_C)\left(|F_n(x+\varepsilon) - F(x+\varepsilon)| + |F(x+\varepsilon) - F(x-\varepsilon)|\right. \\ &\left. + |F(x-\varepsilon) - F_n(x-\varepsilon)|\right).\end{aligned}$$

Letting $\varepsilon = h^\delta$ for $\delta \in (0,1)$,

$$\begin{aligned}\sup_{x\in\mathbb{R}} |\hat{F}(x) - F_n(x)| \leq \phi_U(h^{\delta-1}) + (1 + \|U\|_C)\Bigg(&\sup_{x\in\mathbb{R}} \left|F_n(x+h^\delta) - F(x+h^\delta)\right| + \sup_{x\in\mathbb{R}} \omega(x, h^\delta) \\ &+ \sup_{x\in\mathbb{R}} \left|F(x-h^\delta) - F_n(x-h^\delta)\right|\Bigg).\end{aligned}$$

By the continuity of F and the uniform, almost sure convergence of F_n to F, $\sup_{x\in\mathbb{R}} |\hat{F}(x) - F_n(x)| \to 0$ almost surely as $h \to 0$, which completes the proof. □

Theorem 2. *Proof.* Let $Z_{in} = \frac{1}{n}U\left(\frac{X_i-x}{h}\right)$, $\mu_n = E(Z_{in})$ and $s_n^2 = \sum_{i=1}^n E(Z_{in}-\mu_n)^2$. Then $s_n^2 = V(\hat{F}(x))$ and

$$\frac{\hat{F}(x) - E(\hat{F}(x))}{\sqrt{V(\hat{F}(x))}} = \sum_{i=1}^n \frac{Z_{in} - \mu_n}{s_n} = \sum_{i=1}^n X_{in},$$

where $X_{in} = \frac{Z_{in}-\mu_n}{s_n}$, $E(X_{in}) = 0$, $V(X_{in}) = \frac{1}{s_n^2}E(Z_{in} - \mu_n)^2$, and $\sum_{i=1}^n V(X_{in}) = 1$. By Liapounov's central limit theorem, $\sum_{i=1}^n X_{in} \xrightarrow{d} \mathcal{Z} \sim N(0,1)$, provided that $\lim_{n\to\infty} \sum_{i=1}^n E\left(|X_{in}|^{2+\theta}\right) = 0$ for some

$\theta > 0$ (see Davidson, 1994, p. 369-372). It follows from the assumption that the X_i are independent and identically distributed that

$$\sum_{i=1}^{n} E\left(|X_{in}|^{2+\theta}\right) = nV(\hat{F}(x))^{-1-\theta/2}E(|Z_{in}-\mu_n|^{2+\theta})$$

and, by the c_r inequality, $E(|Z_{in}-\mu_n|^{2+\theta}) \leq 2^{1+\theta}(E(|Z_{in}|^{2+\theta}) + |\mu_n|^{2+\theta})$. From Theorem 1, if $h \to 0$, $n|\mu_n|^{2+\theta} = \frac{1}{n^{1+\theta}}|F(x)+o(1)|^{2+\theta}$. Consequently,

$$\begin{aligned}\sum_{i=1}^{n} E\left(|X_{in}|^{2+\theta}\right) &\leq nV(\hat{F}(x))^{-1-\theta/2}2^{1+\theta}\left(\frac{1}{n^{2+\theta}}E\left(\left|U\left(\frac{X_1-x}{h}\right)\right|^{2+\theta}\right) + |\mu_n|^{2+\theta}\right)\\ &= (nV(\hat{F}(x)))^{-1-\theta/2}2^{1+\theta}n^{-\theta/2}\left(E\left(\left|U\left(\frac{X_1-x}{h}\right)\right|^{2+\theta}\right) + |F(x)+o(1)|^{2+\theta}\right)\\ &\to 0 \text{ as } n\to\infty\end{aligned}$$

since $nV(\hat{F}(x)) \to F(x)(1-F(x))$ and $E\left(\left|U\left(\frac{X_1-x}{h}\right)\right|^{2+\theta}\right) = O(1)$ by the fact that $U^{2+\theta}$ satisfies Assumption 1. Now, noting that $\sqrt{n}(\hat{F}(x)-F(x)) = \sqrt{n}(\hat{F}(x)-E(\hat{F}(x))) + \sqrt{n}(E(\hat{F}(x))-F(x))$, $\sqrt{n}(\hat{F}(x)-F(x)) \overset{d}{\to} N(0,(1-F(x))F(x))$ provided that $\sqrt{n}(E(\hat{F}(x))-F(x)) \to 0$. But from Theorem 1,

$$|\sqrt{n}(E(\hat{F}(x))-F(x))| \leq \sqrt{n}\phi_U(h^{\delta-1}) + \sqrt{n}\,\omega(x,h^{\delta})(1+||U||_C).$$

Since $\sqrt{n}\,\phi_U(h^{\delta-1}) \to 0$ and $\sqrt{n}\,\omega(x,h^{\delta}) \to 0$ by assumption, the proof is complete. □

Lemma 2. *Proof.* We set $\left[\frac{z-y}{h},\frac{z-x}{h}\right] \equiv [a,b]$ and write

$$\begin{aligned}\int_{\mathbb{R}} G([a,b])\,dF(z) - \int_x^y dF(z) &= \int_{z<x-\varepsilon} G([a,b])\,dF(z) + \int_{x+\varepsilon}^{y-\varepsilon}[G([a,b])-1]\,dF(z)\\ &+ \int_{z>y+\varepsilon} G([a,b])\,dF(z) + \int_{|z-x|\leq\varepsilon} G([a,b])\,dF(z)\\ &+ \int_{|z-y|\leq\varepsilon} G([a,b])\,dF(z) + \int_x^{x+\varepsilon} dF(z) + \int_{y-\varepsilon}^{y} dF(z).\end{aligned}$$

In the first integral, $\frac{z-y}{h} < \frac{z-x}{h} < -\frac{\varepsilon}{h}$, so

$$\left|\int_{z<x-\varepsilon} G([a,b])\,dF(z)\right| \leq \sup_{b<-\varepsilon/h} G([a,b])\,F(x) \leq \phi_G(\varepsilon/h)F(x).$$

In the second integral, $\frac{z-y}{h} < -\frac{\varepsilon}{h}$, $\frac{z-x}{h} > \frac{\varepsilon}{h}$ and

$$\left|\int_{x+\varepsilon}^{y-\varepsilon}[G([a,b])-1]\,dF(z)\right| \leq \sup_{a<-\varepsilon/h,\ b>\varepsilon/h}|G([a,b])-1|(F(y)-F(x)) \leq \phi_G(\varepsilon/h)(F(y)-F(x)).$$

In the third integral, $\frac{z-x}{h} > \frac{z-y}{h} > \frac{\varepsilon}{h}$ and

$$\left|\int_{z>y+\varepsilon} G([a,b])\, dF(z)\right| \le \sup_{a>\varepsilon/h} G([a,b])\,(1-F(y)) \le \phi_G(\varepsilon/h)(1-F(y)).$$

The sum of the remaining four integrals is obviously bounded by $(1+\|G\|_C)\,[\omega(x,\varepsilon)+\omega(y,\varepsilon)]$. Thus, to finish the proof it remains to add the bounds and set $\varepsilon = h^{\delta}$. □

Theorem 3. *Proof.* Part (i) immediately follows from Lemma 2 because $E\hat{F}(x,y) = \int_{\mathbb{R}} G\left(\left[\frac{z-y}{h}, \frac{z-x}{h}\right]\right) dF(z)$.

Part (ii) follows from

$$V(\hat{F}(x,y)) = \frac{1}{n}\left\{EG^2\left(\left[\frac{X-y}{h}, \frac{X-x}{h}\right]\right) - \left[EG\left(\left[\frac{X-y}{h}, \frac{X-x}{h}\right]\right)\right]^2\right\},$$

where G^2 satisfies Assumption 2 and, by Lemma 2,

$$\left|EG^2\left(\left[\frac{X-y}{h}, \frac{X-x}{h}\right]\right) - F(x,y)\right| \le \phi_{G^2}(h^{\delta-1}) + (1+\left\|G^2\right\|_C)\left[\omega(x,h^{\delta})+\omega(y,h^{\delta})\right] \to 0 \text{ as } h \to 0.$$

□

Lemma 3. *Proof.* (i) By the additivity of probability for any $h_1 \in (0,h)$,

$$p_x \le \int_{|z-x|<h_1} dF(z) \le \int_{|z-x|<h} dF(z) = p_x + \delta_F(h). \tag{10}$$

This implies that

$$\int_{h_1\le|z-x|<h} dF(z) = \int_{|z-x|<h} dF(z) - \int_{|z-x|<h_1} dF(z) \le \delta_F(h) \tag{11}$$

and

$$\left|\int_{|z-x|<h_1} dF(z) - p_x\right| \le \delta_F(h). \tag{12}$$

We wish to bound

$$\begin{aligned}\int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x \quad &= \quad \int_{|z-x|<h_1} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x \\ &+ \quad \int_{h_1\le|z-x|<h} W\left(\frac{z-x}{\lambda}\right) dF(z) + \int_{|z-x|\ge h} W\left(\frac{z-x}{\lambda}\right) dF(z). \end{aligned} \tag{13}$$

Obviously,

$$\left|\int_{|z-x|\ge h} W\left(\frac{z-x}{\lambda}\right) dF(z)\right| \le \phi_W(h/\lambda). \tag{14}$$

By Equations (10) and (12),

$$
\begin{aligned}
\left|\int_{|z-x|<h_1} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x\right| &\leq \left|\int_{|z-x|<h_1} \left[W\left(\frac{z-x}{\lambda}\right) - W(0)\right] dF(z)\right| \\
&+ \left|W(0)\int_{|z-x|<h_1} dF(z) - p_x\right| \\
&\leq \omega_W(h_1/\lambda)\int_{|z-x|<h_1} dF(z) + \left|\int_{|z-x|<h_1} dF(z) - p_x\right| \\
&\leq \omega_W(h_1/\lambda)\left[p_x + \delta_F(h)\right] + \delta_F(h). \qquad (15)
\end{aligned}
$$

By Equation (11)

$$
\left|\int_{h_1\leq|z-x|<h} W\left(\frac{z-x}{\lambda}\right) dF(z)\right| \leq \|W\|_C \int_{h_1\leq|z-x|<h} dF(z) \leq \|W\|_C\, \delta_F(h). \qquad (16)
$$

Combining (14)–(16),

$$
\left|\int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x\right| \leq \omega_W(h_1/\lambda)\left[p_x + \delta_F(h)\right] + \left(1 + \|W\|_C\right)\delta_F(h) + \phi_W(h/\lambda).
$$

It suffices to set $h = \lambda^{1-\varepsilon_1}$ and $h_1 = h\lambda^{\varepsilon_2} = \lambda^{1+\varepsilon_2-\varepsilon_1} < h$ to finish the proof.

(ii) Instead of Equation (13), write

$$
\int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x = \int_{|z-x|<h} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x + \int_{|z-x|\geq h} W\left(\frac{z-x}{\lambda}\right) dF(z).
$$

For all small h, $\int_{|z-x|<h} W\left(\frac{z-x}{\lambda}\right) dF(z) = W(0)p_x = p_x$. By (14) we see that the statement in the lemma is true. □

Theorem 4. *Proof.* Both (i) and (ii) follow from Lemma 3. Part (iii) follows from (i) since

$$
V(\hat{p}_x) = \frac{1}{n}\left\{EW^2\left(\frac{X-x}{h}\right) - \left[EW\left(\frac{X-x}{h}\right)\right]^2\right\}
$$

and the fact that W^2 satisfies Assumption 3. □

Theorem 5. *Proof.* In the proof of Theorem 2, let $Z_{in} = \frac{1}{n}W\left(\frac{X_i-x}{h}\right)$ and proceed with the same arguments to conclude that $\sqrt{n}(\hat{p}_x - E(\hat{p}_x)) \xrightarrow{d} N(0, (1-p_x)p_x)$. Use Theorem 4 to obtain

$$
|\sqrt{n}(E(\hat{p}_x) - p_x)| \leq \sqrt{n}\left(\omega_W(h^{\varepsilon_2-\varepsilon_1})\left[p_x + \delta_F(h^{1-\varepsilon_1})\right] + \left(1 + \|W\|_C\right)\delta_F(h^{1-\varepsilon_1}) + \phi_W(h^{-\varepsilon_1})\right).
$$

Hence, given the conditions stated in the theorem, $\sqrt{n}(E(\hat{p}_x) - p_x) \to 0$ as $n \to \infty$. □

Theorem 6. *Proof.* Since F_1 is purely discrete, its conjugate distribution, defined by $F_2(x) = 1 - F_1(-x_-)$, is also discrete. If F_1 has a jump p_j at ξ_j, then F_2 has the same jump at $-\xi_j$. By Theorem 3.3.3 of Lukacs (1970), the convolution $F = F_1 * F_2$ has a jump $\sum_j p_j^2$ at 0. Using Equation (3.3.2) from Lukacs (1970), we have that $(\mathcal{F}F) = (\mathcal{F}F_1)(\mathcal{F}F_2) = (\mathcal{F}F_1)\overline{(\mathcal{F}F_1)} = |(\mathcal{F}F_1)|^2$, where the bar denotes complex conjugation. Thus, (5) implies that

$$\int_{\mathbb{R}} (\mathcal{F}K)\left(\frac{z}{\lambda}\right) dF(z) = \int_{\mathbb{R}} (\mathcal{F}F)(t)\lambda K(\lambda t)dt = \int_{\mathbb{R}} |(\mathcal{F}F_1)(t)|^2 \lambda K(\lambda t)dt. \tag{17}$$

By Lemma 3 and Equation (17),

$$\begin{aligned}
\left|\int_{\mathbb{R}} |(\mathcal{F}F_1)(t)|^2 \lambda K(\lambda t)dt - \sum_j p_j^2\right| &= \left|\int_{\mathbb{R}} (\mathcal{F}K)\left(\frac{z}{\lambda}\right) dF(z) - \sum_j p_j^2\right| \\
&\leq \omega_{\tilde{K}}(\lambda^{\varepsilon_2 - \varepsilon_1}) \left[\sum_j p_j^2 + \delta_F(\lambda^{1-\varepsilon_1})\right] + (1 + \|(\mathcal{F}K)\|_C)\, \delta_F(\lambda^{1-\varepsilon_1}) \\
&+ \sup_{|z| \geq \lambda^{-\varepsilon_1}} |(\mathcal{F}K)(z)| \to 0.
\end{aligned}$$

□

Theorem 7. *Proof.* (i) We derive two equivalent formulas for the expression under the limit sign in Equation (9). Denoting

$$I_{x,y}(h,t) = \mathcal{F}^{-1}\left[H(h\cdot)(\mathcal{F}g_{x,y})\right](t),$$

we have that

$$\mathcal{F}^{-1}\left[H(h\cdot)(\mathcal{F}g_{x,y})(\mathcal{F}F)\right](0) = \int_{\mathbb{R}} I_{x,y}(h,-t)dF(t), \tag{18}$$

where the change in integration order is made possible by the integrability of H. For the same reason,

$$\begin{aligned}
I_{x,y}(h,-t) &= \left[(\mathcal{F}^{-1}H(h\cdot)) * g_{x,y}\right](-t) = \int_{\mathbb{R}} (\mathcal{F}^{-1}H(h\cdot))(-t-u)g_{x,y}(u)du \\
&= \frac{1}{2\pi}\int \left(\int_{\mathbb{R}} e^{i(t+u)s}H(hs)ds\right) g_{x,y}(u)du \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} (\mathcal{F}H(h\cdot))(t+u)g_{x,y}(u)du.
\end{aligned}$$

Writing $(\mathcal{F}H(h\cdot))(s) = (\mathcal{F}H)(s/h)/h$,

$$\begin{aligned} I_{x,y}(h,-t) &= \frac{1}{2\pi h}\int_{\mathbb{R}} (\mathcal{F}H)\left(\frac{t+u}{h}\right)\chi_{[x,y]}(-u)du \quad \left(\text{replacing } \frac{t+u}{h} = v\right) \\ &= \frac{1}{2\pi}\int_{(t-y)/h}^{(t-x)/h} (\mathcal{F}H)(v)\, dv = G\left(\left[\frac{t-y}{h}, \frac{t-x}{h}\right]\right). \end{aligned} \tag{19}$$

On the other hand, since $(\mathcal{F}g_{x,y})(t) = \int_{\mathbb{R}} e^{ist}\chi_{[x,y]}(-s)ds = \frac{e^{-ixt}-e^{-iyt}}{it}$,

$$\mathcal{F}^{-1}\left[H(h\cdot)(\mathcal{F}g_{x,y})\psi\right](0) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{-ixt}-e^{-iyt}}{it} H(ht)\psi(t)dt. \tag{20}$$

Equations (18)–(20) imply that

$$\frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{-ixt}-e^{-iyt}}{it} H(ht)\psi(t)dt = \int_{\mathbb{R}} G\left(\left[\frac{t-y}{h}, \frac{t-x}{h}\right]\right) dF(t).$$

An application of Lemma 2 provides the stated bound, finishing the proof of part (i).

(ii) If $H = \chi_{[-1,1]}$, then $(\mathcal{F}H)(s) = \frac{e^{is}-e^{-is}}{is} = \frac{2\sin s}{s}$ and

$$G([a,b]) = \frac{1}{\pi}\int_a^b \frac{\sin s}{s} ds \to \begin{cases} 1, & a \to -\infty,\ b \to \infty \\ 0, & a \to \infty \\ 0, & b \to -\infty \end{cases}$$

by Lemma 3.2.1 of Lukacs (1970). Assumption 2 is satisfied and the use of Lemma 2 completes the proof.

(iii) For $H(t) = e^{-t^2}$, $(\mathcal{F}H)(s) = \int e^{ist-t^2}dt = \sqrt{\pi}e^{-s^2/4} = 2\pi\phi(s;0,2)$ (Cramér, 1946, p.99), where $\phi(s;0,2)$ is the value of a Gaussian density with parameters $\mu = 0$ and $\sigma^2 = 2$, evaluated at s. Hence,

$$G([a,b]) = \int_a^b \phi(s;0,2)ds,$$

which satisfies Assumption 2. Again, the use of Lemma 2 completes the proof. □

Theorem 8. *Proof.* The proof follows from Lemma 3, noting that

$$\int_{\mathbb{R}} e^{-itx}(\mathcal{F}F)(t)\lambda K(\lambda t)dt = \int_{\mathbb{R}} (\mathcal{F}K)\left(\frac{z-x}{\lambda}\right) dF(z). \tag{21}$$

□

Acknowledgement

This research has been funded by grant AP09561264 from the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan.

References

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

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika,* 68, 326–328.

Berg, A. & Politis, D. (2009). Cdf and survival function estimation with infinite order kernels. *Electronic Journal of Statistics,* 3, 1436–1454.

Borovkov, A. (2009). *Probability Theory.* Universitext. Springer, London.

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

Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star wars: the empirics strike back. *American Economic Journal: Applied Economics,* 8, 1–32.

Chacón, J. E. & Rodríguez-Casal, A. (2010). A note on the universal consistency of the kernel distribution function estimator. *Statistics and Probability Letters,* 80, 1414–1419.

Cheng, F. (2017). Strong uniform consistency rates of kernel estimators of cumulative distribution functions. *Communications in Statistics - Theory and Methods,* 46, 6803–6807.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Asia Publishing House, Bombay.

Davidson, J. (1994). *Stochastic limit theory.* Oxford University Press, New York.

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

Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika,* 38, 481–482.

Giné, E. & Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probability Theory and Related Fields,* 143, 569–596.

Iosida, K. (1980). *Functional Analysis.* Springer, Berlin.

Li, Q. & Racine, J., 2007. *Nonparametric econometrics: theory and practice.* Princeton University Press, Princeton, NJ.

Lukacs, E. (1970). *Characteristic Functions.* Griffin, London, UK.

Murthy, V. K. (1965). Estimation of jumps, reliability and hazard rate. *Annals of Mathematical Statistics,* 36, 1032–1040.

Mynbaev, K. (2012). Distributions escaping to infinity and the limiting power of the Cliff-Ord test for autocorrelation. *ISRN Probability and Statistics,* (article number 926164), 1–39.

Nadaraya, E. A. (1964). Some new estimates for distribution functions. *Theory of Probability and Applications,* 15, 497–500.

Natanson, I. P. (1955). *Theory of functions of a real variable. Vol. 1.* Frederick Ungar Publishing Co., New York, NY.

Shirahata, S. & Chu, I.-S. (1992). Integrated squared error of kernel-type estimator of distribution function. *Annals of the Institute of Statistical Mathematics,* 44, 579–591.

Shiryaev, A. N. (1996). *Probability.* 2nd Edition. No. 95 in Graduate Texts in Mathematics. Springer, Berlin.

Sterling, T. D. (1959). Publication decision and their possible effects on inferences drawn from tests of significance or vice-versa. *Journal of the American Statistical Association,* 54, 30–34.

Swanepoel, J. & Van Graan, F. (2005). A new kernel distribution function estimator based on a non-parametric transformation of the data. *Scandinavian Journal of Statistics,* 32, 551–562.

Tenreiro, C. (2013). Boundary kernels for distribution function estimation. *REVSTAT-Statistical Journal,* 11, 169–190.

Watson, G. S. & Leadbetter, M. (1964). Hazard analysis ii. *Sankhya,* A 26, 101–116.

Winter, B. (1973). Strong uniform consistency of integrals of density estimators. *The Canadian Journal of Statistics,* 1, 247–253.

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

Zabotin, N. A., Zavarotny, V. U., & Rietveld, M. T. (2014). Physical mechanisms associated with long-range propagation of the signals from ionospheric heating experiments. *Radio Science,* 49, 987–995.