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

Terms of use:

© Institute of Information Theory and Automation AS CR, 1987

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these Terms of use.

This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project DML-CZ: The Czech Digital Mathematics Library http://project.dml.cz
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 $q \geq 1$), 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 $\psi: \mathbb{R} \to \mathbb{R}$, such that

\begin{align*}
(1) & \int_{\mathbb{R}} \psi(x) \, dF(x) = 0, \\
(2) & \psi(x) = \psi_1(x) + \psi_2(x), \quad x \in \mathbb{R},
\end{align*}

where both $\psi_1$ and $\psi_2$ are monotone functions. Consider the process

\begin{equation}
N_j(t) = \sum_{i=1}^{n} c_i \left[ \psi(Y_i + n^{-1/2} t \epsilon_i) - \psi(Y_i) \right], \quad t \in \mathbb{R}^p, \quad j = 1, \ldots, p,
\end{equation}

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

\begin{equation}
n^{-1} \sum_{i=1}^{n} c_i c_i' = n^{-1} C \to Q, \quad n \to \infty.
\end{equation}

Keeping (2) in mind, we denote by

\begin{equation}
N_j^{(i)}(t) = \sum_{i=1}^{n} c_i \left[ \psi_i(Y_i + n^{-1/2} t \epsilon_i) - \psi_i(Y_i) \right], \quad t \in \mathbb{R}^p, \quad j = 1, \ldots, p;
\end{equation}

where both $\gamma_i = \gamma(\psi_i, F)$ depends of $\psi_i$ and $F$ (viz., $\gamma_i = \int \psi_i(x) \, dF(x)$). The asymptotic behaviour of $N_j(t) - E_j(t)$ 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

\begin{equation}
N_j^{(0)}(t) = N_j(t) - E_j(t), \quad j = 1, \ldots, p, \quad t \in \mathbb{R}^p,
\end{equation}

and

\begin{equation}
N_j^{(i)(0)}(t) = N_j^{(i)}(t) - E_j^{(i)}(t), \quad t \in \mathbb{R}^p, \quad j = 1, \ldots, p, \quad i = 1, 2.
\end{equation}

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

\begin{equation}
N^*_\epsilon = \max_{1 \leq j \leq p} \sup_{t \in T} |N_j^{(0)}(t)|.
\end{equation}
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 $t, u \in T$ (with $t \leq u$, coordinatewise), $n \geq n_0$ and every $\lambda > 0$,

$$P(\|N^{(1,0)}(u) - N^{(1,0)}(t)\| \geq \lambda) \leq K^* \lambda^{-2} \|u - t\|^2, \quad l = 1, 2; \quad f = 1, \ldots, p,$$

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

$$P(N^{(0)} \geq \lambda) \leq L_p \lambda^{-2}, \quad \text{for every } \lambda > 0.$$

**Remarks.** For $p = 1$, Theorem 1 is very much comparable with Theorem 12.2 of [2]. For $p \geq 2$, an extension of this Billingsley-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 $\|u - t\|$ 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 $c_t$. 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 $p = 1, 2, \ldots$).

**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 $t \in R^p$, let us denote by

$$N^{(0)}(t) = \sum_{t_1 = 0, 1} \ldots \sum_{t_p = 0, 1} (-1)^{t_1 + \ldots + t_p} N^{(0)}(t_1 - \varepsilon t_1, \ldots, t_p - \varepsilon t_p),$$

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

$$N^{(j)}(t) = N^{(0)}(t) + [N^{(j)} - N^{(0)}(t)], \quad t \in T,$$

where $N^{(j)}(t) - N^{(0)}(t)$ is a linear combination of restrictions of $N^{(j)}(t)$ to $(p - 1)$ parameters (or less). Thus, by our induction hypothesis, $N^{(j)}(t) - N^{(0)}(t)$ satisfies (11). Hence, it suffices to establish (11) for the process $N^{(0)}(t)$. Notice that $N^{(0)}(t) = 0$ if $t, = 0$ for some $r = 1, \ldots, p$, 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 $\{N^{(0)}(t) : t \in T \}$, for some $r = 1, \ldots, p$, and use the elementary inequality to extend the result for all $r = 1, \ldots, p$. Hence, we specifically choose $r = 1$. 384
For fixed $s, h, u \in \mathcal{U}$ (with $s t = u y = U_j$), consider the processes

\begin{equation}
Y(t_2) = N_t^0(t_1, t_2) - N_t^0(t_1, t_2),
\end{equation}

\begin{equation}
Z(t_2) = N_t^0(u_1, t_2) - N_t^0(t_1, t_2),
\end{equation}

$t_2 = (t_2, \ldots, t_p) \in T_2 \times \ldots \times T_p$.

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

\begin{equation}
P(|Y(t_2)| \geq \lambda) \leq \lambda^{-2} K^{**}(t_1 - s),
\end{equation}

\begin{equation}
P(|Z(t_2)| \geq \lambda) \leq \lambda^{-2} K^{**}(u_1 - t_1), \quad \forall t_2 \in T_2,
\end{equation}

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

\[ N_t^0(t) = N_t^0(t) - \sum_{i=0}^{n-1} \sum_{r \in \mathcal{N}} (-1)^{i+r} N_t(t - e_i, \ldots, t_p - e_p).
\]

Then, by (5), we may rewrite $N_t^0(t)$ as a linear combination of $2^p$ functions of the form (with coefficients +1 or -1):

\begin{equation}
\sum_{i=1}^{n} c_i [\psi_i(Y_i + n^{-1/2}(c_{i_1} t_{i_1} + \ldots + c_{i_p} t_{i_p}))) - \psi_i(Y_i)]
\end{equation}

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 $\psi_i$ are monotone. Also, the components of the vector

\begin{equation}
(c_{i_1} c_{i_2}, \ldots, c_{i_p} c_{i_p})
\end{equation}

have $2^p$ possible combination of signs. Hence, the index set $\mathcal{N} = \{i: 1 \leq i \leq N\}$ can be decomposed into $2^p$ subsets corresponding to the same combinations of signs. Thus, there exists a positive integer $M(\leq 2^p - 1)$, such that $N_t^0(t)$ can be expressed as a signed-sum over $M$ subsets of terms of the form (18) (with $i \in \mathcal{N}$ 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 a $\psi_i$ in $t_{i_s}$ if $c_{i_s} c_{i_s}$ is positive and it is a $\psi_i$ in $t_{i_s}$ if $c_{i_s} c_{i_s}$ is negative, for $k = 1, \ldots, r; s = 1, \ldots, M$. Hence, we may write

\begin{equation}
Y(t_2) = Y_1(t_2) + \ldots + Y_M(t_2),
\end{equation}

\begin{equation}
Z(t_2) = Z_1(t_2) + \ldots + Z_M(t_2),
\end{equation}

$t_2 \in T_2 \times \ldots \times T_p$,

where $Y_i(t_2)$ and $Z_i(t_2)$ are monotone in all components of $t_2$, for $l = 1, \ldots, M$.

Now, for each $j = 2, \ldots, p$, we consider a partitioning of $T_j$ as $\mathcal{U}(z_r, z_{r+1})$, where $0 = z_0 < z_1 < \ldots < z_p = 1$. 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 $Y_i(t_2)$ (in the elements of $t_2$), we readily obtain that

\begin{equation}
\sup \{|Y(t_2)| : t_2 \in T_2 \times \ldots \times T_p\} \leq
\end{equation}

\begin{equation}
\leq \max \{|Y(z_{r_1}, \ldots, z_{r_p})| : 0 \leq r_j \leq k ; \quad j = 2, \ldots, p\} +
+ 2 \sum_{i=1}^{M} \max \{|Y_i(z_{r_1}, \ldots, z_{r_p})| : 0 \leq r_j \leq k ; \quad j = 2, \ldots, p\}.
\end{equation}
Note that \( Y(z_{r_2}, ..., z_{r_p}) \) or the \( Y_{r_j} \) vanishes when \( z_{r_2} = ... = z_{r_p} = 0 \). Let us order the grid-points \((z_{r_1}, ..., z_{r_p})\), \(0 < r_j < k, j = 2, ..., p\) lexicographically and denote them by \( \xi_1, ..., \xi_{M^*} \), where \( M^* = (k + 1)^{p-1} - 1 \). Then, it follows from (16) and (22) that for every \( \lambda > 0 \),

\[
P\{ \sup_{t_2 \in T_2} |Y(t_2)| > \lambda \} \leq \sum_{i=1}^{M^*} P\{ |Y(\xi_i)| > \lambda_1 \} + \sum_{i=1}^{M^*} \sum_{j=1}^{M^*} P\{ |Y(\xi_i)| \geq (2M)^{-1} \lambda_2 \} \leq M^*K^*(t_1 - s_1)^2 (\lambda_1^{-2} + 4M^2\lambda_2^{-2}), \quad \forall \lambda_1, \lambda_2 > 0; \quad \lambda_1 + \lambda_2 = \lambda.
\]

It is easy to verify that for positive \( c_1, c_2 \),

\[
\min \{ (c_1\lambda_1^{-2} + c_2\lambda_2^{-2}) \lambda_1 + \lambda_2 = \lambda \} = \lambda^{-2}(c_1^{1/3} + c_2^{1/3})^3,
\]

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

\[
P\{ \sup_{t_2 \in T_2} |Y(t_2)| \geq \lambda \} \leq \lambda^{-2}M^*K^*(1 + 4^{1/2}M)^3 (t_1 - s_1)^2 = \mathcal{K}^0\lambda^{-2}(t_1 - s_1)^2,
\]

where \( \mathcal{K}^0 \) is a finite positive constant. It follows similarly that for every \( \lambda > 0 \),

\[
P\{ \sup_{t_2 \in T_2} |Z(t_2)| \geq \lambda \} \leq \mathcal{K}^0\lambda^{-2}(u_1 - t_1)^2.
\]

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

\[
P\{ \min_{t_2} \{ \sup_{t_2 \in T_2} |N_{11}^o(t_1, t_2) - N_{11}^o(s_1, t_2)| \}, \sup_{t_2 \in T_2} |N_{12}^o(t_1, t_2)| \geq \lambda \} \leq \mathcal{K}^0\lambda^{-2}(u_1 - s_1)^2.
\]

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 \( N_0^o(t) \) 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 \( p \geq 2 \), 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 \( N_0^o(t) \), and this would in turn require more stringent regularity conditions of the score function \( \psi \),
the regression vectors $c_t$ 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 $K^2 \sum \left( w - f \right)^2$, for some $r > \frac{1}{2}$, and in that case, in (11), we need to replace $\lambda^{-2}$ by $\lambda^{-2r}$, $r > \frac{1}{2}$. 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 $D^r$-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 $\psi$ 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


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, Chapel Hill, N.C. 27514. U.S.A.