3 Distribution function estimation over $\mathbb{R}$

It is natural to define an estimator for the distribution function at $F(x)$ at x by integrating $\hat{f}$. Hence, we have

Definition 8. *Let $\hat{f}(x)$ be the kernel density estimator. The kernel distribution function estimator is given by*

$$\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(z)dz. \tag{20}$$

Theorem 17. *Assume that $\{X_i\}_{i=1}^n$ is an IID sequence and that K is a density function. Then,*

$$E(\hat{F}(x)) - F(x) = \int (F(x + h_n\gamma) - F(x))K(\gamma)d\gamma \tag{21}$$

and

$$V(\hat{F}(x)) = \frac{1}{n}\left(2\int F(x + h_n\gamma)(1 - G(\gamma))K(\gamma)d\gamma - \left(\int F(x + h_n\gamma)K(\gamma)d\gamma\right)^2\right) \tag{22}$$

where $G(\gamma) = \int_{-\infty}^{\gamma} K(z)dz$.

Proof. By definition $\hat{F}(x) = \frac{1}{nh_n}\sum_{i=1}^n \int_{-\infty}^{x} K\left(\frac{X_i - y}{h_n}\right) dy$. By a change of variable in integration we write $\hat{F}(x) = \frac{1}{nh_n}\sum_{i=1}^n h_n \int_{\frac{X_i - x}{h_n}}^{\infty} K(z)dz$. Consequently, by the assumption that the sequence $\{X_i\}_{i=1}^n$ is IID we have

$$\begin{aligned}
E(\hat{F}(x)) &= \frac{1}{n}\sum_{i=1}^{n}\int\int_{\frac{y-x}{h_n}}^{\infty} K(z)dzf(y)dy = \int\int_{\frac{y-x}{h_n}}^{\infty} K(z)dzf(y)dy \\
&= 1 - \int\int_{-\infty}^{\frac{y-x}{h_n}} K(z)dzf(y)dy \text{ given that } \int K(z)dz = 1 \\
&= 1 - h_n\int\int_{-\infty}^{\gamma} K(z)dzf(x + h_n\gamma)d\gamma \text{ by letting } \gamma = \tfrac{y-x}{h_n} \\
&= 1 - \int G(\gamma)dF(x + h_n\gamma)
\end{aligned}$$

Integrating by parts, we have $\int G(\gamma)dF(x+h_n\gamma) = \lim_{\gamma\to\infty} G(\gamma)F(x+h_n\gamma) - \lim_{\gamma\to-\infty} G(\gamma)F(x+h_n\gamma) - \int F(x+h_n\gamma)dG(\gamma) = 1 - \int F(x+h_n\gamma)K(\gamma)d\gamma$. Thus, $E(\hat{F}(x)) = \int F(x+h_n\gamma)K(\gamma)d\gamma$ and given that $\int K(\gamma)d\gamma = 1$

$$E(\hat{F}(x)) - F(x) = \int (F(x + h_n\gamma) - F(x))K(\gamma)d\gamma. \tag{23}$$

Now,

$$
\begin{aligned}
V(\hat{F}(x)) &= E\left(\left(\left(1-\frac{1}{n}\sum_{i=1}^{n}G\left(\frac{X_i-x}{h_n}\right)\right)-\left(1-E\left(G\left(\frac{X_i-x}{h_n}\right)\right)\right)\right)^2\right)\\
&= E\left(\frac{1}{n^2}\left(\sum_{i=1}^{n}\left(G\left(\frac{X_i-x}{h_n}\right)-E\left(G\left(\frac{X_i-x}{h_n}\right)\right)\right)\right)^2\right)\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(\left(G\left(\frac{X_i-x}{h_n}\right)-E\left(G\left(\frac{X_i-x}{h_n}\right)\right)\right)\right.\\
&\times\left.\left(G\left(\frac{X_j-x}{h_n}\right)-E\left(G\left(\frac{X_j-x}{h_n}\right)\right)\right)\right)\\
&= \frac{1}{n}V\left(G\left(\frac{X_i-x}{h_n}\right)\right)=\frac{1}{n}\left(E\left(G^2\left(\frac{X_i-x}{h_n}\right)\right)-\left(E\left(G\left(\frac{X_i-x}{h_n}\right)\right)\right)^2\right)
\end{aligned}
$$

where the penultimate equality follows from the assumption that $\{X_i\}_{i=1}^n$ is IID. Using integration by parts and a change of variable as in obtaining the expression for $E(\hat{F}(x))$ we have

$$E\left(G^2\left(\frac{X_i-x}{h_n}\right)\right)=1-2\int F(x+h_n\gamma)K(\gamma)G(\gamma)d\gamma.$$

Hence, given that $E\left(G\left(\frac{X_i-x}{h_n}\right)\right)=1-\int F(x+h_n\gamma)K(\gamma)d\gamma$ we have

$$V(\hat{F}(x))=\frac{1}{n}\left(2\int F(x+h_n\gamma)(1-G(\gamma))K(\gamma)d\gamma-\left(\int F(x+h_n\gamma)K(\gamma)d\gamma\right)^2\right). \quad (24)$$

□

Remarks. 1. Note that $E(\hat{F}(x))-F(x)$ is expressed with the same algebraic structure as the right-hand side of $E(\hat{f}(x))-f(x)=\int(f(x+h_n\gamma)-f(x))K(\gamma)d\gamma$ with F replacing f. In the case of $\hat{f}$ we used Theorem 7, which required $\int|f(\gamma)|d\gamma<\infty$, to show that $E(\hat{f}(x))-f(x)\to 0$ as $n\to\infty$. In the case of $\hat{F}$, it is not generally true that $\int F(\gamma)d\gamma<\infty$, thus we cannot make use of Theorem 7. Additional assumptions are needed to study the behavior of the bias and the variance of $\hat{F}$.

Theorem 18. *Assume that a) $\{X_i\}_{i=1}^n$ is an IID sequence; b) K is a density function with $\int\gamma K(\gamma)d\gamma=0$ and $\int\gamma^2K(\gamma)d\gamma<\infty$; c) $F(x)$ has bounded second derivative on $\mathbb{R}$ and $F^{(1)}(x)$ is continuous on $\mathbb{R}$; d) $0<h_n\to 0$ as $n\to\infty$. Then, $E(\hat{F}(x))=F(x)+O(h_n^2)$ and $V(\hat{F}(x))=\frac{F(x)(1-F(x))}{n}-\frac{h_n}{n}f(x)\alpha+o\left(\frac{h_n}{n}\right)$ where $\alpha=2\int\gamma K(\gamma)G(\gamma)d\gamma$.*

Proof. By c), for any $x \in \mathbb{R}$ and some $0 \le \lambda \le 1$ we have $F(x+h_n\gamma) = F(x) + F^{(1)}(x)h_n\gamma + \frac{1}{2}F^{(2)}(x+\lambda h_n\gamma)h_n^2\gamma^2$. Then, from Theorem 17

$$\int F(x+h_n\gamma)K(\gamma)d\gamma = F(x) + \frac{h_n^2}{2}\int F^{(2)}(x+\lambda h_n\gamma)\gamma^2 K(\gamma)d\gamma = F(x) + O(h_n^2)$$

where the last integral exists given that $F^{(2)}$ is bounded and $\int \gamma^2 K(\gamma)d\gamma < \infty$. Similarly,

$$\begin{aligned}
\int 2F(x+h_n\gamma)K(\gamma)(1-G(\gamma))d\gamma = F(x)\int 2K(\gamma)(1-G(\gamma))d\gamma \\
&+ h_nF^{(1)}(x)\int 2\gamma K(\gamma)(1-G(\gamma))d\gamma \\
&+ h_n^2\int F^{(2)}(x+\lambda h_n\gamma)(1-G(\gamma))\gamma^2K(\gamma)d\gamma \\
&= 2F(x) - F(x)\,2\int G(\gamma)K(\gamma)d\gamma \\
&+ 2h_nF^{(1)}(x)\int \gamma K(\gamma)d\gamma - h_nF^{(1)}(x)2\int \gamma K(\gamma)G(\gamma)d\gamma \\
&+ h_n^2\int F^{(2)}(x+\lambda h_n\gamma)\gamma^2K(\gamma)d\gamma \\
&- h_n^2\int F^{(2)}(x+\lambda h_n\gamma)\gamma^2K(\gamma)G(\gamma)d\gamma.
\end{aligned}$$

Note that all integrals on the right hand side of the last equality exist since: integrating by parts we have that $2\int G(\gamma)K(\gamma)d\gamma = 1$, by b) $\int \gamma K(\gamma)d\gamma = 0$, $\left|\int \gamma K(\gamma)G(\gamma)d\gamma\right| \le \int |\gamma|K(\gamma)d\gamma < \infty$ by b) and the fact that $G(\gamma) \le 1$, $\left|\int F^{(2)}(x+\lambda h_n\gamma)\gamma^2K(\gamma)G(\gamma)d\gamma\right| \le C\int \gamma^2K(\gamma)d\gamma < \infty$ by b), c) and $G(\gamma) \le 1$. Hence, noting that $F^{(1)}(x) = f(x)$ and letting $\alpha = 2\int \gamma K(\gamma)G(\gamma)d\gamma$ we have

$$\int 2F(x+h_n\gamma)K(\gamma)(1-G(\gamma))d\gamma = F(x) - h_nf(x)\alpha + O(h_n^2). \tag{25}$$

Consequently, $E(\hat{F}(x)) - F(x) = O(h_n^2)$ and

$$\begin{aligned}
V(\hat{F}(x)) &= \frac{1}{n}\left(F(x) - h_nf(x)\alpha + O(h_n^2) - (F(x)+O(h_n^2))^2\right) \\
&= \frac{F(x)(1-F(x))}{n} - \frac{h_n}{n}f(x)\alpha + \frac{h_n^2}{n}O(1) + \frac{h_n^4}{n}O(1) \\
&= \frac{F(x)(1-F(x))}{n} - \frac{h_n}{n}f(x)\alpha + o\left(\frac{h_n}{n}\right) \text{ since } h_n \to 0.
\end{aligned}$$

$\square$

Remarks. 1. Under the conditions of Theorem 18 $\hat{F}(x)$ converges in quadratic mean to $F(x)$ and consequently $\hat{F}(x) \xrightarrow{p} F(x)$ for all x.

2. Different from the case of density estimation, there is no requirement that $nh_n \to \infty$ in the case of distribution estimation.

3. Given that $\{X_i\}_{i=1}^n$ is IID, the empirical distribution function $F_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leq x)$ is such that $E(F_n(x)) = F(x)$, $V(F_n(x)) = \frac{1}{n}F(x)(1-F(x))$ and using Lévy's central limit theorem $\sqrt{n}(F_n(x) - F(x)) \xrightarrow{d} N(0, F(x)(1-F(x))$. Hence, the use of $\hat{F}(x)$ as an estimator for $F(x)$ introduces a bias term compared to $F_n(x)$. However, since $V(F_n(x)) - V(\hat{F}(x)) = \frac{h_n}{n}f(x)\alpha - o\left(\frac{h_n}{n}\right)$, it is possible to obtain a reduction of the variance. Consequently, $E((\hat{F}(x) - F(x))^2)$ could be smaller than $E((F_n(x) - F(x))^2)$. An advantage of $\hat{F}(x)$ over $F_n(x)$ is its smoothness.

4. $E((\hat{F}(x) - F(x))^2) \leq \frac{F(x)(1-F(x))}{n} - \frac{h_n}{n}f(x)\alpha + o\left(\frac{h_n}{n}\right) + h_n^4 C$. Choosing h_n to minimize the bound on the righthand side of the inequality we obtain $h_n^* = n^{-1/3}\left(\frac{f(x)\alpha}{4C}\right)^{1/3}$. This rate of decay for h_n is different from that obtained for the estimation of f when $\beta \geq 2$ in Theorem 12.

Theorem 19. *Assume that a) $\{X_i\}_{i=1}^n$ is an IID sequence; b) K is a density function with $\int \gamma K(\gamma)d\gamma = 0$ and $\int \gamma^2 K(\gamma)d\gamma < \infty$; c) $F(x)$ has bounded second derivative on $\mathbb{R}$ and $F^{(1)}(x)$ is continuous on $\mathbb{R}$; d) $0 < h_n \to 0$ as $n \to \infty$. Then,*

$$\sqrt{n}(\hat{F}(x) - E(\hat{F}(x))) \xrightarrow{d} N(0, F(x)(1-F(x))). \tag{26}$$

Proof. $-(\hat{F}(x) - E(\hat{F}(x))) = \frac{1}{n}\sum_{i=1}^n \left(G\left(\frac{X_i - x}{h_n}\right) - E\left(G\left(\frac{X_i-x}{h_n}\right)\right)\right)$. Let $Z_{ni} = \frac{1}{n}G(\frac{X_i - x}{h_n})$, $E(Z_{ni}) = \mu_n$ and $s_n^2 = \sum_{i=1}^n E(Z_{ni} - \mu_n)^2$. Note that

$$s_n^2 \;=\; \sum_{i=1}^n \frac{1}{n^2} E\left(G\left(\frac{X_i - x}{h_n}\right) - E\left(G\left(\frac{X_i - x}{h_n}\right)\right)\right)^2 = \frac{1}{n}V\left(G\left(\frac{X_i - x}{h_n}\right)\right) = V\left(\hat{F}(x)\right).$$

Hence, $\frac{\hat{F}(x) - E(\hat{F}(x))}{\sqrt{V(\hat{F}(x))}} = \sum_{i=1}^n \left(\frac{Z_{ni} - \mu_n}{s_n}\right) = \sum_{i=1}^n X_{ni}$ with $E(X_{ni}) = 0$, $E(X_{ni}^2) = \frac{1}{s_n^2}E(Z_{ni} - \mu_n)^2$ and $\sum_{i=1}^n E(X_{in}^2) = 1$. Consequently, by Liapounov's CLT (see Davidson (1994), p. 372)

$$\frac{\hat{F}(x) - E(\hat{F}(x))}{\sqrt{V(\hat{F}(x))}} \xrightarrow{d} N(0,1)$$

provided $lim_{n\to\infty}\sum_{i=1}^{n} E|X_{ni}|^{2+\delta} = 0$ for some $\delta > 0$. We now verify that this limit is in fact zero.

$$\begin{aligned}\sum_{i=1}^{n} E|X_{ni}|^{2+\delta} &= \sum_{i=1}^{n} E\left(\left|\frac{Z_{ni}-\mu_n}{s_n}\right|^{2+\delta}\right) = \sum_{i=1}^{n} V(\hat{F}(x))^{-1-\frac{\delta}{2}} E\left(|Z_{ni}-\mu_n|^{2+\delta}\right) \\ &= V(\hat{F}(x))^{-(1+\frac{\delta}{2})} nE\left(|Z_{ni}-\mu_n|^{2+\delta}\right)\end{aligned}$$

By the c_r inequality and the fact that by Theorem 18 $\mu_n = O(n^{-1})$, we have $E\left(|Z_{ni}-\mu_n|^{2+\delta}\right) \leq 2^{1+\delta} E\left(|Z_{ni}|^{2+\delta}\right)$. Therefore,

$$\begin{aligned}\sum_{i=1}^{n} E|X_{ni}|^{2+\delta} &\leq V(\hat{F}(x))^{-(1+\frac{\delta}{2})} n 2^{1+\delta} E\left(|Z_{nt}|^{2+\delta}\right) \\ &= V(\hat{F}(x))^{-(1+\frac{\delta}{2})} 2^{1+\delta} n^{-1-\delta} E\left(\left|G\left(\frac{X_t - x}{h_n}\right)\right|^{2+\delta}\right) \\ &\leq V(\hat{F}(x))^{-(1+\frac{\delta}{2})} n^{-1-\delta} \text{ since } |G(x)| \leq 1 \text{ and } \int f(\gamma)d\gamma = 1 \\ &= (nV(\hat{F}(x)))^{-1-\frac{\delta}{2}} n^{-\delta/2}\end{aligned}$$

By Theorem 18 as $n \to \infty$, $nV(\hat{F}) \to F(x)(1-F(x))$ and since $n^{-\delta/2} \to 0$ for all $\delta > 0$ we have $\sum_{i=1}^{n} E|X_{ni}|^{2+\delta} = o(1)$. Since $V(\hat{F}(x)) - \frac{F(x)(1-F(x))}{n} = -\frac{h_n}{n} f(x)\alpha + o\left(\frac{h_n}{n}\right) = o(1)$ we have

$$\frac{\hat{F}(x) - E(\hat{F}(x))}{\sqrt{\frac{F(x)(1-F(x))}{n}}} \xrightarrow{d} N(0,1). \tag{27}$$

$\square$

Remark. $\sqrt{n}(\hat{F}(x) - F(x)) = \sqrt{n}(\hat{F}(x) - E(\hat{F}(x))) + \sqrt{n}(E(\hat{F}(x)) - F(x)) = \sqrt{n}(\hat{F}(x) - E(\hat{F}(x))) + n^{1/2}h_n^2 O(1)$, where the last equality follows from Theorem 18. Hence, if $n^{1/2}h_n^2 \to 0$ we have $\sqrt{n}(\hat{F}(x) - F(x)) \xrightarrow{d} N(0, F(x)(1-F(x)))$. If $h_n = h_n^*$ on remark 4 following Theorem 18 then $n^{1/2}(h_n^*)^2 = n^{-1/6}C \to 0$ as $n \to \infty$ and $\sqrt{n}(\hat{F}(x) - F(x)) \xrightarrow{d} N(0, F(x)(1-F(x)))$. Hence, under suitable choice for the order of the bandwidth, F_n and $\hat{F}$ have the same asymptotic distribution.

4 Estimation of discontinuous distributions and their jumps

We will now consider the estimation of distribution functions that are continuous, but not absolutely continuous, or even discontinuous at x. We first study the case where F is continuous but does not have a density at x. In what follows, since F is a distribution function, we take it to be right-continuous, L_1 denotes the space of functions on $\mathbb{R}$ with a finite norm $\|f\|_{L_1} = \int_{\mathbb{R}} |f(t)|dt$, C_b is the space of uniformly bounded continuous functions on $\mathbb{R}$ with norm $\|f\|_C = \sup_{t\in\mathbb{R}} |f(t)|$.

4.1 Estimation at points of continuity of F

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

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} U\left(\frac{X_i - x}{h}\right) \tag{28}$$

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. *$U \in C$, $\lim_{x\to-\infty} U(x) = 1$ 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 $\hat{F}(x) = \frac{1}{n}\sum_{i=1}^n L\left(\frac{x-X_i}{h}\right)$ where $L(x) = \int_{-\infty}^x K(t)dt$.

Letting

$$\phi_U(N) = \max\left\{\sup_{z<-N} |U(z) - 1|, \sup_{z>N} |U(z)|\right\}$$

for $N > 0$, we note that Assumption 1 implies $\phi_U(N) \to 0$ as $N \to \infty$. Also, letting $\omega(x,\epsilon) \equiv \sup_{0<h\leq\epsilon} (F(x+h) - F(x-h))$, we note that if x is a continuity point of F, then $\omega(x,\epsilon) \to 0$ as $\epsilon \to 0$. We now state the following lemma.

Lemma 3. *If U satisfies Assumption 1 and $\delta \in (0,1)$, then for all $h > 0$*

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

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 $\epsilon > 0$ we have

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

Then, we obtain the following bounds

$$\left|\int_{y<x-\epsilon}\left[U\left(\frac{y-x}{h}\right)-1\right]dF(y)\right| \le \sup_{z<-\epsilon/h}|U(z)-1|\int_{y<x-\epsilon}dF(y) \le \phi_U(\epsilon/h)F(x)$$

$$\left|\int_{|x-y|\le\epsilon}U\left(\frac{y-x}{h}\right)dF(y)\right| \le \|U\|_C\left[F(x+\epsilon)-F(x-\epsilon)\right] \le \|U\|_C\,\omega(x,\epsilon),$$

$$\left|\int_{y>x+\epsilon}U\left(\frac{y-x}{h}\right)dF(y)\right| \le \sup_{z>\epsilon/h}|U(z)|\int_{y>x+\epsilon}dF(y) \le \phi_U(\epsilon/h)(1-F(x)),$$

and $\int_{x-\epsilon}^{x}dF(y)\le\omega(x,\epsilon)$. Collecting the above inequalities we have

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

To finish the proof, we put $\epsilon=h^\delta$.

□

Lemma 3 is the basis for the following theorem.

Theorem 20. *Suppose U satisfies Assumption 1 and let x be a continuity point of F. Then,*

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

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

Proof. a) follows directly from Lemma 3 since

$$\left|E\left(\hat{F}(x)\right)-F(x)\right| = \left|\int_{\mathbb{R}}U\left(\frac{t-x}{h}\right)dF(t)-F(x)\right|.$$

b) We have $V(\hat{F}(x))=\frac{1}{n}\left\{E\left(U^2\left(\frac{X-x}{h}\right)\right)-\left[E\left(U\left(\frac{X-x}{h}\right)\right)\right]^2\right\}$. From part a) we know that $E\left(U\left(\frac{X-x}{h}\right)\right)=F(x)+o(1)$ as $h\to0$. U^2 satisfies Assumption 1, so by Lemma 3

$$\left|E\left(U^2\left(\frac{X-x}{h}\right)\right)-F(x)\right| \le \phi_{U^2}(h^{\delta-1})+\omega(x,h^\delta)(1+\left\|U^2\right\|_C)\to0,\ \ h\to0.$$

If $F(x)\in(0,1)$ this proves the statement. □

The following corollary is a direct consequence of Theorem 20.

Corollary 2. *If U satisfies Assumption 1 and $F\in C$, then*

$$\sup_{x\in\mathbb{R}}\omega(x,\epsilon)\to0,\ \ \textit{as } \epsilon\to0, \tag{29}$$

and $\sup_{x\in\mathbb{R}}\left|E(\hat{F}(x))-F(x)\right| \le \phi_U(h^{\delta-1})+\sup_{x\in\mathbb{R}}\omega(x,h^\delta)(1+\|U\|_C)\to0,\ \ h\to0.$

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 segments $[x - \epsilon, x + \epsilon]$ will be small for small ϵ and all $x \in [-N, N]$. Thus (29) is true and the statement follows from Theorem 20 a). □

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

Theorem 21. *Suppose U satisfies Assumption 1 and 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 $n \to \infty$, we have*

$$\sqrt{n}(\hat{F}(x) - F(x)) \xrightarrow{d} N(0, (1 - F(x))F(x)). \tag{30}$$

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 $\lim_{n\to\infty} \sum_{i=1}^n E\left(|X_{in}|^{2+\theta}\right) = 0$ for some $\theta > 0$ (see Davidson, 1994, p. 369-372). It follows from the fact that $\{X_i\}_{i=1,2,\cdots}$ is assumed to be 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 $E(|Z_{in} - \mu_n|^{2+\theta}) \le 2^{1+\theta}(E(|Z_{in}|^{2+\theta}) + |\mu_n|^{2+\theta})$. From Theorem 20, if $h \to 0$ we have that $n|\mu_n|^{2+\theta} = \frac{1}{n^{1+\theta}}|F(x) + o(1)|^{2+\theta}$. Consequently,

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

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 $\sqrt{n}(\hat{F}(x) - F(x)) = \sqrt{n}(\hat{F}(x) - E(\hat{F}(x))) +$

$\sqrt{n}(E(\hat{F}(x)) - F(x))$ we have that $\sqrt{n}(\hat{F}(x) - F(x)) \xrightarrow{d} N(0, (1 - F(x))F(x))$ provided that $\sqrt{n}(E(\hat{F}(x)) - F(x)) \to 0$. But from Theorem 20,

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

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. □

Remark. 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 compact support, then $\sqrt{n}\,\phi_U(h^{\delta-1}) \equiv 0$ for any $\delta \in (0, 1)$ and all $0 < h < h(\delta)$. If K has 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 $\alpha > 0$, we have that for some positive constant M_x and $u \in (0, h^\delta]$

$$F(x + u) - F(x - u) < 2M_x u^\alpha \leq 2M_x h^{\alpha\delta}$$

and $\sqrt{n}\,\omega(x, h^\delta) \leq 2M_x\sqrt{n}\,h^{\alpha\delta} \to 0$ 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 $F(x, y) := F(y) - F(x)$ for $x < y$. From Theorem 20 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 20. To obtain it we make the following assumption.

Assumption 2. *$G = G([a, b])$ is a bounded continuous function of intervals $[a, b] \subset \mathbb{R}$ or, more precisely, of an ordered pair (a, b), where $a < b$ with*

$$\lim_{b\to-\infty} G([a, b]) = 0, \quad \lim_{a\to-\infty,\ b\to\infty} G([a, b]) = 1 \text{ and } \lim_{a\to\infty} G([a, b]) = 0.$$

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. Besides, if $U_1,\ U_2$ satisfy Assumption 1, then $G([a, b]) = U_1(a) - U_2(b),\ -\infty < a < b < \infty$, 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 4. *If G satisfies Assumption 2 and $\delta \in (0, 1)$, then for all $h > 0$*

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

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

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

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

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

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

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

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

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

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

Theorem 22. *Suppose G satisfies Assumption 2, $x < y$ are continuity points of F and let $\delta \in (0,1)$. Then*

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

b) If $F(x,y) \in (0,1)$ *then* $V(\hat{F}(x,y)) = \frac{1}{n}[F(x,y)(1 - F(x,y)) + o(1)]$.

Proof. Part a) immediately follows from Lemma 4 because

$$E\left(\hat{F}(x,y)\right) = \int_{\mathbb{R}} G\left(\left[\frac{z-y}{h}, \frac{z-x}{h}\right]\right) dF(z).$$

Part b) 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 4,

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

as $h \to 0$. □

5 Estimation of jumps of a distribution function

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

$$\hat{p}_x = \frac{1}{n} \sum_{i=1}^{n} W\left(\frac{X_i - x}{h}\right), \tag{31}$$

where the function W satisfies the following assumption.

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

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

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

Lemma 5. *a) Let $\epsilon_1 \in (0,1)$, $\epsilon_2 > \epsilon_1$. For any $\lambda \in (0,1)$*

$$\left| \int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x \right| \le \omega_W(\lambda^{\epsilon_2 - \epsilon_1}) \left[p_x + \delta_F(\lambda^{1-\epsilon_1})\right] + (1 + \|W\|_C)\, \delta_F(\lambda^{1-\epsilon_1}) + \phi_W(\lambda^{-\epsilon_1}).$$

b) 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 $\lambda > 0$

$$\left| \int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x \right| \le \phi_W(\lambda^{-\epsilon_1}).$$

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

Proof. a) By additivity of probability for any $h_1 \in (0,h)$

$$p_x \le \int_{|z-x|<h_1} dF(z) \le \int_{|z-x|<h} dF(z) = p_x + \delta_F(h). \tag{32}$$

This implies

$$\int_{h_1 \le |z-x| < h} dF(z) = \int_{|z-x|<h} dF(z) - \int_{|z-x|<h_1} dF(z) \le \delta_F(h), \tag{33}$$

$$\left|\int_{|z-x|<h_1} dF(z) - p_x\right| \leq \delta_F(h). \tag{34}$$

We wish to bound

$$\int_{\mathbb{R}} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x = \int_{|z-x|<h_1} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x + \int_{h_1\leq|z-x|<h} W\left(\frac{z-x}{\lambda}\right) dF(z) + \int_{|z-x|\geq h} W\left(\frac{z-x}{\lambda}\right) dF(z). \tag{35}$$

Obviously,

$$\left|\int_{|z-x|\geq h} W\left(\frac{z-x}{\lambda}\right) dF(z)\right| \leq \phi_W(h/\lambda). \tag{36}$$

By (32) and (34)

$$\begin{aligned}
\left|\int_{|z-x|<h_1} W\left(\frac{z-x}{\lambda}\right) dF(z) - p_x\right| &\leq \left|\int_{|z-x|<h_1}\left[W\left(\frac{z-x}{\lambda}\right) - W(0)\right] dF(z)\right| \\
&+ \left|W(0)\int_{|z-x|<h_1} dF(z) - p_x\right| \\
&\leq \omega_W(h_1/\lambda)\int_{|z-x|<h_1} dF(z) + \left|\int_{|z-x|<h_1} dF(z) - p_x\right| \\
&\leq \omega_W(h_1/\lambda)\left[p_x + \delta_F(h)\right] + \delta_F(h). \qquad (37)
\end{aligned}$$

By (33)

$$\left|\int_{h_1\leq|z-x|<h} W\left(\frac{z-x}{\lambda}\right) dF(z)\right| \leq \|W\|_C \int_{h_1\leq|z-x|<h} dF(z) \leq \|W\|_C\, \delta_F(h). \tag{38}$$

Combining (36)-(38) we get

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

It remains to put here $h = \lambda^{1-\epsilon_1}$, $h_1 = h\lambda^{\epsilon_2} = \lambda^{1+\epsilon_2-\epsilon_1} < h$ to finish the proof.
b) Instead of (35) 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 (36) we see that the statement in the enunciation of the lemma is true. □

Theorem 23. *Suppose W satisfies Assumption 3.*

a) Let $\epsilon_1 \in (0,1)$, $\epsilon_2 > \epsilon_1$. Then, for any $h \in (0,1)$,

$$|E(\hat{p}_x) - p_x| \leq \omega_W(h^{\epsilon_2-\epsilon_1})\left[p_x + \delta_F(h^{1-\epsilon_1})\right] + (1 + \|W\|_C)\,\delta_F(h^{1-\epsilon_1}) + \phi_W(h^{-\epsilon_1}).$$

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

$$|E(\hat{p}_x) - p_x| \leq \phi_W(h^{-\epsilon_1}).$$

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

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

Proof. a) and b) follow from Lemma 5. c) obtains from a), given that

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

and the fact that W^2 satisfies Assumption 3. $\square$

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

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

Using this equation in (37) we get

$$\begin{aligned}
\left|\int_{|z-x|<h_1} \tilde{K}\left(\frac{z-x}{\lambda}\right) dF(z) - p_x\right| \quad &\leq \quad \frac{\sup\left|\tilde{K}^{(m)}\right|}{m!}\left(\frac{h_1}{\lambda}\right)^m \left|\int_{|z-x|<h_1} dF(z)\right| \\
&+ \quad \left|\int_{|z-x|<h_1} dF(z) - p_x\right| \\
&\leq \quad c\lambda^{m(\epsilon_2-\epsilon_1)}\left[p_x + \delta_F(\lambda^{1-\epsilon_1})\right] + \delta_F(\lambda^{1-\epsilon_1})
\end{aligned}$$

with the same choice of h_1, h as before. The resulting bound

$$|E(\hat{p}_x) - p_x| \quad \leq \quad c\lambda^{m(\epsilon_2-\epsilon_1)}\left[p_x + \delta_F(\lambda^{1-\epsilon_1})\right] + \left(1 + \|K\|_{L_1}\right)\delta_F(\lambda^{1-\epsilon_1}) + \sup_{|z|\geq\lambda^{-\epsilon_1}}\left|\tilde{K}(z)\right|$$

shows that imposing higher-order kernel conditions, without restricting the class of distribution functions, has a limited effect on convergence rate, because of the term $\delta_F(\lambda^{1-\epsilon_1})$. The following theorem gives asymptotic normality of $\hat{p}_x$ when suitably centered and normalized.

Theorem 24. *Suppose W satisfies Assumption 3 and F has a jump p_x at x where $p_x \in (0,1)$. For $\epsilon_1 \in (0,1)$ and $\epsilon_2 > \epsilon_1$, if $h \to 0$, $\sqrt{n}\,\phi_W(h^{-\epsilon_1}) \to 0$, $\sqrt{n}\,\omega_W(h^{\epsilon_2-\epsilon_1}) \to 0$ and $\sqrt{n}\delta_F(h^{1-\epsilon_1}) \to 0$ as $n \to \infty$, we have*

$$\sqrt{n}(\hat{p}_x - p_x) \xrightarrow{d} N(0, (1-p_x)p_x). \tag{39}$$

Proof. In the proof of Theorem 21, 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 23 to obtain

$$|\sqrt{n}(E(\hat{p}_x) - p_x)| \leq \sqrt{n}\left(\omega_W(h^{\epsilon_2-\epsilon_1})\left[p_x + \delta_F(h^{1-\epsilon_1})\right] + (1 + \|W\|_C)\,\delta_F(h^{1-\epsilon_1}) + \phi_W(h^{-\epsilon_1})\right).$$

Hence, given the conditions on the enunciation of the theorem we have $\sqrt{n}(E(\hat{p}_x) - p_x) \to 0$ as $n \to \infty$.

$\square$

Remark. The requirements that $\sqrt{n}\,\phi_W(h^{-\epsilon_1}) \to 0$, $\sqrt{n}\,\omega_W(h^{\epsilon_2-\epsilon_1}) \to 0$ as $n \to \infty$ always hold with appropriately chosen $\{h_n\}$. If $W(t) = \int_{\mathbb{R}} e^{itz}K(z)dz$ and K is a kernel such that W has compact support (see, e.g., W and K used in subsection 5.2), then $\sqrt{n}\,\phi_W(h^{-\epsilon_1}) \equiv 0$ for all small h. The condition that $\sqrt{n}\delta_F(h^{1-\epsilon_1}) \to 0$ can always be satisfied if h_n tends to 0 sufficiently quickly. However, to obtain a tractable dependence of the bandwidth on n, we note that

$$\delta_F(h^{1-\epsilon_1}) = F(x + h^{1-\epsilon_1}) - \left(F(x - h^{1-\epsilon_1}) + p_x\right)$$

and if a Lipschitz condition of order α is imposed on the difference on the righthand side of the equality, we have $\sqrt{n}\delta_F(h^{1-\epsilon_1}) < 2M_x\sqrt{n}h^{(1-\epsilon_1)\alpha})$. As in Remark 1, if $h \propto n^{-\gamma}$, it suffices to choose $\gamma > 1/(2(1-\epsilon_1)\alpha)$ to satisfy the condition.