isid/ms/2004/08
March 26, 2004
http://www.isid.ac.in/~statmath/eprints

Moment inequalities for supremum of empirical processes of u-statistic structure and application to density estimation

B. L. S. Prakasa Rao

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

MOMENT INEQUALITIES FOR SUPREMUM OF EMPIRICAL PROCESSES OF U-STATISTIC STRUCTURE AND APPLICATION TO DENSITY ESTIMATION

B.L.S. PRAKASA RAO
Indian Statistical Institute, New Delhi

Abstract

We derive moment inequalities for the supremum of empirical processes of U-Statistic structure and give application to kernel type density estimation and estimation of the distribution function for functions of observations.

Mathematics Subject classification : Primary 62G07.

Keywords and Phrases : ; Moment inequality; Empirical process of U-statistic structure; Kernel type density estimation; Functions of random variables. [1]

1 Introduction

Moment inequalities for the supremum of empirical processes with applications to kernel type estimation of a density function and a distribution function for identically distributed observations were investigated by Ahmad (2002) and by Prakasa Rao (2003) for ϕ-mixing stationary processes. Kernel type density estimation was extensively discussed in Prakasa Rao (1983). Let $X_n, n \geq 1$ be independent and identically distributed (i.i.d.) random variables and for a given integer $m \geq 1$, $g : R^m \to R$ be a real-valued function symmetric in its arguments. Motivated by examples to study the measure of degree of spatial randomness in a spatial point pattern (cf. Diggle (1983)) through study of the distribution function of the inter-point distances with $g(x_1, x_2) = |x_1 - x_2|$ with $m = 2$ and to study the distribution of sum of insurance claims with $g(x_1, \ldots, x_m) = x_1 + \ldots + x_m$ for some fixed $m \geq 1$, such as, for instance, in health insurance (cf. Panjer and Willmot (1992)), Frees (1994) investigated the problem of estimation of the density function of the random variable $g(X_1, \ldots, X_m)$ using the kernel method of density estimation. We will derive some moment inequalities for the deviation of this density estimator from the true density of the function $g(X_1, \ldots, X_m)$ assuming that it exists and obtain the optimal rate for the bandwidth in the sense of the supremum norm. Ahmad and Fan (2001) obtained optimal bandwidths, optimal in the sense of minimization of the asymptotic mean integrated squared error, for the kernel density estimator proposed by Frees (1994).

[1]Corresponding address: Indian Statistical Institute, 7,S.J.S.Sansanwal Marg, New Delhi 110016, India; e-mail: blsp@isid.ac.in

2 Empirical process of U-Statistic structure

Let $X_n, n \geq 1$ be independent and identically distributed (i.i.d.) random variables and let $g : R^m \to R$ be a real-valued function. As a generalization of the usual empirical distribution function, we consider the empirical distribution function $H_n(t)$ defined by the random variables

$$\{g(X_{i_1}, \ldots, X_{i_m}) : 1 \leq i_1 < \ldots < i_m \leq n\},$$

that is, for $-\infty < t < \infty$,

$$H_n(t) = (n_{C_m})^{-1} \sum_c I(g(X_{i_1}, \ldots, X_{i_m}) \leq t) \tag{2. 1}$$

hereafter called the *empirical kernel distribution function* (EKDF). Here $I(.)$ denotes the indicator function of a set, $\{i_1, \ldots, i_m\}$ is an ordered subset of $\{1, \ldots, n\}$ and $\sum_c$ denotes that the sum is taken over all n_{C_m} such subsets. For a fixed t , $H_n(t)$ is a U-Statistic. Let $H(t)$ be the distribution function of $g(X_1, \ldots, X_m)$. The process

$$\{\sqrt{n}(H_n(t) - H(t)), n \geq 1\}$$

is called the *empirical process of U-Statistic structure*. Observe that this process reduces to the usual empirical process of the i.i.d. random variables $\{g(X_n), n \geq 1\}$ in case $m = 1$. Silverman (1976) established the weak convergence of the process. Sen (1983) discussed the properties of EKDF from the view point of U-Statistic theory and established Glivenko-Cantelli lemma type result for the EKDF and studied the weak convergence of the related empirical process as defined above. Serfling (1984) proved the central limit theorems for functionals of empirical distribution function of U-Statistic structure in his study of genaralized L-Statistics. Silverman (1983) extended these results by proving the weak convergence of the sequence of processes

$$\{\sqrt{n}(H_n(t) - H(t)), -\infty < t < \infty\}, n \geq 1$$

to a continuous Gaussian process in weighted sup-norm metrics in $D(-\infty, \infty)$. Dehling et al. (1987) obtained an almost sure approximation of the process $n(H_n - H)$ by a suitable Gaussian process with a remainder term of the order $O(n^{\frac{1}{2}-\lambda}), \lambda > 0$ extending the Kiefer'result for the empirical distribution function of i.i.d. random variables. As a consequence of Corollary 2 in Dehling et al. (1987), it follows that

$$D_n = \sup_{-\infty<t<\infty} |H_n(t) - H(t)| = O(n^{-1/2}(\log \log n)^{1/2}) \tag{2. 2}$$

almost surely.

Let $\alpha \equiv (\alpha(1), \ldots, \alpha(n))$ denote an arbitrary permutation of $(1, \ldots, n)$. Following the blocking arguments of Hoeffding in connection with his study of U-statistics, define i.i.d. random variables

$$Y_{j+1}^{\alpha} \equiv g(X_{\alpha(jm+1)}, \ldots, X_{\alpha(jm+m)}), j = 0, 1, \ldots, [\frac{n}{m}] - 1 \tag{2. 3}$$

where $[\frac{n}{m}]$ denotes the largest integer less than or equal to $\frac{n}{m}$. Then $Y_1^\alpha, \ldots, Y_{[\frac{n}{m}]}^\alpha$ are i.i.d. random variables with the distribution function H and the empirical distribution function

$$H_n^\alpha(t) = [\frac{n}{m}]^{-1} \sum_{j=1}^{[\frac{n}{m}]} I[Y_j^\alpha \leq t]. \tag{2. 4}$$

Furthermore, it is known that

$$H_n(t) = (n!)^{-1} \sum_\alpha H_n^\alpha(t) \tag{2. 5}$$

which is also known as the Hoeffding identity. Hence

$$H_n(t) - H(t) = (n!)^{-1} \sum_\alpha \{H_n^\alpha(t) - H(t)\}$$

which implies that

$$\sup_t |H_n(t) - H(t)| \leq (n!)^{-1} \sum_\alpha \sup_t |H_n^\alpha(t) - H(t)|.$$

Therefore

$$E(\sup_t |H_n(t) - H(t)|) \leq (n!)^{-1} \sum_\alpha E(\sup_t |H_n^\alpha(t) - H(t)|).$$

Since the random variables $X_n, n \geq 1$ are i.i.d. random variables. it follows that

$$E(\sup_t |H_n^\alpha(t) - H(t)|)$$

is the same for any permutation $\alpha \equiv (\alpha(1), \ldots, \alpha(n))$ of $(1, \ldots, n)$. Hence

$$E(\sup_t |H_n(t) - H(t)|) \leq E(\sup_t |H_n^\alpha(t) - H(t)|) \tag{2. 6}$$

for any permutation α of $(1, \ldots, n)$. Therefore, from the results given in the inequality 2.1 of Ahmad (2002) for the deviation of the empirical distribution function from the true distribution function for i.i.d. random variables, it follows that

$$E(\sup_t |H_n^\alpha(t) - H(t)|) \leq K_r[\frac{n}{m}]^{-1/2} \tag{2. 7}$$

for any integer $r \geq 1$ where

$$K_r = \{\frac{r}{2^{r/2}}\Gamma(r/2)\}^{1/r}. \tag{2. 8}$$

Furthermore, the inequality 2.2 of Ahmad (2002) implies that

$$E(\exp\{t\sqrt{[\frac{n}{m}]} \sup_y |H_n^\alpha(y) - H(y)|\}) \leq 1 + \sqrt{2\pi} t e^{2t^2} \tag{2. 9}$$

for every $n \geq m$. Combining the above facts, we have the following result for the EKDF H_n.

Theorem 2.1: Let $X_i, 1 \leq i \leq n$ be i.i.d. random variables and H_n be the empirical kernel distribution function corresponding to a kernel $g(x_1, \ldots, x_m)$ as defined earlier and let H be the distrbution function of $g(X_1, \ldots, X_m)$. Then, for every integer $r \geq 1$, and for all $n \geq m$,

$$E(\sup_t |H_n(t) - H(t)|) \leq K_r[\frac{n}{m}]^{-1/2} \tag{2. 10}$$

where K_r is an absolute constant as defined in (2.8) and for every $n \geq m$,

$$E(\exp\{t\sqrt{[\frac{n}{m}]}\sup_y |H_n(y) - H(y)|\}) \leq 1 + \sqrt{2\pi} t e^{2t^2}. \tag{2. 11}$$

3 Application to Density Estimation

Following the notation introduced in the previous section, suppose that the distribution function $H(t)$ of the random variable $g(X_1, \ldots, X_m)$ has the probability density function $h(t)$. Suppose that the function $h(t)$ has a continuous and bounded second derivative with

$$\sup_t |h''(t)| = C_h < \infty.$$

Frees (1994) studied the problem of estimation of the density function $h(t)$ by the kernel method of density estimation (cf. Prakasa Rao (1983)). Let $w(.)$ be a bounded symmetric probability density function with mean zero and finite variance σ_w^2. Further suppose that it is of bounded variation with total variation V_w.

The kernel type density estimator $h_n(t)$ of $h(t)$ introduced by Frees (1994) is given by

$$h_n(t) = ({}_nC_m b_n)^{-1} \sum_c w(\frac{t - g(X_{i_1}, \ldots, X_{i_m})}{b_n}) \tag{3. 1}$$

where b_n is the bandwidth such that $b_n \to 0$ as $n \to \infty$. Let

$$J_n = \sup_t |h_n(t) - h(t)|.$$

Applying the Taylor's expansion for the function $h(t - ub_n)$ around the point t and observing that $w(.)$ is a bounded symmetric probability density function with mean zero and finite variance, it is easy to check that

$$\begin{aligned} |h_n(t) - h(t)| &\leq |\frac{1}{b_n}\int_{-\infty}^{\infty} w(\frac{t-y}{b_n})dH_n(y) - \frac{1}{b_n}\int_{-\infty}^{\infty} w(\frac{t-y}{b_n})dH(y)| \\ &\quad + |\frac{1}{b_n}\int_{-\infty}^{\infty} w(\frac{t-y}{b_n})dH(y) - h(t)| \\ &\leq \frac{1}{b_n}\sup_{-\infty<y<\infty} |H_n(y) - H(y)|| \int_{-\infty}^{\infty} dw(\frac{t-y}{b_n})| + \frac{b_n^2}{2}\sigma_w^2 C_h \\ &\leq \frac{1}{b_n}\sup_{-\infty<y<\infty} |H_n(y) - H(y)| V_w + \frac{b_n^2}{2}\sigma_w^2 C_h \\ &\leq \frac{1}{b_n} D_n V_w + \frac{b_n^2}{2}\sigma_w^2 C_h \end{aligned} \tag{3. 2}$$

where D_n is as defined by (2.2). Hence

$$E(J_n) \leq \frac{1}{b_n}E(D_n)V_w + \frac{b_n^2}{2}\sigma_w^2 C_h. \tag{3. 3}$$

Applying the bound on $E(D_n)$ derived in Theorem 2.1, we have

$$E(J_n) \leq K_r\frac{1}{b_n}[\frac{n}{m}]^{-1/2}V_w + \frac{b_n^2}{2}\sigma_w^2 C_h. \tag{3. 4}$$

Choosing b_n such that

$$\frac{1}{b_n}[\frac{n}{m}]^{-1/2} = b_n^2,$$

that is,

$$b_n = [\frac{n}{m}]^{-1/6},$$

we get an optimum bound on $E(J_n)$ as far as the rate of convergence is concerned and for a such choice of the bandwidth b_n,

$$E(J_n) = O([\frac{n}{m}]^{-1/3}). \tag{3. 5}$$

Following equation (12) of Ahmad and Fan (2001), observe that the optimal bandwidth for minimizing the asymptotic mean integrated squared error (AMISE) of the density estimator $h_n(t)$ is of the order $(n_{C_m})^{-1/3}$ as opposed to the optimal bandwidth for minimizing the bound on the mean uniform deviation of the density estimator obtained above which is of the order $[\frac{n}{m}]^{-1/6}$. If $m = 1$, then these orders reduce to $n^{-1/3}$ and $n^{-1/6}$ respectively.

Let us now consider the problem of estimation of the functional

$$I(h) = \int_{-\infty}^{\infty} h^2(t)dt. \tag{3. 6}$$

An estimator of $I(h)$ is $I(h_n)$. Note that

$$\begin{aligned} |I(h_n) - I(h)| &= |\int_{-\infty}^{\infty}(h_n(t) - h(t))(h_n(t) + h(t))dt| \\ &\leq \int_{-\infty}^{\infty}|h_n(t) - h(t)||h_n(t) + h(t)|dt \\ &\leq \sup_{-\infty<t<\infty}|h_n(t) - h(t)|\int_{-\infty}^{\infty}|h_n(t) + h(t)|dt \\ &= 2\sup_{-\infty<t<\infty}|h_n(t) - h(t)| = 2J_n. \end{aligned}$$

The last equality follows from the fact that the functions $h_n(t)$ and $h(t)$ are probability density functions. Hence

$$\begin{aligned} E|I(h_n) - I(h)| &\leq 2[\frac{1}{b_n}E(D_n)V_w + \frac{b_n^2}{2}\sigma_w^2 C_h] \\ &\leq 2[K_r\frac{1}{b_n}[\frac{n}{m}]^{-1/2}V_w + \frac{b_n^2}{2}\sigma_w^2 C_h]. \end{aligned} \tag{3. 7}$$

If $b_n = [\frac{n}{m}]^{-1/6}$, then the above bound reduces to

$$E|I(h_n) - I(h)| = O([\frac{n}{m}]^{-1/3}). \tag{3. 8}$$

4 Application to Estimation of Distribution Function

Let us now consider the problem of estimating the distribution function $H(.)$ of $g(X_1, \ldots, X_m)$ based on the i.i.d. observations $\{X_i, 1 \leq i \leq n\}$ for a fixed integer $m \geq 1$. Let $R_n(x)$ be a sequence of distribution functions converging weakly to the distribution function $R(x)$ degenerate at zero such that

$$\sup_{-\infty<x<\infty} |R_n(x) - R(x)| = o(\delta_n) \tag{4. 1}$$

where $\delta_n \to 0$ as $n \to \infty$. Define

$$\hat{H}_n(t) = (n_{C_m})^{-1} \sum_c R_n(t - g(X_{i_1}, \ldots, X_{i_m})) \tag{4. 2}$$

Let

$$\begin{aligned} Z_n &= \sup_{-\infty<t<\infty} |\hat{H}_n(t) - H(t)| \\ &\leq \sup_{-\infty<t<\infty} |\hat{H}_n(t) - E\hat{H}_n(t)| + \sup_{-\infty<t<\infty} |E\hat{H}_n(t) - H(t)|. \end{aligned} \tag{4. 3}$$

But

$$\begin{aligned} \sup_{-\infty<t<\infty} |\hat{H}_n(t) - E\hat{H}_n(t)| &= \sup_{-\infty<t<\infty} |\int_{-\infty}^{\infty} R_n(t-y)dH_n(y) \\ &\quad - \int_{-\infty}^{\infty} R_n(t-y)dH(y)| \\ &= \sup_{-\infty<t<\infty} |\int_{-\infty}^{\infty} (H_n(y) - H(y))dR_n(t-y)| \\ &\leq D_n. \end{aligned} \tag{4. 4}$$

Therefore

$$Z_n \leq D_n + \sup_{-\infty<t<\infty} |E\hat{H}_n(t) - H(t)|. \tag{4. 5}$$

It can be checked that

$$\begin{aligned} \sup_{-\infty<t<\infty} |E\hat{H}_n(t) - H(t)| &\leq \sup_{-\infty<t<\infty} |\int_{-\infty}^{\infty} |R_n(t-y) - R(t-y)|h(y)dy \\ &\leq \sup_{-\infty<t<\infty} |R_n(t) - R(t)| \int_{-\infty}^{\infty} h(y)dy \\ &\leq \delta_n. \end{aligned} \tag{4. 6}$$

Hence

$$E(Z_n) \leq E(D_n) + \delta_n. \tag{4. 7}$$

Applying Theorem 2.1 , we get that

$$E(Z_n) \leq K_r[\frac{n}{m}]^{-1/2} + \delta_n. \tag{4. 8}$$

Remarks: It is interesting to note that the bounds obtained in Theorem 2.1 and hence the bounds given in the inequalities (3.4), (3.7) and (4.8) do not depend on the function $g(x_1, \ldots, x_m)$. Further more the results agree with the optimal rate for bandwidth obtained in Ahmad (2002) for the case $m = 1$, that is, for the standard problem of density estimation.

References

Ahmad, I.A. (2002) On moment inequalities of the supremum of empirical processes with applications to kernel estimation, *Statist. Probab. Lett.*, **57**, 215-220.

Ahmad, I.A. and Fan, Y. (2001) Optimal bandwidths for kernel density estimators for functions of observations, *Statist. Probab. Lett.*, **51**, 245-251.

Dehling, H., Denker, M., and Philipp, W. (1987) The almost sure invariance principle for the empirical process of U-Statistic structure, *Ann. Inst. Henri Poincare*, **23**, 121-134.

Diggle, P. (1983) *Statistical Analysis of Spatial Point Pattern*, Academic Press, New York.

Frees, E.W. (1994) Estimating densities of functions of observations, *J. Amer. Statist. Assoc.*, **89**, 517-525.

Panjer, H.H., and Willmot, G.E. (1992) *Insurance Risk Models*, Society of Actuaries, Schaumburg, Illinois.

Prakasa Rao, B.L.S. (1983) *Nonparametric Functional Estimation*, Academic Press, New York.

Prakasa Rao, B.L.S. (2003) Moment inequalities for supremum of empirical processes for ϕ-mixing sequences, *Communications in Statistics - Theory and Methods*, **32**, 1695-1701.

Sen, P.K. (1983) On the limiting behavior of the empirical kernel distribution function, *Calcutta Statistical Association Bulletin*, **32**, 1-8.

Serfling, R. (1984) Generlized $L-, M-$, and R- statistics, *Ann. Statist.*, **12**, 76-86.

Silverman, B.W. (1976) Limit theorems for dissociated random variables, *Adv. Appl. Prob.*, **8**, 806-819.

Silverman, B.W. (1983) Convergence of a class of empirical distribution functions of dependent random variables, *Ann. Probab.*, **11**, 745-751.