REMARKS ON SOME SMOOTHED EMPIRICAL DISTRIBUTION FUNCTIONS AND PROCESS

Mack, Y. P.
Division of Statistics, University of California

https://doi.org/10.5109/13353

出版情報：Bulletin of informatics and cybernetics. 21 (1/2), pp.29-35, 1984-03. Research Association of Statistical Sciences

Bulletin of Informatics and Cybernetics Vol. 21, No. 1~2, 1984

REMARKS ON SOME SMOOTHED EMPIRICAL DISTRIBUTION FUNCTIONS AND PROCESSES*

By

Y. P. MACK**

Abstract

It is shown that under mild assumptions, a convolution-smoothed empirical process exhibits essentially the same asymptotic properties as the standard empirical process such as: a pointwise law of iterated logarithm, weak convergence to Brownian bridge, and the Chung-Smirnov property. Some remarks of statistical and probabilistic interests are made. A list of open questions is also included.

I. Introduction

Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with common distribution function (d.f.) F on the real line. Denote by F_n the standard empirical d.f., i.e., $F_n(x)$=proportion of observations$\leqq x$. By now there is a rich body of literature on the study of F_n and the related empirical process $Z_n=n^{1/2}(F_n-F)$ (see Gaenssler and Stute [8]). When F is absolutely continuous with density f, it is natural to look for some smooth estimates $\hat{F}_n$ of F. For example, Kronmal and Tarter [12] proposed using a trigonometric series estimate in connection with address calculation sorting. Recently, Efron [7] and Boos and Monahan [5] have suggested using $\hat{F}_n$ instead of F_n to generate bootstrap samples. Their initial Monte Carlo experiments appeared to be encouraging.

Typically $\hat{F}_n$ is constructed by taking an indefinite integral (if it exists) of some density estimates based on some "delta sequence" as studied by Walter and Blum [27]. Among these which are not convolution-based are, for instance, approaches based on orthogonal expansions (e.g., Schwartz [19], Kronmal and Tarter [12], and Walter [26]); approaches based on spline interpolations (e.g., Boneva, Kendall and Stefanov [4], Wahba [25], Lii and Rosenblatt [13]); and approaches based on nonparametric maximum likelihood considerations (e.g., Geman and Huang [9], Blum and Walter [3]). Another "delta sequence" approach is the kernel method, by far the most studied in density estimation, which dates back to Rosenblatt [17] in 1956. The kernel density estimates have the form

$$f_n(x)=\int \frac{1}{b_n} w\left(\frac{x-t}{b_n}\right) dF_n(t)$$

* This research was carried out in part while the author was visiting the Institute of Applied Mathematics, University of Heidelberg, West Germany, in the summer of 1980 under the sponsorship of Sondersforschungsbereich 123. This research is also supported in part by NSF research grant #MCS 82-13801.

** Division of Statistics, University of California, Davis, California 95616, U.S.A.

$$=\frac{1}{nb_n}\sum_{i=1}^{n} w\left(\frac{x-X_i}{b_n}\right)$$

where w is a density and $\{b_n\}$ is a so-called "bandwidth" sequence which tends to 0 as n tends to ∞. The smoothed estimate $\hat{F}_n$ of F based on the kernel method is thus given by a convolution:

(1)
$$\begin{aligned}\hat{F}_n(x)&=\int_{-\infty}^{x} f_n(u)du\\&=\iint_{-\infty}^{x}\frac{1}{b_n}w\left(\frac{u-t}{b_n}\right)du\,dF_n(t)\\&=\int W\left(\frac{x-t}{b_n}\right)dF_n(t)\\&=\frac{1}{n}\sum_{i=1}^{n}W\left(\frac{x-X_i}{b_n}\right)\end{aligned}$$

where $W(u)=\int_{-\infty}^{u} w(t)dt$. Note that $\left\{\frac{1}{b_n}w\left(\frac{u}{b_n}\right)\right\}$ is a "delta sequence", while $\left\{W\left(\frac{u}{b_n}\right)\right\}$ is a sequence of d.f.'s converging weakly to the d.f. of the unit mass at 0. (In the spirit of Walter and Blum [27], we may call such a sequence a "Heaviside sequence".)

While some estimates of the non-convolution type seem to have computational appeal (such as the trigonometric series method), and others may have global optimality at the finite sample level (such as the spline estimates and MLE's), they frequently suffer from the drawback of not being d.f.'s themselves. In addition, many estimates based on global optimality are constructed implicity, making statistical analysis quite intractible. The convolution-based approach is easier to handle, and the computation issue is made less serious because of current advances in computer science, and by using fast Fourier transform technique as described by Silverman [21].

For the remainder of this discussion, we shall concentrate our efforts on the convolution-based estimates (1) or, more generally,

(2)
$$\begin{aligned}\hat{F}_n(x)&=\int W_n(x-t)dF_n(t)\\&=\frac{1}{n}\sum_{i=1}^{n}W_n(x-X_i)\,,\end{aligned}$$

where $\{W_n\}$ is a Heaviside sequence. Note that there are delta sequences not of the kernel type. For example, the Landau sequence described in Walter and Blum [27], or the Fejer sequence encountered in time series.

A detailed listing of research on (1) or (2) is given in the references. Most results seem to indicate that the asymptotic behaviour of the smoothed estimates and that of the standard empirical d.f. are quite similar. For example, Yamato [31] showed that (2) has the same asymptotic covariance as F_n, while Winter [30] showed that (1) has the Chung-Smirnov property.

In this article, we consider the smoothed empirical processes

$$\hat{Z}_n(x)=n^{1/2}[\hat{F}_n(x)-E\hat{F}_n(x)]$$

or

$$Z_n^*(x)=n^{1/2}[\hat{F}_n(x)-F(x)]$$

with $\hat{F}_n$ given by (1) or (2). We shall demonstrate that under mild conditions, a pointwise law of iterated logarithm holds for (1) by using a result of Hall [10], and that $\hat{Z}_n$ and Z_n are uniformly close by appealing to a result of Stute [24]. Consequently, $\hat{Z}_n$ inherits many asymptotic properties of Z_n. We shall also make remarks on these results which have both statistical and probabilistic interests,

2. Main Results and Remarks

In what follows, $\|\ \|$ will stand for the supremum norm over $\boldsymbol{R}^1$, and "w. p. 1." will be an abbreviation for "with probability one". F is assumed to have a density.

THEOREM 1. *The following pointwise law of iterated logarithm holds for $\hat{Z}_n$ with $\hat{F}_n$ given by* (2):

$$(3)\qquad \overline{\lim_{n\to\infty}}\pm[2\log\log n\cdot F(x)(1-F(x))]^{-1/2}\hat{Z}_n(x)=1$$

*w.p.*1 *for each x such that $F(x)\neq 0$,* 1.

PROOF OF THEOREM 1. We verify that the hypothesis of Theorem 1 in Hall [10] is satisfied. Let

$$\sigma_{mn}=\operatorname{cov}[W_m(x-X_1),\ W_n(x-X_l)]$$

$$\sigma_n^2=\sigma_{nn}\,.$$

Since $\{W_n\}$ is obviously a sequence of functions of bounded variation on $\boldsymbol{R}^1$, it remains to check that

$$(4)\qquad \sigma_{mn}/\sigma_n^2\longrightarrow 1\qquad \text{as}\quad m,\ n\to\infty \text{ with } n/m\to 1$$

and that

$$(5)\qquad (\log n)^4\Big[\int dW_n(t)\Big]^2\Big/n\sigma_n^2\log\log n\longrightarrow 0\qquad \text{as}\quad n\to\infty\,.$$

Now as mentioned earlier, $\sigma_n^2\to F(x)(1-F(x))$ as $n\to\infty$ (see Yamato [31]), so clearly (5) holds. To see that (4) holds, it is enough to show that $|\sigma_{mn}-\sigma_n^2|\to 0$, or equivalently, that

$$(6)\qquad |E[W_m(x-X_1)\cdot W_n(x-X_1)]-E[W_n^2(x-X_1)]|\longrightarrow 0$$

as $m,\ n\to\infty$. Writing the expression in (6) as

$$|E\{W_n(x-X_1)[W_m(x-X_1)-W_n(x-X_1)]\}|$$

we see that the result follows easily since W_n is bounded and that $EW_n(x-X_1)$ and $EW_m(x-X_1)$ both tend to $F(x)$ in the limit as $n,\ m\to\infty$.

THEOREM 2. *As in Stute* [24], *let $\{a_n\}$ be a sequence of positive real numbers satisfying condition*

$$\text{(A)}\qquad \begin{aligned}&\text{(i)}\quad a_n\downarrow 0\quad \text{as } n\to\infty\\ &\text{(ii)}\quad \log\frac{1}{a_n}=o(na_n)\end{aligned}$$

(iii) $\log \log n = o\left(\log \frac{1}{a_n}\right)$.

Let $\tau_n = \int_{|t|>a_n} dW_n(t)$ *and* $\theta_n = \left(a_n \log \frac{1}{a_n}\right)^{1/2}$. *Then*

$$\|\hat{Z}_n - Z_n\| = O[\theta_n + \tau_n (\log \log n)^{1/2}] \tag{6}$$

w.p. 1.

COROLLARY 1. *Suppose* $\{a_n\}$ *satisfies* (A) (i)-(iii) *and* $\{W_n\}$ *satisfies the tail condition*

$$\tau_n = o((\log \log n)^{-1/2}). \tag{T}$$

Then $\hat{Z}_n$ *converges weakly to the rescaled Brownian bridge* $B^\circ(F)$.

COROLLARY 2. *Under condition* (A) (i)-(iii), *on a rich enough probability space (à la Komlós, Major, Tusnády* [11]), *there exist a version of* $\hat{Z}_n$ *and a sequence of Brownian bridges* B_n° *such that*

$$\|\hat{Z}_n - B_n^\circ(F)\| = O[n^{-1/2} \log n + \|\hat{Z}_n - Z_n\|] \tag{7}$$

w.p. 1, *where* $\|\hat{Z}_n - Z\|$ *is given by* (6).

(There is a similar result using a Kiefer process instead of Brownian bridges.)

COROLLARY 3. *Suppose* $\{a_n\}$ *satisfies* (A) (i)-(iii) *and suppose* $\tau_n = o(1)$

Then

$$\overline{\lim_{n\to\infty}} (\log \log n)^{-1/2} \|\hat{Z}_n\| = 2^{-1/2} \tag{8}$$

w.p. 1.

(This result is in particular an improvement over Theorems 3.2 and 3.3 of Winter [30] on the Chung-Smirnov property of $\hat{Z}_n$ with $\hat{F}_n$ defined by (1).)

PROOF OF THEOREM 2. Write

$$\begin{aligned} |\hat{Z}_n(x) - Z_n(x)| &= \left| \int Z_n(x-t) dW_n(t) - Z_n(x) \int dW_n(t) \right| \\ &\leqq \int |Z_n(x-t) - Z_n(x)| \, dW_n(t) \\ &= \int_{|t|\leqq a_n} + \int_{|t|>a_n}. \end{aligned}$$

$$\begin{aligned} \left\| \int_{|t|\leqq a_n} \right\| &\leqq \sup_{\substack{|t|\leqq a_n \\ \text{all } x}} |Z_n(x-t) - Z_n(x)| \cdot \int_{|t|\leqq a_n} dW_n(t) \\ &= O(\theta_n) \end{aligned}$$

w.p. 1. by Theorem 0.2 of Stute [24], while

$$\begin{aligned} \left\| \int_{|t|>a_n} \right\| &\leqq 2 \|Z_n\| \cdot \int_{|t|>a_n} dW_n(t) \\ &= O[\tau_n (\log \log n)^{1/2}] \end{aligned}$$

w.p.1. by the law of iterated logarithm of Z_n. Thus (6) is prorved. The corollaries follow by standard results on Z_n (see for example Billingsley [2] or Csörgő and Révész [6].)

REMARKS.

(a) $a_n=n^{-\lambda}\log n$, $0<\lambda<1$, is an example of $\{a_n\}$ satisfying (A) (i)-(iii) in Theorem 2.

(b) For kernel-type smoothers (1), if W (as a d.f.) has absolute k^{th} moment, $k>0$, and is symmetric about 0, then

$$\tau_n=\int_{|v|>a_n/b_n} dW(v)\leq O\left[\left(\frac{b_n}{a_n}\right)^k\right].$$

So we can choose, for example,

$$b_n=o[a_n(\log\log n)^{-1/2k}]$$

in order to satisfy condition (T). Various combinations of choices of $\{a_n\}$, $\{b_n\}$ and k are possible for the corollaries to hold. As an illustration, take $k=1$, $a_n=n^{-1/10}$, $b_n=n^{-1/5}$, then Corollary (3) holds.

For kernels W with compact support, we can allow $a_n=c\ b_n$ with $0<c<\infty$ chosen so that w vanishes outside $[-c, c]$. In this case $\tau_n=0$ and $\{b_n\}$ should satisfy the same condition as $\{a_n\}$.

(c) There are refinements on the oscillation behaviour of Z_n (see Shorack and Wellner [20].) For example, let $a_n=n^{-1}(\log n)^{\delta}$, $-\infty<\delta<1$. Then the rate θ_n in Theorem 2 is replaced by $a_n^{1/2}(\log n)^{1-\delta/2}(\log\log n)^{-1}$. However, with $\delta=-1$, $na_n\not\to\infty$ as $n\to\infty$. A kernel density estimate constructed with this choice of the bandwidth sequence will not be consistent. Nevertheless, since $b_n\to 0$ as $n\to\infty$ is both necessary and sufficient for $\left\{W\left(\frac{t}{b_n}\right)\right\}$ to be a Heaviside sequence, the choice of $\{a_n\}$ indicated above when W has compact support will still give rise to consistent estimates $\hat{F}_n$.

(d) The difference between $\hat{Z}_n$ and Z_n^* lies in the bias term. Under additional assumptions on F and $\{W_n\}$, the bias can be made to decrease at a certain rate. (See Winter [30], Singh [22] and Reiss [16].) Similarly, the mean square error issue can be analysed just as the density estimate case. (See Azzalini [1].) However, Read [15] has shown that the standard empirical d.f. F_n is inadmissible w.r.t. integrated square loss by exhibiting a (biased) continuous piecewise linear estimate of F dominating F_n. It would be of interest to investigate if versions of $\hat{F}_n$ exist which dominate F_n.

3. Open Questions

There are obviously many interesting open questions regarding the statistical and probabilistic behaviour of $\hat{F}_n$ or $\hat{Z}_n$. We list a few here, hoping that there will be further research work on this topic:

(i) What kinds of optimality properties can $\hat{F}_n$ inherit from F_n?

(ii) Is there a Hájek-Beran type representation theorem for $\hat{Z}_n$ in the context of regular estimators?

(iii) What is the Prokhorov distance between $\hat{F}_n$ and F? (This has relevance in robustness.)

(iv) Would similar smoothing procedures be useful in the setting with censoring? (i. e., replace F_n by the Kaplan-Maier product limit function.)

(v) How does $\hat{F}_n$ compare with other possible competitors? For instance, how does (1) or (2) compare with the estimates constructed by orthogonal series expansions as in Kronmal and Tarter [12]?

(vi) How do bootstrap estimates behave when the bootstrap sample is generated according to $\hat{F}_n$ rather than F_n?

Acknowledgement

I would like to thank Professor T. Gasser for making my visit to Heidelberg a rewarding experience. I would like also to thank Professor Jon Wellner for many useful discussions on the oscillation of the empirical processes.

References

[1] AZZALINI, A.: *A Note on the Estimation of a Distribution Function and Quantiles by a Kernel Method*, Biometrika, **68** (1981), 326-328.

[2] BILLINGSLEY, I.: Convergence of Probability Measures, Wiley, (1968).

[3] BLUM, J. and WALTER, G.: *A Simple Solution to a Non-Parametric Maximum Likelihood Estimation Problem*, Technical report, (1983).

[4] BONEVA, L. I., KENDALL, D. G. and STEFANOV, I.: *Spline Transformations*, JRSS, **B33**, (1971), 1-70.

[5] BOOS, D. D. and MONAHAN, J. F.: *A Robust Bayesian Approach Using Bootstrapped Likelihoods*, Technical report, (1982).

[6] CSÖRGŐ, M. and RÉVÉSZ, P.: Strong Approximations in Probability and Statistics, Academic Press, (1981).

[7] EFRON, B.: *Bootstrap Methods: Another Look at the Jackknife*. Ann. Statist., **7** (1979), 1-26.

[8] GAENSSLER, P. and STUTE, W.: *Empirical Processes: A Survey of Results for Independent and Identically Distributed Random Variables*. Ann. Prob., **7** (1979), 193-243.

[9] GEMAN, S. and HUANG, C.-R.: *Non-Parametric Maximum Likelihood Estimation by the Method of Sieves*. Ann. Statist., **10** (1982), 401-414.

[10] HALL, P.: *Laws of the Iterated Logarithm for Non-Parametric Density Estimators*, Z. Wahr. Verw. Geb., **56** (1981), 47-61.

[11] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G.: *An Approximation of Partial Sums of Independent Random Variables, and the Sample Distribution Function, I*, Z. Wahr. Verw. Geb., **32** (1975), 111-131.

[12] KRONMAL, R. and TARTER, M.: *The Estimation of Probability Densities and Cumulatives by Fourier Series Methods*, JASA, **63** (1968), 925-952.

[13] LII, K. S. and ROSENBLATT, M.: *Asymptotic Behavior of Spline Estimates of a Density Function*, Comp. Math. Appl., **1** (1975), 223-235.

[14] NADARAYA, E. A.: *Some New Estimates for Distribution Functions*, Th. Prob. Appl., **15** (1964), 497-500.

[15] READ, R. R.: *The Asymptotic Inadmissibility of the Sample Distribution Function*, Ann. Math. Statist., **43** (1972), 89-95.

[16] REISS, R.-D.: *Nonparametric Estimation of Smooth Distribution Functions*, Scand. J. Statist. **8** (1981), 116-119.

[17] ROSENBLATT, M.: *Remarks on Some Nonparametric Estimation of a Density Function*, Ann. Math. Statist., **27** (1956), 832-837.

[18] SCHUSTER, E. F.: *On the Rate of Convergence of an Estimate of a Functional of a Pro-*

bability Density, Scand. Actuar. J., (1974), 103-107.
[19] SCHWARTZ, S. C.: *Estimation of Probability Density by an Orthogonal Series*, Ann. Math. Statist., **38** (1967), 1261-1265.
[20] SHORACK, G. and WELLNER, J. A.: *Limit Theorems and Inequalities for the Uniform Empirical Process Indexed by Sets*, Ann. Prob., **10** (1982), 639-652.
[21] SILVERMAN, B. W.: *Density Estimation for Univariate and Bivariate Data*, In Interpreting Multivariate Data, ed. V. Barnett, Wiley, (1981).
[22] SINGH, R. S.: *Uniform Almost Sure and Mean Square Convergence of Nonparametric Estimates of Distributions with Rates*, Technical report, (1980).
[23] SMAGA, E.: *Gtadka Dystrybuanta Empiryczna*, Przeglad Statystyczny, **25** (1978), 121-131. (In Polish.)
[24] STUTE, W.: *The Oscillation Behaviour of Empirical Processes*, Ann. Prob., **10** (1982), 86-107.
[25] WAHBA, G.: *Interpolating Spline Methods for Density Estimation I. Equi-Spaced Knots*, Ann. Statist., **3** (1975), 30-48.
[26] WALTER, G.: *Properties of Hermite Series Estimates of Probability Densities*, Ann. Statist., **5** (1977), 1258-1264.
[27] WALTER, G. and BLUM, J.: *Probability Density Estimation Using Delta Sequences*, Ann. Statist., **7** (1979), 328-340.
[28] WATSON, G. S. and LEADBETTER, M. R.: *Hazard Analysis II*, Sankhyā, **A26**, (1964), 101-116.
[29] WINTER, B. B.: *Strong Uniform Consistency of Integrals of Density Estimators*, Can. J. Statist., **1** (1973), 247-253.
[30] WINTER, B. B.: *Convergence Rate of Perturbed Empirical Distribution Functions*, J. Appl. Prob., **16** (1979), 163-173.
[31] YAMATO, H.: *Some Statistical Properties of Estimators of Density and Distribution Functions*, Bull. Math. Statist., **14** (1972), 113-131.
[32] YAMATO, H.: *Uniform Convergence of an Estimator of a Distribution Function*, Bull. Math. Statist., **15** (1973), 69-78.

Communicated by S. Kano
Received December 24, 1982
Revised December 1, 1983