Metrika, Volume 33, 1986, page 197–202

Rates of Convergence for the Distance Between Distribution Function Estimators

By D. D. Boos[1]

Summary: The normed difference between "kernel" distribution function estimators $\hat{F}_n$ and the empirical distribution function F_n is investigated. Conditions on the kernel and bandwidth of $\hat{F}_n$ are given so that $a_n \|\hat{F}_n - F_n\| \xrightarrow{wp1} 0$ as $n \to \infty$ for both the sup-norm $\|g\|_\infty = \sup_x |g(x)|$ and L_1 norm $\|g\|_1 = \int |g(x)|\, dx$. Applications include equivalence in asymptotic distribution of $T(\hat{F}_n)$ and $T(F_n)$ (to order a_n) for certain robust functionals $T(\cdot)$.

1 Introduction

The empirical distribution function $F_n(x) = n^{-1} \sum I(X_i \leqslant x)$ is the most widely used nonparametric distribution function estimator. However, integrals of kernel density estimators form a large class of smooth competitors. These estimators may be expressed as

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b_n}\right) = \int_{-\infty}^{\infty} K\left(\frac{x-y}{b_n}\right) dF_n(y), \tag{1.1}$$

where K is a distribution function on $(-\infty,\infty)$ and $b_n > 0$ is the "bandwidth". Nadaraya (1964), Winter (1973, 1979), Yamato (1973), Azzalini (1981), Reiss (1981), and Falk (1983) have studied the convergence of $\hat{F}_n$ to F, the distribution function of the observations. The purpose of this present note is to give sufficient conditions on F, K, and b_n so that $a_n \|\hat{F}_n - F_n\| \xrightarrow{wp1} 0$ for several norms $\|\cdot\|$. Of particular interest is the sup-norm $\|\hat{F}_n - F_n\|_\infty = \sup_{-\infty < x < \infty} |\hat{F}_n(x) - F_n(x)|$ and $a_n = n^{1/2}$. Then $n^{1/2} \|\hat{F}_n - F_n\|_\infty \xrightarrow{wp1} 0$ implies weak convergence of $n^{1/2}[\hat{F}_n(F^{-1}(t)) - t]$ as well as the Chung-Smirnov property (law of the iterated logarithm for $\|\hat{F}_n - F\|_\infty$) of Winter (1979) and the asymptotic equivalence in distribution of $n^{1/2}[T(\hat{F}_n) - T(F)]$ and $n^{1/2}[T(F_n) - T(F)]$ for the many robust functionals $T(\cdot)$ which are locally Lipschitz with respect to $\|\cdot\|_\infty$, i.e., $|T(G) - T(H)| \leqslant C \|G - H\|_\infty$ for all $\|G - H\|_\infty$ sufficiently small. More details and

[1] Dennis D. Boos, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203 USA

0026–1335/86/03040197–202 $2.50

applications are given in Section 3. The main results proved in Section 2 are simple and rely on the usual integration by parts representation of $\hat{F}_n$ and Serfling's (1980) generalization of a theorem of Sen and Ghosh (1971) on the uniform convergence of $F_n(x+t) - F(x+t) - [F_n(x) - F(x)]$ as $t \to 0$.

2 Main Results

For functions G and H on $(-\infty, \infty)$ define $\|G - H\|_\infty = \sup_{-\infty < x < \infty} |G(x) - H(x)|$ and $\|G - H\|_1 = \int_{-\infty}^{\infty} |G(x) - H(x)|\,dx$. The sampling situation is

(S) $X_1, \ldots, X_n$ are independent real-valued random variables having the distribution function F.

The first two theorems require

(C) K has support in a compact interval $[c, d], -\infty < c < d < \infty$.

Lemma 2.1 of Winter (1979) justifies the use of integration by parts to reexpress $\hat{F}_n$ of (1.1) as

$$\hat{F}_n(x) = \int_{-\infty}^{\infty} F_n(x - b_n y)\,dK(y).$$

Theorem 1: Suppose that (S) and (C) hold and F satisfies the Lipschitz condition $|F(x) - F(y)| \leqslant L|x - y|$ on $(-\infty, \infty)$. If $a_n b_n \xrightarrow{wp1} 0$ and $nb_n/\log n \xrightarrow{wp1} \infty$, then

$$a_n \|\hat{F}_n - F_n\|_\infty \xrightarrow{wp1} 0 \quad \text{as } n \to \infty. \tag{2.1}$$

Proof: Let $e(c, d) = \max\{|c|, |d|\}$. Then

$$a_n \|\hat{F}_n - F_n\|_\infty \leqslant a_n \sup_x \int_c^d |F_n(x - b_n y) - F_n(x)|\,dK(y)$$

$$\leqslant a_n \sup_x \int_c^d |F_n(x - b_n y) - F(x - b_n y) - [F_n(x) - F(x)]|\,dK(y)$$

$$+ a_n \sup_x \int_c^d |F(x - b_n y) - F(x)|\,dK(y) \tag{2.2}$$

$$\leqslant a_n \sup_x \sup_{|t| \leqslant e(c,d) b_n} |F_n(x+t) - F(x+t) - [F_n(x) - F(x)]|$$

$$+ a_n b_n L \int_c^d |y|\,dK(y). \tag{2.3}$$

Rewrite the first term of (2.3) as $a_n Q_n$. Lemma 2.2 of Serfling (1980) yields

$$\left[\frac{n}{b_n \log n}\right]^{1/2} Q_n = 0(1) \quad wp1 \text{ as } n \to \infty$$

for uniform random variables. (See also Stute 1982, Theorem 0.1.) Following the remark on page 194 of Sen and Ghosh (1971), this result also holds for all F which are Lipschitz. That is, if $G_n(x) = n^{-1} \sum_{i=1}^{n} I(F(X_i) \leq x)$, then

$$\sup_{-\infty<x<\infty} \sup_{|t|\leq e(c,d)b_n} |F_n(x+t) - F(x+t) - [F_n(x) - F(x)]|$$

$$\overset{wp1}{=} \sup_{-\infty<x<\infty} \sup_{|t|\leq e(c,d)b_n} |G_n(F(x+t)) - F(x+t) - [G_n(F(x)) - F(x)]|$$

$$\leq \sup_{0<u<1} \sup_{|v|\leq Le(c,d)b_n} |G_n(u+v) - (u+v) - [G_n(u) - u]|,$$

and Serfling's result applies. Then $a_n Q_n \xrightarrow{wp1} 0$ since $a_n Q_n = a_n b_n [\log n/(nb_n)]^{1/2} [n/(b_n \log n)]^{1/2} Q_n$.

Remarks 2.1: The proof shows that the weaker result, $a_n \|\hat{F}_n - F_n\|_\infty = 0(1)\ wp1$, holds if only $a_n b_n = 0(1)\ wp1$ is assumed. The bandwidth b_n is allowed to be random since b_n must be estimated in most applications. The usual choice of a_n is $n^{1/2}$ (see Section 3) so that Theorem 1 allows $b_n \sim n^{-\alpha}, \alpha > 1/2$.

The next theorem strengthens conditions on F and K in order to reduce conditions on b_n. In particular, if $a_n = n^{1/2}$ then Theorem 2 allows $b_n \sim n^{-\alpha}, \alpha > 1/4$.

Theorem 2: Suppose that (S) and (C) hold and $\int_c^d x dK(x) = 0$. Let F have derivatives f and f' which exist everywhere on $(-\infty, \infty)$ with $\|f\|_\infty < \infty$ and $\|f'\|_\infty < \infty$. If $a_n b_n^2 \xrightarrow{wp1} 0$, $nb_n/\log n \xrightarrow{wp1} \infty$, and $a_n[b_n \log n/n]^{1/2} \xrightarrow{wp1} 0$, then (2.1) holds.

Proof: The proof is the same as for Theorem 1 except that the second term of (2.2) is expanded by Taylor's theorem to yield

$$a_n \sup_x \left|\int_c^d [F(x - b_n y) - F(x)] dK(y)\right|$$

$$= a_n \sup_x \left|\int_c^d [-f(x)(b_n y) + 1/2 f'(t_n^*)(b_n y)^2] dK(y)\right|$$

$$\leq a_n b_n^2 (1/2) \|f'\|_\infty \int_c^d y^2 dK(y).$$

Theorems 1 and 2 are similar in spirit to Theorems 3.2 and 3.3 of Winter (1979) which conclude that

$$\limsup_{n\to\infty} \{2n/\log\log n\}^{1/2} \|\hat{F}_n - F\|_\infty \leqslant 1 \ wp1. \tag{2.4}$$

Here, condition (C) and a slightly stronger condition on b_n than Winter required yield (2.1) which in turn yields (2.4).

Remarks 2.2: i) The extra condition in Theorem 2, $a_n[b_n \log n/n]^{1/2} \xrightarrow{wp1} 0$, is required to have $a_n Q_n \xrightarrow{wp1} 0$. When $a_n = 0(n^{1/2})$, this is implied by $b_n \log n \xrightarrow{wp1} 0$.

ii) It is clear that we could allow K to be more general than a distribution function as in Reiss (1981, Condition A). That is, if (C) holds and $K(x) = \int_{-\infty}^{x} k(y)dy$ with $\lim_{x\to-\infty} K(x) = 0$, $\lim_{x\to\infty} K(x) = 1$, $\int x^i k(x)dx = 0$ for $i = 1, \ldots, m$ and $\int |x^{m+1} k(x)| dx < \infty$, and F has $m + 1$ bounded derivatives, then there is a constant $C(m, k, F)$ such that

$$a_n \|\hat{F}_n - F_n\|_\infty \leqslant a_n Q_n \int_{-\infty}^{\infty} |k(x)| dx + a_n b_n^{m+1} C(m, k, F).$$

Reiss (1981), Remark 1.1) gives such a kernel for $m = 3$. In the case that $a_n = n^{1/2}$ and the optimal bandwidth rate $b_n \sim n^{-1/(2m+1)}$ is used, then (2.1) holds since $(m + 1)/(2m + 1) > 1/2$.

The last theorem in this section applies to $\|\cdot\|_1$.

Theorem 3: Suppose that (S) holds and $\int |x| dK(x) < \infty$. If $a_n b_n \xrightarrow{wp1} 0$, then

$$a_n \|\hat{F}_n - F_n\|_1 \xrightarrow{wp1} 0 \quad \text{as } n \to \infty. \tag{2.5}$$

Proof:

$$a_n \int_{-\infty}^{\infty} |\hat{F}_n(x) - F_n(x)| dx \leqslant a_n \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |F_n(x - b_n y) - F_n(x)| dK(y)dx$$

$$= a_n \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |F_n(x - b_n y) - F_n(x)| dx dK(y)$$

$$= a_n b_n \int_{-\infty}^{\infty} |y| dK(y).$$

The interchange of integrals is justified by Fubini's theorem, and the last step follows since $\int_{-\infty}^{\infty} [G(x + a) - G(x)]dx = a$ for any distribution function G and constant $a \geqslant 0$ (see Chung 1974, p. 49, prob. 16).

3 Applications

A) Weak convergence of $n^{1/2}[\hat{F}_n(F^{-1}(t)) - t]$ to W^0, the Brownian bridge. Choose $a_n = n^{1/2}$. If K is continuous, then the convergence may be carried out in $C[0, 1]$ using Theorems 4.1 and 13.1 of Billingsley (1968). If K is not continuous, then the space $D[0, 1]$ is appropriate and Theorems 4.1 and 16.4 of Billingsley yield the result. (In verifying the latter it helps to note that the uniform metric $\rho(x, y) = \|x - y\|_\infty$ dominates either of the Skorohod metrics given by Billingsley, see page 150.)

B) Theorem 4.2 of Sen and Ghosh (1971) mentioned in the proof of Theorem 1 yields for F Lipschitz

$$\sup_{-\infty < x < \infty} \sup_{|t| \leqslant d_n} n^{1/2} |F_n(x + t) - F(x + t) - [F_n(x) - F(x)]| \xrightarrow{wp1} 0, \tag{3.1}$$

where $n^{1/2} d_n$ increases at a rate not slower than that of $\log n$ but not faster than that of n^α, $\alpha < 1/4$. The results of Theorems 1 and 2 with $a_n = n^{1/2}$ allow (3.1) to hold with F_n replaced by $\hat{F}_n$.

C) Statistical functions $T(F)$. If $|T(\hat{F}_n) - T(F_n)| \leqslant C_n \|\hat{F}_n - F_n\|_\infty$ and $C_n = 0_p(1)$, then $a_n[T(\hat{F}_n) - T(F_n)] \xrightarrow{p} 0$ and $T(\hat{F}_n)$ and $T(F_n)$ have the same asymptotic distribution up to order a_n. A trivial extension is to replace X_i by the perturbed random variable $X_i + Y_n$ (e.g., Pitman location alternatives). Some specific $T(\cdot)$ are given below.

1. Quantile estimation, $T(F) = F^{-1}(p) = \inf\{x : F(x) \geqslant p\}$. Suppose that $F'(F^{-1}(p)) > 0$. For distribution functions G and H and $\|G - H\|_\infty$ sufficiently small, one can verify that

$$|G^{-1}(p) - H^{-1}(p)| \leqslant \left(\epsilon + \frac{1}{H'(H^{-1}(p))}\right) \|G - H\|_\infty. \tag{3.1}$$

   Letting $G = \hat{F}_n$ and $H = F$ in (3.1) and $a_n = 1$, we get $\hat{F}_n^{-1}(p) \xrightarrow{wp1} F^{-1}(p)$. Under the conditions of Theorem A of Silverman (1978) which include uniform continuity of F' and K', we have $\hat{F}_n'(\hat{F}_n^{-1}(p)) = F'(\hat{F}_n^{-1}(p)) + [\hat{F}_n'(\hat{F}_n^{-1}(p)) - F'(\hat{F}_n^{-1}(p))] \xrightarrow{wp1} F'(F^{-1}(p))$. Thus applying (3.1) with $G = F_n$ and $H = \hat{F}_n$ and setting $a_n = n^{1/2}$ yields $n^{1/2}[T(\hat{F}_n) - T(F_n)] \xrightarrow{p} 0$. This approach along with Theorem 2 yields an asymptotic normality result for $T(\hat{F}_n)$ comparable to that of Nadaraya (1964). Azzalini (1981) notes that the optimal bandwidth rate is $b_n \sim n^{-1/3}$. Theorem 2 requires $b_n = o(n^{-1/4})$.
2. L-estimators with smooth score function, $T(F) = \int F^{-1}(t)J(t)dt$. Boos (1979), Theorem 1, showed that $T(\cdot)$ has a Frechet differential with respect to $\|\cdot\|_\infty$ under certain conditions on J and F. Similarly, it can be shown that if J is bounded and integrable on $[0, 1]$ and equal to zero in neighborhoods of 0 and 1, then for all n sufficiently large $wp1$ there exists a constant $C(J, F)$ such that

$$|T(\hat{F}_n) - T(F_n)| \leqslant C(J, F) \|\hat{F}_n - F_n\|_\infty.$$

Alternatively, if J is merely bounded and integrable on [0, 1],

$$|T(\hat{F}_n) - T(F_n)| \leqslant \|J\|_\infty \|\hat{F}_n - F_n\|_1 .$$

Each of these bounds follows directly from the representation

$$T(G) - T(H) = \int_{-\infty}^{\infty} [S(H(x)) - S(G(x))]dx \quad \text{where } S(t) = \int_0^t J(u)du.$$

3. Other examples include classes of R-estimators, minimum distance estimators, and one-step M-estimators.

Acknowledgement: The author wishes to thank the referee for helpful comments.

References

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

Billingsley P (1968) Convergence of probability measures. John Wiley, New York

Boos DD (1979) A differential for L-statistics. Annals of Statistics 7:955–959

Chung KL (1974) A course in probability theory. Academic Press, New York

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

Nadaraya EA (1964) Some new estimates for distribution functions. Theory of Probability and its Applications 9:497–500

Reiss R-D (1981) Nonparametric estimation of smooth distribution functions. Scandinavian Journal of Statistics 8:116–119

Sen PK, Ghosh M (1971) On bounded length sequential confidence intervals based on one-sample rank order statistics. Annals of Mathematical Statistics 42:189–203

Serfling RJ (1980) Properties and applications of metrics on nonparametric density estimators. Proceedings of the International Colloquium on Nonparametric Statistical Inference, Budapest. North Holland, pp 859–873

Silverman BW (1978) Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. Annals of Statistics 6:177–184

Stute W (1982) The oscillation behavior of empirical processes. Annals of Probability 10:86–107

Winter BB (1973) Strong uniform consistency of integrals of density estimators. Canadian Journal of Statistics 1:247–253

Winter BB (1979) Convergence rate of perturbed empirical distribution functions. Journal of Applied Probability 16:163–173

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

Received September 13, 1983
(Revised version June 12, 1984)