Kybernetika

Jana Jurečková; Pranab Kumar Sen

An extension of Billingsley's uniform boundedness theorem to higher-dimensional M-processes

Kybernetika, Vol. 23 (1987), No. 5, 382--387

Persistent URL: http://dml.cz/dmlcz/125881

KYBERNETIKA — VOLUME 23 (1987), NUMBER 5

AN EXTENSION OF BILLINGSLEY'S UNIFORM BOUNDEDNESS THEOREM TO HIGHER DIMENSIONAL M-PROCESSES

JANA JUREČKOVÁ, PRANAB KUMAR SEN*

For multi-dimensional M-processes, an extension of an uniform boundedness theorem of Billingsley is considered under regularity conditions weaker than the usual ones pertaining to weak convergence (and, in particular, tightness) of such processes. Useful applications of this theorem are also stressed.

1. INTRODUCTION

For random functions belonging to the $C[0, 1]$ or $D[0, 1]$ space, Theorem 12.2 of Billingsley's monograph [2] relates to a probability inequality for the supremum, and it plays a vital role in the proof of the tightness of these processes. In the context of tightness of multi-parameter processes (i.e., for random functions belonging to the $D[0, 1]^q$ space, for some $q \geqq 1$), various extensions of the Billingsley inequality have been considered by a host of workers (viz., Bickel and Wichura [1] and references cited therein). For robust estimation in general linear models (viz., Jurečková and Sen [5] and the references cited therein), it may be convenient to consider some general multi-parameter M-processes and to exploit their asymptotic linearity results in the study of the properties of the derived estimators. In this context, a basic requirement is the *uniform boundedness in probability* of such M-processes. Such a result can, of course, be derived through the weak convergence of such processes (viz., Jurečková and Sen [3], [4] and others). However, this may demand comparatively more stringent regularity conditions. For this reason, for a general class of multi-dimensional M-processes, an extension of Billingsley's uniform boundedness theorem is considered under less stringent regularity conditions, and applications of this result in statistical inference are stressed. Along with the preliminary notions, the main theorem is presented in Section 2. Applications are considered in the last section.

* Work of the second author was partially supported by the (U.S.) Office of Naval Research, Contract No. N 00014-83-K-0387.

2. THE MAIN RESULT

Keeping in mind the general setup of robust estimation in general linear models (viz [5], [6]), including the first and second order asymptotic distributional results, we consider the following random functions. Let $Y_1, \ldots, Y_n$ be n independent and identically distributed (i.i.d.) random variables with a distribution function (d.f.) F, defined on the real line $\mathbb{R}(=(-\infty, \infty))$. Consider a *score function* ψ: $\mathbb{R} \to \mathbb{R}$, such that

$$(1) \qquad \int_{\mathbb{R}} \psi(x)\,\mathrm{d}F(x) = 0\,,$$

and

$$(2) \qquad \psi(x) = \psi_1(x) + \psi_2(x)\,, \quad x \in \mathbb{R}\,,$$

where both ψ_1 and ψ_2 are monotone functions. Consider the process

$$(3) \qquad N_j(\boldsymbol{t}) = \sum_{i=1}^{n} c_{ij}[\psi(Y_i + n^{-1/2}\boldsymbol{c}_i'\boldsymbol{t}) - \psi(Y_i)]\,, \quad \boldsymbol{t} \in \mathbb{R}^p\,, \quad j = 1, \ldots, p\,,$$

where the $\boldsymbol{c}_i = (c_{i1}, \ldots, c_{ip})'$ are vectors of known (regression) constants, p is a positive integer and we assume that there exists a positive definite (p.d.) matrix $\boldsymbol{Q}$, such that

$$(4) \qquad n^{-1}\sum_{i=1}^{n} \boldsymbol{c}_i\boldsymbol{c}_i' = n^{-1}\boldsymbol{C}_n \to \boldsymbol{Q}\,, \quad \text{as} \quad n \to \infty\,.$$

Keeping (2) in mind, we denote by

$$(5) \qquad N_j^{(l)}(\boldsymbol{t}) = \sum_{i=1}^{n} c_{ij}[\psi_l(Y_i + n^{-1/2}\boldsymbol{t}'\boldsymbol{c}_i) - \psi_l(Y_i)]\,, \quad \boldsymbol{t} \in \mathbb{R}^p\,, \quad j = 1, \ldots, p\,; \quad l = 1, 2\,.$$

Under general regularity conditions, $\mathsf{E}N_j^{(l)}(\boldsymbol{t})$ can be approximated by the first term of its Taylor expansion, which is typically of the form:

$$(6) \qquad \mathsf{E}_j^{(l)}(\boldsymbol{t}) = n^{-1/2}\sum_{i=1}^{n} c_{ij}\boldsymbol{t}'\boldsymbol{c}_i\gamma_l\,,$$

where the functional $\gamma_l = \gamma(\psi_l, F)$ depends of ψ_l and F (viz., $\gamma_l = \int \psi_l'(x)\,\mathrm{d}F(x)$ or $\int f(x)\,\mathrm{d}\psi_l(x)$). The asymptotic behaviour of $N_j(\boldsymbol{t}) - \mathsf{E}_j(\boldsymbol{t})$ has been studied in a variety of contexts (in the form of asymptotic linearity) in [3], [4], [5], [6]. The current note provides a clarification of a technical point underlying these developments. Denote by

$$(7) \qquad N_j^{(0)}(\boldsymbol{t}) = N_j(\boldsymbol{t}) - \mathsf{E}_j(\boldsymbol{t})\,, \quad j = 1, \ldots, p, \boldsymbol{t} \in \mathbb{R}^p\,,$$

and

$$(8) \qquad N_j^{(l,0)}(\boldsymbol{t}) = N_j^{(l)}(\boldsymbol{t}) - \mathsf{E}_j^{(l)}(\boldsymbol{t})\,, \quad \boldsymbol{t} \in \mathbb{R}^p\,, \quad j = 1, \ldots, p\,, \quad l = 1, 2\,.$$

Finally, let $\boldsymbol{T}(=T_1 \times \ldots \times T_p)$ be the unit hypercube in $\mathbb{R}^p$ (i.e., $\boldsymbol{T} = [0, 1]^p$), and let

$$(9) \qquad N_*^{(0)} = \max_{1 \leq j \leq p} \sup_{\boldsymbol{t} \in \boldsymbol{T}} |N_j^{(0)}(\boldsymbol{t})|\,.$$

Our main result may then be stated as follows:

Theorem 1. If there exist positive constants $K^*(<\infty)$ and n_0, such that for all $\boldsymbol{t}, \boldsymbol{u} \in T$ (with $\boldsymbol{t} \leqq \boldsymbol{u}$, coordinatewise), $n \geqq n_0$ and every $\lambda > 0$,

$$\text{(10)} \quad \mathsf{P}\{|N_j^{(l,0)}(\boldsymbol{u}) - N_j^{(l,0)}(\boldsymbol{t})| \geqq \lambda\} \leqq K^* \lambda^{-2} \|\boldsymbol{u} - \boldsymbol{t}\|^2, \quad l = 1, 2; \ j = 1, \dots, p,$$

where $\|\cdot\|$ stands for the Euclidean norm, then there exists a positive constant L_p, such that for every $n \geqq n_0$,

$$\text{(11)} \qquad \mathsf{P}\{N_*^{(0)} \geqq \lambda\} \leqq L_p \lambda^{-2}, \quad \text{for every} \quad \lambda > 0.$$

Remarks. For $p = 1$, Theorem 1 is very much comparable with Theorem 12.2 of [2]. For $p \geqq 2$, an extension of this Billigsley-type inequality has been considered by Bickel and Wichura in [1]. They have considered the inequality in the context of tightness of D^p-valued processes, and for this comparatively stronger result, they have to consider Lebesgue (or other products) measures of *blocks* in T and their regularity conditions include an inequality similar to (10) but involving the increments of the process over blocks and the Euclidean distance $\|\boldsymbol{u} - \boldsymbol{t}\|$ being replaced by such product measures. If our main objective is to derive a probability inequality on the uniform boundedness of M-processes [i.e., (11)], then we may replace such product measures by a quadratic measure [as in (10)], so that the verification of this regularity condition may demand less restrictive (moment) conditions on the score function as well as the regression vectors $\boldsymbol{c}_i$. This is the main idea of the current study. We do not, however, claim that (10) suffices for the compactness of the M-processes under consideration.

Proof. We shall prove the theorem by induction (over $p = 1, 2, \dots$).

Step I. If $p = 1$, we have a one-parameter process, and (11) follows directly from Theorem 12.2 of Billingsley [2].

Step II. For $\boldsymbol{t} \in \mathbb{R}^p$, let us denote by

$$\text{(12)} \qquad N_j^{*0}(\boldsymbol{t}) = \sum_{\varepsilon_1 = 0,1} \dots \sum_{\varepsilon_p = 0,1} (-1)^{\varepsilon_1 + \dots + \varepsilon_p} N_j^0(t_1 - \varepsilon_1 t_1, \dots, t_p - \varepsilon_p t_p),$$

for $j = 1, \dots, p$. Then,

$$\text{(13)} \qquad N_j^0(\boldsymbol{t}) = N_j^{*0}(\boldsymbol{t}) + [N_j^0(\boldsymbol{t}) - N_j^{*0}(\boldsymbol{t})], \quad \boldsymbol{t} \in T,$$

where $N_j^0(\boldsymbol{t}) - N_j^{*0}(\boldsymbol{t})$ is a linear combination of restrictions of $N_j^0(\boldsymbol{t})$ to $(p-1)$ parameters (or less). Thus, by our induction hypothesis, $N_j^0(\boldsymbol{t}) - N_j^{*0}(\boldsymbol{t})$ satisfies (11). Hence, it suffices to establish (11) for the process $N_j^{*0}(\boldsymbol{t})$. Notice that $N_j^{*0}(\boldsymbol{t}) = 0$ if $t_r = 0$ for some $r = 1, \dots, p$, and this is one of the conditions in Theorem 1 of Bickel and Wichura [1] which we shall make use of. Also, we may note that for proving (11), it suffices to consider $\sup\{|N_r^{*0}(\boldsymbol{t})| : \boldsymbol{t} \in T\}$, for some r $(=1, \dots, p)$, and use the elementary inequality to extend the result for all $r = 1, \dots, p$. Hence, we specifically choose $r = 1$.

For fixed $s_1, t_1, u_1 \in T_1$ (with $s_1 \leqq t_1 \leqq u_1$), consider the processes

$$Y(\boldsymbol{t}_2) = N_1^{*0}(t_1, \boldsymbol{t}_2) - N_1^{*0}(s_1, \boldsymbol{t}_2), \tag{14}$$

$$Z(\boldsymbol{t}_2) = N_1^{*0}(u_1, \boldsymbol{t}_2) - N_1^{*0}(t_1, \boldsymbol{t}_2), \quad \boldsymbol{t}_2 = (t_2, \ldots, t_p)' \in T_2 \times \ldots \times T_p. \tag{15}$$

Then, by (10), for every $n \geqq n_0$ and $\lambda > 0$,

$$\mathsf{P}\{|Y(\boldsymbol{t}_2)| \geqq \lambda\} \leqq \lambda^{-2} K^{**}(t_1 - s_1)^2, \tag{16}$$

$$\mathsf{P}\{|Z(\boldsymbol{t}_2)| \geqq \lambda\} \leqq \lambda^{-2} K^{**}(u_1 - t_1)^2, \quad \forall \boldsymbol{t}_2 \in T_2, \tag{17}$$

where K^{**} is a finite positive constant (independent of n and λ). Notice that in (12)

$$N_1^{*0}(\boldsymbol{t}) = N_1^*(\boldsymbol{t}) = \sum_{\varepsilon_1=0,1} \ldots \sum_{\varepsilon_p=0,1} (-1)^{\varepsilon_1+\ldots+\varepsilon_p} N_1(t_1 - \varepsilon_1 t_1, \ldots, t_p - \varepsilon_p t_p).$$

Then, by (5), we may rewrite $N_1^*(\boldsymbol{t})$ as a linear combination of 2^p functions of the form (with coefficients $+1$ or -1):

$$\sum_{i=1}^{n} c_{i1}[\psi_l(Y_i + n^{-1/2}(c_{ij_1}t_{j_1} + \ldots + c_{ij_r}t_{j_r})) - \psi_l(Y_i)] \tag{18}$$

where r ranges over $(0, p)$, $l = 1, 2$, and $j_1, \ldots, j_r$ over all possible combinations of r indices from $(1, \ldots, p)$; for $r = 0$, we have a null subset. Note that by our assumption, the ψ_l are monotone. Also, the components of the vector

$$(c_{i1}c_{ij_1}, \ldots, c_{i1}c_{ij_r}) \tag{19}$$

have 2^r possible combination of signs. Hence, the index set $\mathbb{N} = \{i: 1 \leqq i \leqq N\}$ can be decomposed into 2^r subsets corresponding to the same combinations of signs. Thus, there exists a positive integer $M(\leqq 2^p(2^p - 1))$, such that $N_1^*(\boldsymbol{t})$ can be expressed as a signed-sum over M subsets of terms of the form (18) (with $i \in \mathbb{N}$ replaced by i belonging to each of these subsets); we denote these component sums by $A_1, \ldots, A_M$ respectively. Then, we may note that A_s is $\nearrow$ in t_{j_k} if $c_{i1}c_{ij_k}$ is positive and it is $\searrow$ in t_{j_k} if $c_{i1}c_{ij_k}$ is negative, for $k = 1, \ldots, r$; $s = 1, \ldots, M$. Hence, we may write

$$Y(\boldsymbol{t}_2) = Y_1(\boldsymbol{t}_2) + \ldots + Y_M(\boldsymbol{t}_2), \tag{20}$$

$$Z(\boldsymbol{t}_2) = Z_1(\boldsymbol{t}_2) + \ldots + Z_M(\boldsymbol{t}_2), \quad \boldsymbol{t}_2 \in T_2 \times \ldots \times T_p, \tag{21}$$

where $Y_l(\boldsymbol{t}_2)$ and $Z_l(\boldsymbol{t}_2)$ are monotone in all components of $\boldsymbol{t}_2$, for $l = 1, \ldots, M$. Now, for each j $(=2, \ldots, p)$, we consider a partitioning of T_j as $\bigcup_{s=0}^{k-1} (z_s, z_{s+1}]$, where $0 = z_0 < z_1 < \ldots < z_k = 1$. Thus, we would have a mesh of k^{p-1} grid-points on $T_2 \times \ldots \times T_p$, so that using the monotonicity property of each $Y_l(\boldsymbol{t}_2)$ (in the elements of $\boldsymbol{t}_2$), we readily obtain that

$$\begin{aligned} &\sup\{|Y(\boldsymbol{t}_2)| : \boldsymbol{t}_2 \in T_2 \times \ldots \times T_p\} \leqq \\ &\leqq \max\{|Y(z_{r_2}, \ldots, z_{r_p})| : 0 \leqq r_j \leqq k;\ j = 2, \ldots, p\} + \\ &+ 2\sum_{l=1}^{M} \max\{|Y_l(z_{r_2}, \ldots, z_{r_p})| : 0 \leqq r_j \leqq k;\ j = 2, \ldots, p\}. \end{aligned} \tag{22}$$

Note that $Y(z_{r_2}, \ldots, z_{r_p})$ or the $Y_l(z_{r_2}, \ldots, z_{r_p})$ vanishes when $z_{r_2} = \ldots = z_{r_p} = 0$. Let us order the grid-points $(z_{r_2}, \ldots, z_{r_p})$, $0 < r_j < k, j = 2, \ldots, p$ lexicographically and denote them by $\xi_1, \ldots, \xi_{M^*}$, where $M^* = (k+1)^{p-1} - 1$. Then, it follows from (16) and (22) that for every $\lambda > 0$,

$$(23) \qquad \mathsf{P}\{\sup\{|Y(t_2)| : t_2 \in T_2 \times \ldots \times T_p\} \geqq \lambda\} \leqq$$
$$\leqq \sum_{l=1}^{M^*} \mathsf{P}\{|Y(\xi_l)| > \lambda_1\} + \sum_{l=1}^{M} \sum_{v=1}^{M^*} \mathsf{P}\{|Y_l(\xi_v)| \geqq (2M)^{-1} \lambda_2\} \leqq$$
$$\leqq M^* K^{**}(t_1 - s_1)^2 (\lambda_1^{-2} + 4M^3 \lambda_2^{-2}), \quad \forall \lambda_1, \lambda_2 > 0 ; \quad \lambda_1 + \lambda_2 = \lambda .$$

It is easy to verify that for positive c_1, c_2,

$$(24) \qquad \min\{(c_1\lambda_1^{-2} + c_2\lambda_2^{-2}) : \lambda_1 + \lambda_2 = \lambda ; \lambda_1, \lambda_2 > 0\} = \lambda^{-2}(c_1^{1/3} + c_2^{1/3})^3 ,$$

so that by (23) and (24), we obtain that

$$(25) \qquad \mathsf{P}\{\sup\{|Y(t_2)| : t_2 \in T_2 \times \ldots \times T_p\} \geqq \lambda\} \leqq$$
$$\leqq \lambda^{-2} M^* K^{**}(1 + 4^{1/3}M)^3 (t_1 - s_1)^2 = K^0 \lambda^{-2}(t_1 - s_1)^2 ,$$

where K^0 is a finite positive constant. It follows similarly that for every $\lambda > 0$,

$$(26) \qquad \mathsf{P}\{\sup\{|Z(t_2)| : t_2 \in T_2 \times \ldots \times T_p\} \geqq \lambda\} \leqq K^0 \lambda^{-1}(u_1 - t_1)^2 .$$

By (14), (15), (25) and (26), we conclude that for every $0 \leqq s_1 \leqq t_1 \leqq u_1 \leqq 1$ and $\lambda > 0$, whenever $n \geqq n_0$,

$$(27)$$
$$\mathsf{P}\{\min\{\sup_{t_2} |N_1^{*0}(t_1, t_2) - N_1^{*0}(s_1, t_2)|, \sup_{t_2} |N_1^{*0}(u_1, t_2) - N_1^{*0}(t_1, t_2)|\} \geqq \lambda\} \leqq$$
$$\leqq K^0 \lambda^{-2}(u_1 - s_1)^2 .$$

Now, (27) corresponds to the last inequality in Step 5 of the proof of Theorem 1 of [1], and hence, we may as well use their inductive arguments and conclude that (11) follows from (27). This completes the proof of the theorem.

3. SOME GENERAL REMARKS

The main motivation of Theorem 1 stems out of first and second order asymptotic distributional properties of M-estimators of location and regression parameters (as have been studied in detail by the current authors and others). In this context, typically, we encounter an M-process involving the $N_j^0(t)$ in (3) and (6), where for the remainder term in such a representation, we typically need a uniform boundedness result, and this is provided by (11) under the assumption in (10). For square integrable and 'smooth' score functions, (10) can easily be verified by using the second moment of the two random functions involved. On the other hand, for $p \geqq 2$, verification of the basic condition of Theorem 1 of [1] would typically involve the computation of 4th moment (or $(2 + \delta)$th absolute moments, for some $\delta > 0$) of the $N_j^0(t)$, and this would in turn require more stringent regularity conditions of the score function ψ, the regression vectors $\boldsymbol{c}_i$ and the underlying density function f. In this way, the current theorem serves a very useful role in the asymptotic theory of M-estimators in linear models. It is, however, not necessary to have the squared Euclidean distance in (10). We may replace the right hand side of (10) by $K^*\lambda^{-2r}\|\boldsymbol{u} - \boldsymbol{t}\|^{2r}$, for some $r > \frac{1}{2}$, and in that case, in (11), we need to replace λ^{-2} by λ^{-2r}, $r > \frac{1}{2}$. The proof sketched in the earlier section remains the same under this modification. However, in actual practice, the case of $r = 1$ is the most commonly adapted one, and hence, this refinement is not of much significance. Finally, we may remark that the delicate treatment in (18) through (22) may not generally hold for D^p-processes, and hence, the current theorem is not advocated as a general replacement of Theorem 1 of [1]. Rather, it is proposed as a simplification in the commonly arisen cases where the score functions and related M-processes satisfy (10) under fairly simple setups, so that one does not need to verify the more stringent condition in [1]. In this context, it may be noted that in M-estimation, typically, the score function ψ is not linear, and hence, the computation of the increments over a block involves higher order differences, and thereby, demands extra regularity conditions.

ACKNOWLEDGEMENT

The authors are grateful to Dr. P. Lachout for a valuable discussion leading to an improved presentation of the result.

(Received March 5, 1987.)

REFERENCES

[1] P. Bickel and M. J. Wichura: Convergence criteria for multiparameter stochastic processes and some applications. Ann. Math. Statist. *42* (1971), 1656—1670.

[2] P. Billingsley: Convergence of Probability Measures. J. Wiley, New York 1968.

[3] J. Jurečková and P. K. Sen: Invariance principles for some stochastic processes related to M-estimators and their role in sequential statistical inference. Sankhya, Ser. A *43* (1981), 190—210.

[4] J. Jurečková and P. K. Sen: Sequential procedures based on M-estimators with discontinuous score functions. J. Statist. Plan Infer. *5* (1981), 253—266.

[5] J. Jurečková and P. K. Sen: On adaptive scale-equivariant M-estimators in linear models. Statist. & Dec., Suppl. Issue No. *1* (1984), 31—46.

[6] J. Jurečková and P. K. Sen: A second order asymptotic distributional representation of M-estimators with discontinuous score functions. Ann. Probab. *15* (1987), 814—823.

[7] A. H. Welsh: Bahadur representations for robust scale estimators based on regression residuals. Ann. Statist. *14* (1986), 1246—1251.

Doc. RNDr Jana Jurečková, CSc., Universita Karlova, fakulta matematicko-fyzikální (Charles University, Faculty of Mathematics and Physics) Sokolovská 83, 186 00 Praha 8. Czechoslovakia.
Prof. Pranab Kumar Sen, University of North Carolina, Department of Biostatistics, Chapell Hill, N.C. 27514. U.S.A.