Rates of Convergence for the Distance Between Distribution Function Estimators

Dennis D. Boos

Inst. of Statistics Mimeo Series # 1390

ABSTRACT

Rates of Convergence for the Distance between
Distribution Function Estimators

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 $F_n(x) = n^{-1}\sum I(X_i \le x)$ 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." Winter (1973,1979), Yamato (1973), and Azzalini (1980) 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 $n^{1/2}[\hat{F}_n(F^{-1}(t))-t]$, 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 Lipschitz with respect to $||\cdot||_\infty$, i.e., $|T(G)-T(H)| \le C||G-H||_\infty$. More details and 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 (1981) 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

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

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

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

PROOF.

$$
\begin{aligned}
a_n||\hat{F}_n - F_n||_\infty &\le a_n \sup_x \int_c^d |F_n(x-b_n y) - F_n(x)|\, dK(y) \\
&\le a_n \sup_x \int_c^d |F_n(x-b_n y) - F(x-b_n y) - [F_n(x) - F(x)]|\, dK(y) \\
&\quad + a_n \sup_x \int_c^d |F(x-b_n y) - F(x)|\, dK(y) \qquad (2.2)\\
&\le a_n \sup_x \sup_{|t|\le(d-c)b_n} |F_n(x+t) - F(x+t) - [F_n(x) - F(x)]| \\
&\quad + a_n b_n L \int_c^d |y|\, dK(y)\ . \qquad (2.3)
\end{aligned}
$$

Rewrite the first term of (2.3) as $a_n Q_n$. Lemma 2.2 of Serfling (1981) yields

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

for uniform random variables. Following the remark on page 194 of Sen and Ghosh (1971), this result also holds for all F which are Lipschitz. Then $a_n Q_n \xrightarrow{\text{wp1}} 0$ since $a_n[b_n \log n/n]^{\frac{1}{2}} = a_n b_n [\log n/nb_n]^{\frac{1}{2}} \xrightarrow{\text{wp1}} 0$.

REMARKS. 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 $n^{\frac{1}{2}}$ (see Section 3) so that Theorem 1 allows $b_n \sim n^{-\alpha}$, $\alpha > \frac{1}{2}$.

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

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 $a_n b_n^2 \xrightarrow{wp1} 0$ and $nb_n/\log n \xrightarrow{wp1} \infty$, 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

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

$$\leq a_n b_n^2 (\tfrac{1}{2}) ||f'||_\infty \int y^2 dK(y) .$$

EXAMPLE. Suppose that F is the uniform distribution function on [0,1]. Then Theorem 1 applies but not Theorem 2.

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

$$\lim_{n\to\infty} \sup\{2n/\log\log n\}^{\frac{1}{2}} ||\hat{F}_n - F||_\infty \leq 1 \quad \text{wp1} . \tag{2.4}$$

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

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

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

PROOF.

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

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

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

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 $a \ge 0$ (see Chung (1974), p. 49, prob. 16).

3. Applications

A) Weak convergence of $n^{\frac{1}{2}}[\hat{F}_n(F^{-1}(t))-t]$ to W^o, the Brownian bridge. Choose $a_n = n^{\frac{1}{2}}$. 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.)

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

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

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

C) Statistical functions $T(F)$. If $|T(\hat{F}_n) - T(F_n)| \le C_n ||\hat{F}_n - F_n||_\infty$ and $C_n = O_p(1)$, 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, $T(F) = F^{-1}(p) = \inf\{x : F(x) \ge p\}$. Suppose that $F'(F^{-1}(p)) > 0$. For distribution functions G and H and $||G-H||_\infty$ sufficiently small, one can verify that

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

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 $a_n = n^{\frac{1}{2}}$ yields $n^{\frac{1}{2}}[T(\hat{F}_n) - T(F_n)] \xrightarrow{p} 0$. Azzalini (1980) has apparently shown this result directly and also obtained the optimal rate for b_n of the order $n^{-\frac{1}{4}}$. Theorem 2 requires $b_n < C_1 n^{-\frac{1}{4}}$.

2. L-estimators with smooth score function, $T(F) = \int F^{-1}(t)J(t)dt$. Boos (1979a), 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

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

Alternatively, if J is continuous on $[0,1]$ (and thus bounded), then

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

Each of these bounds follows directly from the representation $T(G) - T(H) = \int_{-\infty}^{\infty} [S(H(x)) - S(G(x))]dx$ 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.

4. For certain cases $T(\hat{F}_n)$ may be easier to handle than $T(F_n)$ because of the smoothness of $\hat{F}_n$. In Boos (1979b) the minimum value of

$$d_{F_n}(\theta) = \int_{-\infty}^{\infty} [1 - F_n(\theta+x) - F_n(\theta-x)]^2 dF_n(\theta+x)$$

was proposed as a test statistic for detecting asymmetry in F. The limiting distribution of the *location* of the minimum $\theta(F_n)$ as well as that of the minimum $T(F_n) = d_{F_n}(\theta(F_n))$ is not hard to find. However, details of proof are made difficult by the fact that $d_{F_n}(\theta)$ is a step function. Consider the following approach. If H is a continuous distribution function, then simple calculations show that

$$\sup_\theta |d_G(\theta) - d_H(\theta)| \leq C||G-H||_\infty$$

and thus

$$\left|\min_\theta d_G(\theta) - \min_\theta d_H(\theta)\right| \leq C||G-H||_\infty . \tag{3.2}$$

First let $G = F$ and $H = \hat{F}_n$. Then (3.2) and (2.1) implies that $\theta(\hat{F}_n) \xrightarrow{wp1} \theta(F)$ (see Lemma 2 of Beran (1978) for the technique of proof). Note that consistency of $\theta(F_n)$ does not follow since we have not shown $|\theta(G)-\theta(H)| \leq C||G-H||_\infty$. However, we can proceed directly and find the limiting distribution of $nT(\hat{F}_n)$ (see Boos (1979b), Section 2, for analogous details). Then using (3.2) again with $G = F_n$ and $H = \hat{F}_n$ and choosing $a_n = n$ in (2.1), we get that $nT(\hat{F}_n)$ and $nT(F_n)$ have the same limiting distribution. Here, Theorem 2 is necessary to get useful results since $nb_n/\log n \to \infty$. Note also that only one $\hat{F}_n$ meeting the above requirements must be found since $T(F_n)$ is the statistic of interest. Koul (1980) has obtained the asymptotic distribution of $\theta(F_n)$ using linear rank methods.

REFERENCES

Azzalini, A. (1981), "A Note on the Estimation of a Distribution Function and Quantiles by a Kernel Method," *Biometrika*, 68, 326-328.

Beran, R. (1978), "An Efficient and Robust Adaptive Estimator of Location," *Annals of Statistics*, 6, 292-313.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: John Wiley.

Boos, D. D. (1979a), "A Differential for L-Statistics," *Annals of Statistics*, 7, 955-959.

Boos, D. D. (1979b), "A Test of Symmetry Associated with the Hodges-Lehmann Estimator," Institute of Statistics Mimeo Series # 1248, North Carolina State University, Raleigh.

Chung, K. L. (1974), *A Course in Probability Theory*, New York: Academic Press.

Koul, H. (1980), "Some Weighted Empirical Inferential Procedures for a Simple Regression Model," preprint.

Sen, P. K., and Ghosh, M. (1971), "On Bounded Length Sequential Confidence Intervals Based on One-Sample Rank Order Statistics, *Annals of Mathematical Statistics*, 42, 189-203.

Serfling, R. J. (1981), "Properties and Applications of Metrics on Nonparametric Density Estimators," Department of Mathematical Sciences Technical Report No. 324, The Johns Hopkins University, Baltimore.

Silverman, B. W. (1978), "Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and its Derivatives," *Annals of Statistics*, 6, 177-184.

Winter, B. B. (1973), "Strong Uniform Consistency of Integrals of Density Estimators," *Canadian Journal of Statistics*, 1, 247-253.

Winter, B. B. (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.