ON THE LIMITING BEHAVIOUR OF THE EMPIRICAL KERNEL DISTRIBUTION FUNCTION

by

Pranab Kumar Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1405

July 1982

ON THE LIMITING BEHAVIOUR OF THE EMPIRICAL KERNEL DISTRIBUTION FUNCTION *

By PRANAB KUMAR SEN
University of North Carolina, Chapel Hill.

For an estimable parameter of degree $m(\geq 1)$, Glivenko-Cantelli lemma type result for the empirical kernel distribution and weak convergence of the related empirical process are studied. Some statistical applications of these results are also considered.

AMS Subject Classifications : 60F17, 62G99

Key Words & Phrases : Blackman-type estimator; degree; Glivenko-Cantelli lemma; kernel; reverse (sub-)martingales; U-statistics; weak convergence.

* Work partially supported by the National Heart, Lung and Blood Institute, Contract NIH-NHLBI-71-2243-L from the National Institutes of Health.

1. Introduction. Let $\{X_i, i \geq 1\}$ be a sequence of independent and identically distributed random vectors (i.i.d.r.v.) with a distribution function (d.f.) F, defined on the real p-space E^p, for some $p \geq 1$. Consider a functional $\theta(F)$ of the d.f. F, for which there exists a function $g(x_1,\ldots,x_m)$, such that

$$(1.1) \qquad \theta(F) = Eg(X_1,\ldots,X_m) = \int\ldots\int g(x_1,\ldots,x_m)dF(x_1)\ldots dF(x_m),$$

for every F belonging to a class $\mathfrak{F}$ of d.f.'s on E^p. Without any loss of generality, we may assume that g(.) is a symmetric function of its m arguments. If $m(\geq 1)$ is the minimal sample size for which (1.1) holds, then $g(X_1,\ldots,X_m)$ is called the kernel and m the degree of $\theta(F)$, and an optimal (symmetric) estimator of $\theta(F)$ is the U-statistic [viz., Hoeffding (1948)]

$$(1.2) \qquad U_n = \binom{n}{m}^{-1} \Sigma_{C_{n,m}} g(X_{i_1},\ldots,X_{i_m}) \; ; \; C_{n,m} = \{ 1 \leq i_1 <\ldots< i_m \leq n\},$$

whenever $n \geq m$. Let us assume that the kernel g(.) is real-valued and denote by

$$(1.3) \qquad H(y) = P\{ g(X_1,\ldots,X_m) \leq y\} , \; y \in E.$$

We are primarily interested in the estimation of the d.f. H(y) in (1.3). Note that $\theta(F)$ $(= \int y dH(y))$ is also a functional of the d.f. H. Analogous to (1.2), we may consider the following estimator of H (to be termed the empirical kernel distribution function (e.k.d.f.)):

$$(1.4) \qquad H_n(x) = \binom{n}{m}^{-1} \Sigma_{C_{n,m}} I(g(X_{i_1},\ldots,X_{i_m}) \leq x) , \; x \in E , \; n \geq m.$$

Note that like U_n in (1.2), H_n, for $m \geq 2$, does not involve independent summands, and hence, the classical results on the asymptotic properties of the sample d.f. may not be directly applicable to H_n. Nevertheless, like the U_n, such asymptotic results can be derived by using some (reverse) sub-martingale theory. The main objective of the present study is to consider the e.k.d.f. H_n and the related empirical process $\{n^{\frac{1}{2}}\{H_n(x) - H(x)\}, x \in E\}$ and to study their asymptotic behaviour.

Section 2 is devoted to the study of the Glivenko-Cantelli type almost sure (a.s.) convergence result on $H_n - H$. The weak convergence of the empirical process $n^{\frac{1}{2}}(H_n - H)$ is studied in Section 3. The concluding section deals with some statistical applications of the results of Sections 2 and 3.

2. Glivenko-Cantelli lemma for H_n

Note that by (1.3) and (1.4),

(2.1) $EH_n(x) = H(x)$, for every x .

We are interested in showing that

(2.2) $\sup\{ |H_n(x) - H(x)| : x \in E \} \rightarrow 0$ a.s., as $n \rightarrow \infty$.

Towards this, let $\mathcal{C}_n$ be the sigma-field generated by the unordered collection $\{X_1,\ldots,X_n\}$ and $X_{n+j}, j \geq 1$, for $n \geq 1$. Note that $\mathcal{C}_n$ is monotone nonincreasing.

Lemma 2.1. $\{ \sup_{x\varepsilon E}|H_n(x) - H(x)| , \mathcal{C}_n ; n \geq m\}$ is a reverse sub-martingale.

Proof. Note that for every $x \varepsilon E$, $n \geq m$,

(2.3) $E[H_n(x) \mid \mathcal{C}_{n+1}] = \binom{n}{m}^{-1} \Sigma_{C_{n,m}} E[I(g(X_{i_1},\ldots,X_{i_m}) \leq x) \mid \mathcal{C}_{n+1}]$

Now, given $\mathcal{C}_{n+1}$, $X_{i_1},\ldots,X_{i_m}$ can be any m of the units $X_1,\ldots,X_{n+1}$ with the equal conditional probability $\binom{n+1}{m}^{-1}$, so that for every $1 \leq i_1 < \ldots < i_m \leq n$,

(2.4) $E[I(g(X_{i_1},\ldots,X_{i_m}) \leq x) \mid \mathcal{C}_{n+1}] = E[I(g(X_1,\ldots,X_m) \leq x) \mid \mathcal{C}_{n+1}]$

$= \binom{n+1}{m}^{-1} \Sigma_{C_{n+1,m}} I(g(X_{j_1},\ldots,X_{j_m}) \leq x) = H_{n+1}(x)$.

Thus, by (2.3) and (2.4), for every $n \geq m$,

(2.5) $E[\{H_n(x) - H(x)\}, x \varepsilon E \mid \mathcal{C}_{n+1}] = \{H_{n+1}(x) - H(x)\}, x \varepsilon E$. (a.e.)

Since $\sup_x (\,.)$ is a convex functional, (2.5) insures the reverse sub-martingale property. Q.E.D.

Let now $n^* = [n/m]$ be the largest integer contained in n/m , for $n \geq m$. Also, for every $n \geq m$, let

(2.6) $H_n^*(x) = (n^*)^{-1} \Sigma_{i=1}^{n^*} I(g(X_{(i-1)m+1},\ldots,X_{im}) \leq x)$, $x \varepsilon E$.

Note that H_n^* involves independent summands, and, as in (2.4),

(2.7) $E[H_n^*(x) \mid \mathcal{C}_n] = H_n(x)$, for every $x \varepsilon E$.

By Lemma 2.1, (2.7) and the Kolmogorov inequality for reverse submartingales, we obtain that for every $\varepsilon > 0$, $n \geq m$,

$$P\{\sup_{N \geq n} \sup_{x \in E} |H_N(x) - H(x)| \geq \varepsilon\}$$

$$\leq \varepsilon^{-1} E\{\sup_{x \in E} |H_n(x) - H(x)|\}$$

(2.8)
$$= \varepsilon^{-1} E\{\sup_{x \in E} |E[H_n^*(x) - H(x) \mid \mathcal{C}_n]|\}$$

$$\leq \varepsilon^{-1} E\{E[\sup_{x \in E} |H_n^*(x) - H(x)| \mid \mathcal{C}_n]\}$$

$$= \varepsilon^{-1} E\{\sup_{x \in E} |H_n^*(x) - H(x)|\}.$$

Now, $H_n^* - H$ relates to the classical case of independent summans for which the results of Dvoretzky, Kiefer and Wolfowitz (1956) insure that for every $r \geq 0$,

(2.9) $$P\{(n^*)^{\frac{1}{2}} \sup_{x \in E} |H_n^*(x) - H(x)| \geq r\} \leq C\, e^{-2r^2}, \forall\, n^* \geq 1,$$

where C is a finite positive constant, independent of r and n^*. Since $n^* \sim n/m$, as $n \to \infty$, (2.9) insures that the right hand side of (2.8) converges to 0 as $n \to \infty$. This completes the proof of (2.2). We may also note that for every $n^* \geq 1$, $\{(n^*)^{\frac{1}{2}}\{H_n^*(x)-H(x)\}/\{1-H(x)\}, x \in E\}$ is a martingale, and hence, by the Hájek-Rényi-Chow inequality, it can be shown that the right hand side of (2.9) may be replaced by a more crude bound r^{-2}, for every $r \geq 1$, so that the convergence of the right hand side of (2.8) (to 0 as $n \to \infty$) remains in tact.

3. Weak convergence of $n^{\frac{1}{2}}(H_n - H)$. For the sake of simplicity, we assume that $H(x)$ is a continuous function of $x \in E$ and denote by $H^{-1}(t) = \inf\{x: H(x) \geq t\}$, $0 \leq t \leq 1$. For every $n (\geq m)$, we then introduce a stochastic process $W_n = \{W_n(t); 0 \leq t \leq 1\}$ by letting

(3.1) $$W_n(t) = n^{\frac{1}{2}}\{H_n(H^{-1}(t)) - t\},\ 0 \leq t \leq 1.$$

Then, W_n belongs to the space D[0,1]. We intend to study the weak convergence of W_n to some appropriate (tied-down) Gaussian function $W = \{W(t); 0 \leq t \leq 1\}$. Towards this, note that for arbitrary $r (\geq 1)$, $0 \leq t_1 < \ldots < t_r \leq 1$ and non-null $\underset{\sim}{\lambda} = (\lambda_1, \ldots, \lambda_r)'$, by (1.4) and (3.1),

$$\Sigma_{j=1}^{r} \lambda_j W_n(t_j) = n^{\frac{1}{2}} \Sigma_{j=1}^{r} \lambda_j [H_n(H^{-1}(t_j)) - t_j]$$

(3.2) $$= n^{\frac{1}{2}} \{\tilde{U}_n - E\tilde{U}_n\}, \text{ say,}$$

where

(3.3) $$\tilde{U}_n = \binom{n}{m}^{-1} \Sigma_{C_{n,m}} \tilde{\phi}(X_{i_1},\ldots,X_{i_m})$$

and

(3.4) $$\tilde{\phi}(X_{i_1},\ldots,X_{i_m}) = \Sigma_{j=1}^{r} \lambda_j \; I(g(X_{i_1},\ldots,X_{i_m}) \leq H^{-1}(t_j)) \; .$$

Thus, $\tilde{U}_n$ is a U-statistic and we may borrow the classical results of Hoeffding (1948) to show that the right hand side of (3.2) converges in law to a normal distribution with 0 mean and a finite variance (depending on $t_1,\ldots,t_m$ and $\tilde{\lambda}$). If we denote by

(3.5) $$\zeta_c(s,t) = P\{ g(X_1,\ldots,X_m) \leq H^{-1}(s),\; g(X_{m-c+1},\ldots,X_{2m-c}) \leq H^{-1}(t) \}- st,$$

for every $(s,t) \in [0,1]^2$ and $c=0,1,\ldots,m$ (note that $\zeta_0(s,t) = 0$), then,

(3.6) $$E\{[H_n(H^{-1}(s)) - s][H_n(H^{-1}(t)) - t]\}$$

$$= \binom{n}{m}^{-1} \Sigma_{c=1}^{m} \binom{m}{c}\binom{n-m}{m-c} \zeta_c(s,t) \text{ , for every } (s,t) \in [0,1]^2.$$

Note that by (3.1) and (3.6), for every $(s,t) \in [0,1]^2$, as $n \to \infty$,

(3.7) $$EW_n(s)W_n(t) \to m^2 \zeta_1(s,t) = \zeta(s,t), \text{ say.}$$

Thus, if we define a Gaussian function $W = \{W(t);0\leq t\leq 1\}$, **such** that $EW = 0$ and the covariance function of W is given by $\{ \zeta(s,t),(s,t) \in [0,1]^2\}$, then from the above discussion it follows that the finite dimensional distributions (f.d.d.) of $\{W_n\}$ converge to those of W. Further, $W_n(0) = 0$ with probsbility 1 and W belongs to the C[0,1] space , in probability. Hence, to establish the weak convergence of $\{W_n\}$ to W, it suffices to show that $\{W_n\}$ is tight. For this, it suffices to show that for every $0 \leq s_1<s_2< s_3\leq 1$, there exist an integer n_o and a finite positive constant K, such that

(3.8) $$E\{[W_n(s_2)- W_n(s_1)]^2[W_n(s_3) - W_n(s_2)]^2\} \leq K(s_2-s_1)(s_3- s_2), \forall n \geq n_o.$$

[See Theorem 15.6 of Billingsley (1968), in this context.] For this, we define $W_n^* = \{W_n^*(t);0\leq t\leq 1\}$ as in (3.1) with H_n being replaced by H_n^* . Then,

$$E\{[W_n(s_2)-W_n(s_1)]^2[W_n(s_3)-W_n(s_2)]^2\}$$
$$=E\{[E(W_n^*(s_2)-W_n^*(s_1)\mid \mathcal{C}_n)]^2[E(W_n^*(s_3)-W_n^*(s_2)\mid \mathcal{C}_n)]^2\}$$
(3.9)
$$\leq E\{E[(W_n^*(s_2)-W_n^*(s_1))^2\mid \mathcal{C}_n]E[(W_n^*(s_3)-W_n^*(s_2))^2\mid \mathcal{C}_n]\}$$
$$=E\{[W_n^*(s_2)-W_n^*(s_1)]^2[W_n^*(s_3)-W_n^*(s_2)]^2\}.$$

On the other hand, $W_n^*(.)$ involves n* independent summands, and hence, using the moment generating function of the multinomial distrubution, we obtain that

$$E\{[W_n^*(s_2)-W_n^*(s_1)]^2[W_n^*(s_3)-W_n^*(s_2)]^2\}$$
(3.10)
$$\leq 5(n/n^*)^2(s_2-s_1)(s_3-s_2)$$
$$\leq 5(m+1)^2(s_2-s_1)(s_3-s_2),\ \text{for every } 0\leq s_1<s_2<s_3\leq 1 \text{ and } n\geq m.$$

Thus, (3.8) follows from (3.9) and (3.10). Hence, we arrive at the following.

Theorem 3.1. W_n in (3.1) converges in law to the Gaussian function W with EW= 0 and covariance function $\zeta(s,t)$, given by (3.7).

Note that $\zeta(s,t)=0$ when s or t is equal to 0 or 1, and hence, W is tied-down at $t=0$ and $t=1$. However, for $m\geq 2$, in general, $\zeta(s,t)$ is not equal to $\min(s,t)-st$, so that W is not necessarily a Brownian bridge.

4. Some applications.

Let $X_1,\ldots,X_n$ be n i.i.d.r.v.'s with a d.f. $F(x)=F_o((x-\mu)/\sigma)$ where F_o is a specified d.f. and the location and scale parameters μ and σ are unknown. Blackman(1955) has considered the estimation of the location parameter (when σ is assumed to be specified) based on the empirical d.f. We consider here a similar estimator of σ^2 when μ is not specified. Note that if we let $g(X_i,X_j)=(X_i-X_j)^2/2$, then $Eg(X_i,X_j)=E(X_1-\mu)^2=\text{Var}(X)=c_o\sigma^2$, where c_o is a specifed positive constant and depends on the specified d.f. F_o. We assume that F_o admits a finite variance, and, without any loss of generality, we may set $c_o=1$. Thus, we have a kernel of degree 2 and the empirical kernel d.f. H_n may be defined as in (1.4) with m=2. We define H(y) as in (1.3) and since F_o is specified, we may rewrite H(y) as

(4.1) $$H(y)=H_o(y/\sigma^2),\ y\in E^+=[0,\infty),$$

where H_o depends on F_o and is of specified form too. Let then

(4.2) $M_n(t) = n \int_0^\infty [\, H_n(ty) - H_o(y)\,]^2 dH_o(y)\ , \ t \in E^+$.

As an estimator of $\theta = \sigma^2$, we consider $\hat{\theta}_n$ which is a solution of

(4.3) $M_n(\hat{\theta}_n) = \inf_t M_n(t)$.

Note that if we rewrite $M_n(t)$ as $\int_0^\infty \{n^{\frac{1}{2}}[H_n(ty) - H(ty)] + n^{\frac{1}{2}}[H_o(ty/\theta)-H_o(y)]\}^2 \cdot dH_o(y)$, then, for t away from θ , $M_n(t)$ blows up as $n \to \infty$, while, for t close to θ , we may proceed as in Pyke(1970) and through some routine steps obtain that as $n \to \infty$,

(4.4) $$n^{\frac{1}{2}}(\hat{\theta}_n - \theta) = \int_0^1 W_n(t)H_o^{-1}(t)(h_o(H_o^{-1}(t))dt \,/\int_0^1[H_o^{-1}(t)]^2[h_o(H_o^{-1}(t))]^2 dt + o_p(1)$$

where h_o is the density function corresponding to H_o and $W_n(.)$ is defined as in (3.1). Hence, the asymptotic normality of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ can be obtained from Theorem 3.1 and (4.4). A similar treatment holds for other Blackman-type estimators of estimable parameters when the underlying d.f. is specified (apart from some unknown parameters).

In the context of tests of goodness of fit when some of the parameters are unknown, an alternative procedure may be suggested as follows. Corresponding to the unknown parameters, obtain the kernels and for these kernels, consider the corresponding empirical kernel d.f.'s. Then, a multivariate version of Theorem 3.1 may be employed for the goodness of fit problem, using either the Kolmogorov-Smirnov or the Cramer-von Mises' type statistics.

The theory also can be extended to the two-sample case on parallel lines.

REFERENCES

BILLINGSLEY, P. (1968). Convergence of Probability Measures. New York: Wiley.

BLACKMAN, J.(1955). On the approximation of a distribution function by an empiric distribution. Ann. Math. Statist. 26, 256-267.

DVORETZKY, A., KIEFER, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and the classical multinomial estimator. Ann. Math. Statist. 27, 642-669.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. Ann. Math. Statist. 19, 293-325.

PYKE, R. (1970). Asymptotic results for rank statistics. In Nonparametric Techniques in Statistical Inference (ed:M.L.Puri), New York: Cambridge Univ. Press, pp.21-37.