On the equivalence of weak and strong convergences of densities[1,2]

Kairat Mynbaev

International School of Economics

Tolebi 59

Kazakh-British Technical University

Almaty 050000, Kazakhstan

email: k_mynbayev@ise.ac

and

Carlos Martins-Filho

Department of Economics

UCB 256

University of Colorado

Boulder, CO 80309-0256, USA

email: carlos.martins@colorado.edu

July, 2026

Abstract. We provide conditions on a sequence of densities $\{f_n\}_{n\in\mathbb{N}}$ and on a density f to assure that weak convergence of $\{f_n\}_{n\in\mathbb{N}}$ to f implies strong convergence of $\{f_n\}_{n\in\mathbb{N}}$ to f. For bounded and continuous f, Walker (2023) provided a new sufficient condition for the equivalence of these two types of convergence that involves the second derivatives of f_n. We follow Walker's approach but replace his condition on $\{f_n\}_{n\in\mathbb{N}}$ by L_1-equicontinuity. In addition, we provide a correct restatement and refined proof of his main supporting lemma.

Keywords and phrases. Weak convergence of densities, strong convergence of densities, kernel density estimation.

MSC Classification. 60F25, 60F99, 62G07.

[1]We thank an anonymous referee for excellent comments.

[2]This research is partially funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No.AP19676673).

1 Introduction

The relationship between different types of convergence of a sequence of densities $\{f_n\}_{n\in\mathbb{N}}$ to a density f has been the subject of several papers dating back to the 1970's. See, *inter alia*, Hettmansperger and Klimko (1974), Dellacherie and Meyer (1978), Visintin (1984), Balder (1991) and Balder (1993). Of particular interest are conditions that assure the equivalence of strong and weak convergences, represented respectively by the L_1 and the Lévy-Prokhorov metrics.

Several sets of sufficient conditions for this equivalence have been suggested. Hettmansperger and Klimko (1974) imposed unimodality on f_n, Dellacherie and Meyer (1978) required $\liminf_{n} (x)$ $\geq f(x)$ almost everywhere, whereas Visintin (1984) established equivalence when f is in the set of all extreme points of the closure of the convex hull of $\{f_n\}_{n\in\mathbb{N}}$ almost everywhere. In a recent paper, Walker (2023) gave a new sufficient condition involving second derivatives of f_n. He noticed that the convergence of $\{f_n\}_{n\in\mathbb{N}}$ to f in L_1 is equivalent to the convergence of a smoothed version of $\{f_n\}_{n\in\mathbb{N}}$ to f in L_1 if f is bounded and continuous.

We follow Walker's approach but replace his condition on $\{f_n\}_{n\in\mathbb{N}}$. In particular, our proof does not require differentiability of f_n nor does it require that f_n have a second-derivative satisfying a $\limsup$ condition on his Theorem 3.2. Our condition on $\{f_n\}_{n\in\mathbb{N}}$ is simply L_1-equicontinuity which is strictly weaker than a local Hölder condition (see Remark 6 below). In fact, given the Kolmogorov–Riesz theorem, L_1-equicontinuity is precisely the missing ingredient for L_1 pre-compactness of a bounded equitight family; for densities converging weakly, both boundedness ($\|f_n\|_1 = 1$) and equitightness are automatic, making L_1-equicontinuity the natural condition to be imposed.

One of our auxiliary results (see Theorem 1 below), needed for our main equivalence theorem, is of independent interest, establishing the asymptotic unbiasedness of the classical Rosenblatt-Parzen density estimator for Lebesgue points of the density f with a bound on the rate of convergence. We also provide a corrected statement of Walker's main lemma, with a refined proof.

The rest of the paper is organized as follows: section 2 contains some technical background and auxiliary lemmas and section 3 contains the main equivalence theorem. Throughout we consider densities on $\mathbb{R}$ with respect to Lebesgue measure and adopt the following notation: a) for an arbitrary function $f : \mathbb{R} \to \mathbb{R}$, we say that f is locally integrable, denoted by $f \in L_1^{loc}(\mathbb{R})$ if $\int_{\mathcal{K}} |f(x)|dx < \infty$ for all compact $\mathcal{K} \subset \mathbb{R}$; b) $\operatorname{supp} f$ denotes the support of the function f; c) χ_A denotes the indicator function of set A; d) $\|f\|_1$ denotes the L_1 norm of the function f; e) for a uniformly bounded $f : D \to \mathbb{R}$, $\|f\|_\infty := \sup\limits_{x \in D} |f(x)|$.

2 Technical preliminaries

We start by noting that by Theorem 5.6.2 in Bogachev (2007), for any function f such that $f \in L_1^{loc}(\mathbb{R})$ we have that $f(x) = \lim\limits_{h \downarrow 0} \frac{1}{2h} \int_{x-h}^{x+h} f(t)dt$ for almost every $x \in \mathbb{R}$. In this case, and when $|f(x)| < \infty$, x is called a Lebesgue point of f. Letting $g_x(t) = \frac{f(x+t)+f(x-t)}{2} - f(x)$, it is easy to see that if x is a Lebesgue point of f then $\lim\limits_{h \downarrow 0} \frac{1}{h} \int_0^h g_x(t)dt = 0$. Letting $\gamma_x(\varepsilon) = \sup\limits_{0<h\le\varepsilon} \frac{1}{h} \left| \int_0^h g_x(t)dt \right|$, we have that

$$\lim_{\varepsilon \to 0} \gamma_x(\varepsilon) = 0 \text{ for almost every } x \in \mathbb{R}. \tag{1}$$

The condition given in equation (1) holds at points of continuity of f, denoted by $C(f)$, since

$$\frac{1}{h} \left| \int_0^h g_x(t)dt \right| \le \sup_{0<|t|\le h} |f(x+t) - f(x)| \to 0, \text{ as } h \downarrow 0 \text{ for } x \in C(f).$$

However, $x \notin C(f)$ may be a Lebesgue point of f. To see this, note that for the Heaviside function H, 0 is a Lebesgue point if we define $H(0) = 1/2 = [H(0-0) + H(0+0)]/2$ and also a point of discontinuity of H.

Some elements of the proofs of the following Lemmas 1 and 2 below have been borrowed from Bochner and Chandrasekharan (1949).

Lemma 1. *Suppose that* $0 \le K \in L_1$ *is an even function, non-increasing on* $[0, \infty)$ *and* $\int_{\mathbb{R}} K(t)dt = c > 0$. *Then,*

$$\lim_{t\to\infty} tK(t) = \lim_{t\to\infty} K(t) = 0, \; tK(t) \le c/2 \textit{ for } t > 0. \tag{2}$$

Proof. $\lim_{t\to\infty} K(t) = 0$ because of the convergence of $\int_{\mathbb{R}} K(t)dt$. Obviously,

$$tK(t) = \int_0^t K(s)\,ds - \int_0^t (K(s) - K(t))\,ds \le \frac{c}{2} - \int_0^\infty (K(s) - K(t))\,\chi_{(0,t)}(s)\,ds \to 0,\ t \to \infty,$$

by the dominated convergence theorem. The bound on $tK(t)$ in (2) is trivial. $\square$

If K has a compact support, where $\operatorname{supp} K = [-a, a]$, then we define $N(\varepsilon) = a/\varepsilon$, $\varepsilon > 0$. If the support of K is not compact, from (2) we know that for each $\varepsilon > 0$ the set $\{N > 0 : N\varepsilon K(N\varepsilon) \le 2\varepsilon^2,\ \int_{\varepsilon N}^\infty K(t)dt \le \varepsilon\}$ is not empty and we can put

$$N(\varepsilon) = \inf\left\{N > 0 : N\varepsilon K(N\varepsilon) \le 2\varepsilon^2,\ \int_{\varepsilon N}^\infty K(t)dt \le \varepsilon\right\}.$$

In this case, if K is right-continuous,

$$N(\varepsilon)K(N(\varepsilon)\varepsilon) \le 2\varepsilon,\ \int_{\varepsilon N(\varepsilon)}^\infty K(t)dt \le \varepsilon. \tag{3}$$

From the second inequality in (3) we see that $\lim\limits_{\varepsilon\to 0} N(\varepsilon) = \infty$.

Lemma 2. *Suppose that K satisfies conditions of Lemma 1 and, in addition, K is right-continuous with derivative K' that exists everywhere on $[0, \infty)$ and is Riemann integrable on every compact interval $[0, b]$, $b > 0$. Let $f \in L_1$ and x be a Lebesgue point of f. Put $I_x(N) = \int_0^\infty g_x(t)NK(Nt)dt$. 1) If the support of K is not compact, then*

$$|I_x(N(\varepsilon))| \le c\gamma_x(\varepsilon) + \varepsilon(\|f\|_1 + |f(x)|),\ \varepsilon > 0. \tag{4}$$

2) If $\operatorname{supp} K = [-a, a]$, then

$$|I_x(N(\varepsilon))| \le c\gamma_x(\varepsilon),\ \varepsilon > 0. \tag{5}$$

Proof. 1) Split the integral $I_x(N(\varepsilon)) = \left(\int_0^\varepsilon + \int_\varepsilon^\infty\right) g_{x(t)}N(\varepsilon)K(N(\varepsilon)t)dt := I_1 + I_2$. Until the end of this proof we write $N := N(\varepsilon)$ for short. Put $G(t) = \int_0^t g_x(s)ds$. g_x is locally integrable on $[0, \infty)$ and G is absolutely continuous. Since G is absolutely continuous and K is monotone in $[0,\infty)$, K is of bounded variation and by Theorem 1 in (Natanson, 1955, Chapter VIII, p. 230), $\int_0^\varepsilon G(t)\, d(NK(Nt))$ exists. In addition, by Theorem 3.3.1 in Hille and Phillips (1974), $\int_0^\varepsilon NK(Nt)dG(t)$ also exists and integrating by parts we have

$$I_1 = \int_0^\varepsilon NK(Nt)dG(t) = NK(N\varepsilon)\, G(\varepsilon) - \int_0^\varepsilon G(t)\, d(NK(Nt)).$$

In addition, since K has a derivative K' everywhere on $[0,\infty)$ and K' is Riemann-integrable on every compact interval $[0,b]$ for $b>0$, by Theorem 2 in (Natanson, 1955, Chapter VIII, p. 231),

$$\int_0^\varepsilon G(t)\, d(NK(Nt)) = \int_0^\varepsilon G(t)\, N^2 K'(Nt)\, dt,$$

and $I_1 = NK(N\varepsilon)G(\varepsilon) - \int_0^\varepsilon G(t)N^2K'(Nt)dt$.

Using (2) for the first term we have $\left| N\varepsilon K(N\varepsilon)\frac{G(\varepsilon)}{\varepsilon}\right| \le \frac{c}{2}\gamma_x(\varepsilon)$. Replacing $Nt=s$ and remembering that $K'(s)\le 0$ we bound the second term as

$$\begin{aligned}\left|\int_0^\varepsilon G(t)N^2K'(Nt)dt\right| &\le \gamma_x(\varepsilon)\int_0^\varepsilon tN^2|K'(Nt)|dt = -\gamma_x(\varepsilon)\int_0^{N\varepsilon} sK'(s)ds \\ &= \gamma_x(\varepsilon)\left(-sK(s)|_0^{N\varepsilon} + \int_0^{N\varepsilon} K(s)ds\right) \le \frac{c}{2}\gamma_x(\varepsilon).\end{aligned}$$

The integration by parts is validated as above. Thus,

$$|I_1| \le c\gamma_x(\varepsilon). \tag{6}$$

For I_2 we split the terms in g_x and use monotonicity:

$$\begin{aligned}|I_2| &\le \int_\varepsilon^\infty \left|\frac{f(x+t)}{2}\right| NK(Nt)dt + \int_\varepsilon^\infty \left|\frac{f(x-t)}{2}\right| NK(Nt)dt + |f(x)|\int_\varepsilon^\infty NK(Nt)dt \\ &\le \frac{NK(N\varepsilon)}{2}\int_{x+\varepsilon}^\infty |f(t)|dt + \frac{NK(N\varepsilon)}{2}\int_{-\infty}^{x-\varepsilon} |f(t)|dt + |f(x)|\int_{N\varepsilon}^\infty K(s)ds.\end{aligned}$$

Applying (3) we get $|I_2| \le \varepsilon(\|f\|_1 + |f(x)|)$, which together with (6) proves the statement.

2) If K has compact support, then $t>\varepsilon$ implies $N(\varepsilon)t > N(\varepsilon)\varepsilon = a$ and $K(N(\varepsilon)t)=0$, so $I_2=0$

and it is enough to apply (6). $\square$

We note that if $\int_{\mathbb{R}} K(t)dt = 1$ then

$$\int_{\mathbb{R}} K(z) f(x+hz)\,dz - f(x) = \frac{1}{h}\int_{\mathbb{R}} K\left(\frac{s}{h}\right) g_x(s)ds = \frac{2}{h}\int_0^{\infty} K\left(\frac{s}{h}\right) g_x(s)ds. \tag{7}$$

For our applications to weak convergence and to density estimation we add to the conditions of Lemma 2 the condition that K integrates to 1, thus obtaining the following condition on K:

$$\begin{aligned}&0 \le K \in L_1,\ \int_{\mathbb{R}} K(t)dt = 1,\ K \text{ is even, right-continuous, non-increasing and its derivative } K' \\ &\text{exists everywhere on } [0,\infty) \text{ and is Riemann integrable on every interval } [0,b],\ b>0.\end{aligned} \tag{8}$$

Non-negativity of K excludes higher-order kernels that are often used to obtain faster convergence rates for the absolute value of the bias of the Rosenblatt-Parzen kernel estimator (see equation (9)) when the density is sufficiently smooth. Condition (8) addresses the case where the density is not sufficiently smooth or the degree of smoothness is unknown.

Define $h(\varepsilon) = 1/N(\varepsilon)$, $\varepsilon > 0$, implying $\lim_{\varepsilon\to 0} h(\varepsilon) = 0$ and consider a sequence of independent and identically distributed random variables $\{X_i\}_{i=1}^n$ with density f. The Rosenblatt-Parzen kernel estimator of $f(x)$, denoted by $\hat{f}(x)$, is given by

$$\hat{f}(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right) \text{ for } h > 0, \tag{9}$$

and since K is even, its expectation, or smoothed version, is

$$E\hat{f}(x) := S_h f(x) = \frac{1}{h}\int_{\mathbb{R}} K\left(\frac{x-z}{h}\right) f(z)\,dz = \int_{\mathbb{R}} K(z) f(x+hz)\,dz. \tag{10}$$

Theorem 1. *Let x be a Lebesgue point of a density f and assume that the kernel K satisfies* (8)*. We choose the smoothing parameter $h := h(\varepsilon)$. Then, we have:*

1. *if* supp K *is not compact,* $\left|E\hat{f}(x) - f(x)\right| \le 2\left(\gamma_x(\varepsilon) + \varepsilon(1+f(x))\right),\ \varepsilon > 0,$

2. if $\operatorname{supp} K = [-a, a]$, $\left|E\hat{f}(x) - f(x)\right| \leq 2\gamma_x(\varepsilon)$, $\varepsilon > 0$.

Proof. The proof follows directly by applying equation (7) and Lemma 2. $\square$

3 Main result

For a distribution function F and corresponding density f let $F(A) = \int_A dF = \int_A f dx$ where $A \subset \mathbb{R}$ is a Borel set. A sequence of densities $\{f_n\}_{n\in\mathbb{N}}$ is said to converge weakly to a density f if $\int_{\mathbb{R}} k(x) f_n(x) dx \to \int_{\mathbb{R}} k(x) f(x) dx$ for all bounded and continuous functions k. Let $A^{\varepsilon} = \bigcup_{a\in A} \{x : |x - a| < \varepsilon\}$ be an ε-neighborhood of A. Given distribution functions F_n, F corresponding to f_n, f the Lévy-Prokhorov metric is defined by

$$d(f_n, f) = \inf\left\{\varepsilon > 0 : \sup_{Borel\ A} \{F_n(A) - F(A^{\varepsilon})\} \leq \varepsilon\right\},$$

and an equivalent definition of weak convergence is that $d(f_n, f) \to 0$. See Billingsley (1999).

The next result is well-known (Devroye and Györfi, 1985, Theorem 1), but here we provide an alternative proof.

Corollary 1. *Let the kernel K satisfy* (8)*. For a density f, we have* $\|S_{h_n} f - f\|_1 := \int_{\mathbb{R}} |S_{h_n} f - f| dx \to 0$*, as* $h_n \to 0$.

Proof. Since $\gamma_x(\varepsilon) \to 0$, as $\varepsilon \to 0$ and $f(x) < \infty$ for almost all x, by Theorem 1 we have that $S_{h_n} f(x) \to f(x)$ almost everywhere. Then, the conclusion of the theorem follows by Scheffé's theorem. $\square$

For completeness, we prove the following lemma and note that the commonly cited reference (Dudley, 2002, Theorem 11.3.3) does not have dependence on a Lipschitz constant.

Lemma 3. *For any two probability measures P and Q on the Borel sets of $\mathbb{R}$, let*

$$d_P(P, Q) := \inf\left\{\varepsilon > 0 : \sup_{Borel\ A} \{P(A) - Q(A^{\varepsilon})\} \leq \varepsilon\right\}.$$

If $d_P(P, Q) \leq \delta$ and $g : \mathbb{R} \to [0, 1]$ has Lipschitz constant L, then $\left|\int g\, dP - \int g\, dQ\right| \leq (L + 1)\delta$.

Proof. For any $0 \le g \le 1$, define the level sets $A_t = \{g > t\}$ for $t \in (0,1)$. Then $\int g\, dP = \int_0^1 P(A_t)\, dt$. By the Lévy–Prokhorov inequality, $P(A_t) \le Q(A_t^\delta) + \delta$. Now, $A_t^\delta = \{y : \exists\, z \in A_t,\ |y - z| < \delta\} \subseteq \{y : g(y) > t - L\delta\} = A_{t-L\delta}$, the inclusion following because g is L-Lipschitz: if $g(z) > t$ and $|y - z| < \delta$, then $g(y) > t - L\delta$. Therefore, $\int g\, dP = \int_0^1 P(A_t)\, dt \le \int_0^1 \big[Q(A_{t-L\delta}) + \delta\big]\, dt = \int_{-L\delta}^{1-L\delta} Q(A_s)\, ds + \delta$. Since $Q(A_s) \le 1$ for all s and $A_s = \mathbb{R}$ for $s < 0$ (as $g \ge 0$), $\int_{-L\delta}^{1-L\delta} Q(A_s)\, ds \le \int_{-L\delta}^{0} 1\, ds + \int_0^1 Q(A_s)\, ds = L\delta + \int g\, dQ$. Hence $\int g\, dP \le \int g\, dQ + (L+1)\delta$. By symmetry (exchanging P and Q), $|\int g\, dP - \int g\, dQ| \le (L+1)\delta$. □

The next lemma is a new correct statement and proof of Lemma 2.1 in Walker (2023).

Lemma 4 (Walker, restated)**.** *Let $S_{h_n} f_n(x) := \int_{\mathbb{R}} K(z) f_n(x + h_n z) dz$ and $S_{h_n} f(x)$ be similarly defined, where K satisfies* (8) *and is bounded with Lipschitz constant L_K. If $d_P(F_n, F) \to 0$ as $n \to \infty$ and f is bounded and continuous, then for some $h_n \to 0$, $\|S_{h_n} f_n - S_{h_n} f\|_1 \to 0$ as $n \to \infty$.*

Proof. Set $\varepsilon_n = d_P(F_n, F)$. First, note that if $\varepsilon_n = 0$ for some n, then $f_n = f$ almost everywhere and $S_{h_n} f_n = S_{h_n} f$. Hence, it suffices to consider that case where $\varepsilon_n > 0$. Let $h_n = \varepsilon_n^{1/3}$, so $h_n \to 0$ and $\varepsilon_n / h_n^2 = \varepsilon_n^{1/3} \to 0$. We also write $\delta_n = \varepsilon_n / h_n$.

Step 1: Rescaling. Fix $x \in \mathbb{R}$. Define the shifted-and-rescaled densities $f_n^*(y) = h_n\, f_n(y h_n + x)$, $f^*(y) = h_n\, f(y h_n + x)$. Note that f_n^* and f^* are densities since $f_n^*, f^* \ge 0$, $\int f_n^*(y)\, dy = \int f_n(u)\, du = 1$ and $\int f^*(y)\, dy = \int f(u)\, du = 1$. Their distribution functions are $F_n^*(t) = F_n(t h_n + x)$, $F^*(t) = F(t h_n + x)$.

Step 2: Lévy–Prokhorov rescaling. Since $d_P(F_n, F) \le \varepsilon_n$, the definition gives $F_n(A) \le F(A^{\varepsilon_n}) + \varepsilon_n$ for all Borel A. For any Borel $B \subseteq \mathbb{R}$, the affine transformation of B is such that $F_n^*(B) = F_n(h_n B + x)$. Note that the ε_n-neighborhood of $h_n B + x$ is by definition, $(h_n B + x)^{\varepsilon_n} = \big\{u : \exists\, b \in B,\ |u - h_n b - x| < \varepsilon_n\big\}$. Substituting $v = (u - x)/h_n$, the condition becomes $|v - b| < \varepsilon_n / h_n = \delta_n$, so $v \in B^{\delta_n}$ and $u \in h_n B^{\delta_n} + x$. Therefore, $F_n^*(B) = F_n(h_n B + x) \le F\big((h_n B + x)^{\varepsilon_n}\big) + \varepsilon_n = F\big(h_n B^{\delta_n} + x\big) + \varepsilon_n = F^*\big(B^{\delta_n}\big) + \varepsilon_n$. Since $\varepsilon_n \le \delta_n$ (for $h_n \le 1$), this gives

$$d_P(F_n^*, F^*) \le \delta_n = \varepsilon_n / h_n. \tag{11}$$

Step 3: Bounded Lipschitz inequality. Apply Lemma 3 to the rescaled distributions. The function K has Lipschitz constant L_K and $\|K\|_\infty < \infty$. Define $g(y) = \big(K(y) + \|K\|_\infty\big)/\big(2\|K\|_\infty\big)$, which satisfies $0 \leq g \leq 1$ with Lipschitz constant $L_g = L_K/(2\|K\|_\infty)$. Since f_n^* and f^* are the densities of F_n^* and F^*, $\big|\int g(y)\, f_n^*(y)\, dy - \int g(y)\, f^*(y)\, dy\big| \leq (L_g + 1)\delta_n$. Unscaling g back to K:

$$\left|\int K(y)\big[f_n^*(y) - f^*(y)\big]\, dy\right| \leq 2\|K\|_\infty(L_g + 1)\delta_n = (L_K + 2\|K\|_\infty)\delta_n =: C_K\, \delta_n. \tag{12}$$

Step 4: From the rescaled integral back to $S_{h_n} f_n - S_{h_n} f$. By the substitution $y = (u-x)/h_n$, $S_{h_n} f_n(x) = \int K(s)\, f_n(x + sh_n)\, ds = \frac{1}{h_n}\int K(y)\, f_n^*(y)\, dy$. Similarly for $S_{h_n} f(x)$. Hence, from (12),

$$|S_{h_n} f_n(x) - S_{h_n} f(x)| = \frac{1}{h_n}\left|\int K(y)\big[f_n^*(y) - f^*(y)\big]\, dy\right| \leq \frac{C_K\, \delta_n}{h_n} = C_K\, \frac{\varepsilon_n}{h_n^2} \to 0 \tag{13}$$

for every $x \in \mathbb{R}$. The constant C_K depends only on K, not on x.

Step 5: L_1 convergence via Scheffé's theorem. We cannot apply Scheffé directly to $S_{h_n} f_n$ and $S_{h_n} f$, because $S_{h_n} f$ itself depends on n. Instead, use the triangle inequality:

$$\|S_{h_n} f_n - S_{h_n} f\|_1 \leq \|S_{h_n} f_n - f\|_1 + \|S_{h_n} f - f\|_1. \tag{14}$$

Bound on $\|S_{h_n} f - f\|_1$. Since f is bounded and continuous, Theorem 1 gives $S_{h_n} f(x) \to f(x)$ for every x. Both $S_{h_n} f$ and f are nonnegative and integrate to 1 (they are densities), so Scheffé's theorem gives $\|S_{h_n} f - f\|_1 \to 0$.

Bound on $\|S_{h_n} f_n - f\|_1$. Inequality (13) gives $S_{h_n} f_n(x) - S_{h_n} f(x) \to 0$ for every x, and Theorem 1 gives $S_{h_n} f(x) \to f(x)$. Hence $S_{h_n} f_n(x) \to f(x)$ for every x. Since $S_{h_n} f_n$ and f are both densities, Scheffé's theorem gives $\|S_{h_n} f_n - f\|_1 \to 0$. Combining via (14) gives $\|S_{h_n} f_n - S_{h_n} f\|_1 \to 0$. □

Remark 1 (Dependence on the boundedness and continuity of f)**.** *Walker states the condition "f bounded and continuous" only after Lemma 2.1, in Lemma 2.2 and in the phrase "we will make this assumption going forward." However, this condition is needed* inside *the proof of Lemma 2.1 at Step 5: specifically, to obtain a fixed (i.e., n-independent) limit for Scheffé's theorem. Without* *it, the pointwise convergence $S_{h_n} f_n(x) - S_{h_n} f(x) \to 0$ between two n-dependent densities does not automatically yield L_1 convergence. The statement of Lemma 2.1 should therefore include the hypothesis that f is bounded and continuous.*

Remark 2. *When K is the uniform kernel $K(s) = \frac{1}{2}\chi_{(-1,1)}(s)$, the function g is not Lipschitz, so the bounded-Lipschitz inequality of Step 3 does not apply directly. This is why Walker treats the uniform kernel separately, using the identity $S_{h_n} f_n(x) = \frac{1}{2h_n}[F_n(x + h_n) - F_n(x - h_n)]$ and the two-sided Lévy–Prokhorov bound $F(x-\varepsilon_n) - \varepsilon_n \le F_n(x) \le F(x+\varepsilon_n) + \varepsilon_n$ to obtain the rate ε_n/h_n.*

Remark 3. *For a non-negative and non-decreasing function φ on $[0, \infty)$ and two distribution functions F and G, Rotar (1975) introduces the metric $\nu(F, G) = \int_{-\infty}^{\infty} \varphi(|x|)|F(x) - G(x)|\, dx$ and proves that under a matched-moment condition*

$$\int_0^{\infty} \varphi(x)\, F^*(x)\, dx = \int_0^{\infty} \varphi(x)\, G^*(x)\, dx, \tag{15}$$

where $F^*(x) = 1 - F(x) + F(-x)$, the ν-distance is controlled by the Lévy distance $L(F, G)$ in that for all $y > 0$, $\nu(F, G) \le 4\left\{\int_y^{\infty} \varphi(x)\, G^*(x)\, dx + \varphi(y)(y + 1)\, L(F, G)\right\}$. Taking $\varphi(x) = x$ (Rotar's Corollary), this gives a quantitative bound on $\int_{-\infty}^{\infty} |x||F(x) - G(x)|\, dx$ in terms of the Lévy distance and the tail integral $\int_{|x|\ge a} x^2\, dG$, for $a \ge 1$.

In our context, Rotar's lemma provides an alternative to the bounded-Lipschitz inequality of Step 3. For the CLT application in Rotar's paper, the matched-variance condition (15) is natural. For Walker's setting, the densities f_n and f do not have matched moments, so the simpler bounded-Lipschitz inequality is the more direct tool. However, if one wanted a rate that accounts for tail behavior (e.g., finite second moments of f_n), Rotar's lemma would give a sharper bound.

Main equivalence theorem

Our condition on $\{f_n\}_{n\in\mathbb{N}}$ is L_1-equicontinuity, which we now define.

Definition 1. *The L_1 modulus of continuity of a function $f \in L_1(\mathbb{R})$ is $\omega_1(f, \delta) = \sup_{|t|\le\delta} \int_{\mathbb{R}} |f(x+$*

$t) - f(x)|\, dx$. A sequence of functions $\{f_n\}_{n\in\mathbb{N}} \subset L_1(\mathbb{R})$ is said to be L_1-equicontinuous if

$$\lim_{\delta\to 0} \sup_{n\in\mathbb{N}} \omega_1(f_n, \delta) = 0. \tag{16}$$

Remark 4. *The L_1 modulus of continuity satisfies $\omega_1(f,\delta) \le 2\|f\|_1$ for all $\delta > 0$, and for densities $\|f_n\|_1 = 1$. Since $\omega_1(f,\delta)$ is nondecreasing in δ, the only content of (16) is the uniformity of the limit (0) as $\delta \to 0$.*

Theorem 2. *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of L_1-equicontinuous densities and f be a bounded and continuous density. Then, weak convergence of $\{f_n\}_{n\in\mathbb{N}}$ to f implies the convergence of $\{f_n\}_{n\in\mathbb{N}}$ to f in L_1.*

Proof. Take a kernel K that satisfies (8) with $\operatorname{supp} K = [-a, a]$ and has Lipschitz constant L_K. Then $\|S_{h_n} f - f\|_1 \to 0$ by Corollary 1. By Lemma 4, $\|S_{h_n}(f_n - f)\|_1 \to 0$. From $f_n - f = (S_{h_n} f - f) + (f_n - S_{h_n} f_n) + S_{h_n}(f_n - f)$, we see that it remains to establish $\|f_n - S_{h_n} f_n\|_1 \to 0$. By Fubini's theorem and the fact that $K \ge 0$, $\int K = 1$, $\operatorname{supp} K = [-a, a]$:

$$\begin{aligned}
\|S_{h_n} f_n - f_n\|_{L_1} &= \int_{\mathbb{R}} \left| \int_{-a}^{a} K(z) \big[f_n(x + h_n z) - f_n(x) \big] \, dz \right| dx \\
&\le \int_{-a}^{a} K(z) \int_{\mathbb{R}} |f_n(x + h_n z) - f_n(x)| \, dx \, dz \le \int_{-a}^{a} K(z)\, \omega_1(f_n, |z| h_n) \, dz.
\end{aligned} \tag{17}$$

Since ω_1 is non-decreasing in its second argument and $|z| \le a$, we have $\omega_1(f_n, |z|h_n) \le \omega_1(f_n, a h_n)$. Therefore,

$$\|S_{h_n} f_n - f_n\|_1 \le \omega_1(f_n, a h_n) \int_{-a}^{a} K(z) \, dz = \omega_1(f_n, a h_n). \tag{18}$$

By L_1-equicontinuity (16), the right-hand side satisfies $\sup_n \omega_1(f_n, a h_n) \to 0$ as $h_n \to 0$, completing the proof. ∎

Remark 5. *(Comparison with Walker's condition) Theorem 3.2 in Walker (2023) obtains the same equivalence but under a different condition. For $f_n^{''}$ denoting the second derivative of f_n he requires*

$$\limsup_n \left| \int_{\mathbb{R}} s^2 K(s) f_n^{''}(x + \lambda s) ds \right| < \infty \text{ for all } x \in \mathbb{R} \text{ and } \lambda \in (0, 1). \tag{19}$$

Hence, the second derivatives of f_n must exist and satisfy (19)*, which is not required by our assumption that $\{f_n\}_{n\in\mathbb{N}}$ is L_1-equicontinuous. However, sequences of densities $\{f_n\}_{n\in\mathbb{N}}$ that satisfy* (19) *are not necessarily L_1-equicontinuous. For example, take $f_n(x) = n\,\phi\,(n(x-n))$ where $\phi(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}$ and $K = \phi$. Then,* (19) *holds at every x but* (16) *fails with $\sup_n \omega_1(f_n, \delta) \geq 2$ for all $\delta > 0$. Note that $\{f_n\}_{n\in\mathbb{N}}$ is not tight, so it does not fall under the weak-convergence premise of Walker (2023); this is intentional. The point is only that condition* (19)*, taken by itself, does not entail L_1-equicontinuity. Thus, whereas the two conditions are not directly comparable for f_n that are sufficiently smooth, ours can accommodate non-smooth f_n.*

Remark 6. *(Comparison with a local Hölder condition) A function f satisfying $|f(x+t)-f(x)| \leq c(x)|t|^\alpha$ for $|t| \leq \varepsilon_0$ with $c \in L_1$ has L_1 modulus of continuity $\omega_1(f,\delta) \leq \|c\|_1\delta^\alpha$ for $\delta \leq \varepsilon_0$. Hence, a sequence $\{f_n\}_{n\in\mathbb{N}}$ with $f_n \in \mathcal{H}(c_n, \alpha)$ and $\sup_n \|c_n\|_1 < \infty$ is L_1-equicontinuous. But L_1-equicontinuity is strictly weaker: it requires $\omega_1(f_n, \delta) \to 0$ uniformly in n, with no rate, and does not assume pointwise regularity of f_n.*

Remark 7. *(Connection with the Kolmogorov–Riesz theorem) By the Kolmogorov–Riesz compactness theorem, a bounded subset of $L_1(\mathbb{R})$ is relatively compact if and only if it is L_1-equicontinuous and equitight (i.e., $\sup_n \int_{|x|>R} f_n(x)\,dx \to 0$ as $R \to \infty$). For a sequence $\{f_n\}_{n\in\mathbb{N}}$ of densities converging weakly to a density f, $\|f_n\|_1 = 1$ (boundedness) and equitightness (from tightness of the measures) are automatic. Thus, L_1-equicontinuity is precisely the missing ingredient for L_1 precompactness, and weak convergence identifies the limit uniquely. This explains why L_1-equicontinuity is the natural condition for the equivalence of weak and strong convergences. This is another proof of Theorem 2, without boundedness or continuity assumption on f. This shows that the L_1-equicontinuity assumption cannot be relaxed further. Although our proof is longer, our Theorem 1, Lemma 3 (whose proof we could not find in the extant literature) and Walker's Lemma 4 are of independent interest.*

Our equivalence result will prove particularly useful in establishing asymptotic properties of Bayesian posterior distributions. See Barron et al. (1999) and Ghosal and van der Vaart (2017).

References

Balder, E. J., 1991. On equivalence of strong and weak convergence in L_1 spaces under extreme point conditions. Israel Journal of Mathematics 75, 21–47.

Balder, E. J., 1993. From weak to strong convergence in L_1 spaces via K convergence. Annali di Matematica Pura ed Applicata CLXV, 337–349.

Barron, A., Schervish, M. J., Wasserman, L., 1999. The consistency of posterior distributions in nonparametric problems. Annals of Statistics 27, 536–561.

Billingsley, P., 1999. Convergence of probability measures, 2nd Edition. Wiley.

Bochner, S., Chandrasekharan, K., 1949. Fourier transforms. Princeton University Press, Princeton, N.J.

Bogachev, V. I., 2007. Measure Theory. Springer, Berlin.

Dellacherie, C., Meyer, P.-A., 1978. Probabilities and potential, 1st Edition. Vol. 29 of Mathematical Studies. North Holland.

Devroye, L., Györfi, L., 1985. Nonparametric density estimation: the L_1 view. John Wiley and Sons, New York, NY.

Dudley, R. M., 2002. Real analysis and probability, 2nd Edition. No. 74 in Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Ghosal, S., van der Vaart, A., 2017. Fundamentals of nonparametric Bayesian inference. No. 44 in Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, UK.

Hettmansperger, T., Klimko, L., 1974. A note on the strong convergence of distributions. Annals of Statistics 2, 597–598.

Hille, E., Phillips, R., 1974. Functional analysis and semi-groups. Vol. XXXI of American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI.

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

Rotar, V. I., 1975. An extension of the lindeberg-feller theorem. Matematicheskie Zametki 18, 129–135.

Visintin, A., 1984. Strong convergence results related to strict convexity. Communications in Partial Differential Equations 9, 439–466.

Walker, S. G., 2023. Comparing weak and strong convergence of density functions. Statistics and Probability Letters 200 (article 109878), 5 pp.