Stochastic Processes and their Applications 11 (1981) 309–312
North-Holland Publishing Company

SHORT COMMUNICATION

A GLIVENKO–CANTELLI THEOREM FOR EMPIRICAL MEASURES OF INDEPENDENT BUT NON-IDENTICALLY DISTRIBUTED RANDOM VARIABLES

Jon A. WELLNER*
Department of Statistics, University of Rochester, Rochester, NY 14627, U.S.A.

Received 10 December 1979
Revised 6 June 1980

The bounded-dual-Lipschitz and Prohorov distances from the 'empirical measure' to the 'average measure' of independent random variables converges to zero almost surely if the sequence of average measures is tight. Three examples are also given.

AMS Subj. Class.: 60F15, 60B10
Bounded-dual-Lipschitz metric
strong law
Prohorov metric
convergence

1. The theorem

Let (S, d) be a separable metric space; let $\mathscr{P}(S)$ be the set of all Borel probability measures on S; and let $X_1, X_2, \ldots$ be independent S-valued random variables with distributions $P_1, P_2, \ldots$ where all $P_n \in \mathscr{P}(S)$. For $x \in S$ let δ_x be the unit mass at x. For $n \geq 1$ define the 'empirical measure' $\mathbb{P}_n$ by

$$\mathbb{P}_n \equiv (\delta_{X_1} + \cdots + \delta_{X_n})/n$$

and the 'average measure' $\bar{P}_n$ by

$$\bar{P}_n = (P_1 + \cdots + P_n)/n.$$

Let ρ and β denote the Prohorov and dual-bounded-Lipschitz metrics on $\mathscr{P}(S)$ respectively: thus for $P, Q \in \mathscr{P}(S)$,

$$\rho(P, Q) = \inf\{\varepsilon > 0\colon P(A) \leq \varepsilon + Q(A^{\varepsilon}) \text{ for all Borel sets } A\}$$

where

$$A^{\varepsilon} \equiv \{y \in S\colon d(x, y) < \varepsilon \text{ for some } x \in A\},$$

* Research supported by the National Science Foundation under MCS 77-02255.

and

$$\beta(P, Q) \equiv \|P-Q\|_{\mathrm{BL}}^{*} \equiv \sup\left\{\left|\int_S f\,\mathrm{d}(P-Q)\right| : \|f\|_{\mathrm{BL}} \le 1\right\}$$

with $\|f\|_\infty \equiv \sup_x |f(x)|$, $\|f\|_{\mathrm{L}} = \sup_{x\neq y}|f(x)-f(y)|/d(x,y)$, and $\|f\|_{\mathrm{BL}} \equiv \|f\|_\infty + \|f\|_{\mathrm{L}}$.

When $P_1 = P_2 = \cdots = P$ (so $\bar{P}_n = P$ for all $n \ge 1$) it is well known that $\rho(\mathbb{P}_n, P) \to 0$ a.s. and $\beta(\mathbb{P}_n, P) \to 0$ a.s.; the latter is due to Fortet and Mourier [5], and the convergence of the Prohorov distance ρ follows from this since ρ and β are equivalent metrics [2, Coroll. 3, p. 1568]. Varadarajan [8] proves that $\mathbb{P}_n$ converges weakly to P with probability one; and ρ and β metrize this convergence [4, Th. 8.3].

In the present case of possibly differing P_n's, the measures $\bar{P}_n$ vary with n, and hence (as suggested by [1, Th. 13]) some restriction is necessary in order to insure convergence. A sufficient condition is that the sequence of measures $\{\bar{P}_n\}_{n\ge1}$ be tight.

Theorem 1. *If $\{\bar{P}_n\}_{n\ge1}$ is tight, then $\rho(\mathbb{P}_n, \bar{P}_n) \to 0$ a.s. and $\beta(\mathbb{P}_n, \bar{P}_n) \to 0$ a.s. as $n \to \infty$.*

Proof. Since ρ and β are equivalent metrics, it suffices to show that $\beta(\mathbb{P}_n, \bar{P}_n) \to 0$ a.s. For f bounded and continuous

$$\int_S f\,\mathrm{d}(\mathbb{P}_n - \bar{P}_n) = \frac{1}{n}\sum_{i=1}^{n} (f(X_i) - \mathbf{E}f(X_i)) \to 0 \text{ a.s.} \tag{1}$$

as $n \to \infty$. This is a consequence of Kolmogorov's strong law of large numbers for independent random variables. Thus if $\mathscr{F}$ is any countable collection of bounded continuous functions,

$$\mathbf{P}\left\{\int_S f\,\mathrm{d}(\mathbb{P}_n - \bar{P}_n) \to 0 \text{ as } n \to \infty \text{ for each } f \in \mathscr{F}\right\} = 1. \tag{2}$$

Now let $\varepsilon > 0$; since $\{\bar{P}_n\}$ is tight there is a compact set $K \subset S$ such that $\bar{P}_n(K) > 1-\varepsilon$ for all $n \ge 1$. Note that the set of functions $B = \{f : \|f\|_{\mathrm{BL}} \le 1\}$, restricted to K, is a compact set of functions for $\|\cdot\|_\infty$. Hence for some finite m there are $f_1, \ldots, f_m \in \mathrm{BL}(S, d)$ such that for any $f \in B$, $\sup_{x\in K}|f(x) - f_j(x)| < \varepsilon$ for some j, and further

$$\sup_{x \in K^\varepsilon} |f(x) - f_j(x)| \le 3\varepsilon. \tag{3}$$

Let $g(x) = \max\{0, (1 - \varepsilon^{-1} d(x, K))\}$. Then $g \in \mathrm{BL}(S, d)$ and $1_K \le g \le 1_{K^\varepsilon}$. Thus

$$\bar{P}_n(K^\varepsilon) \ge \int g\,\mathrm{d}\bar{P}_n \ge \bar{P}_n(K) \ge 1 - \varepsilon \tag{4}$$

and

$$\begin{aligned}\mathbb{P}_n(K^\varepsilon) &\ge \int g\,\mathrm{d}\mathbb{P}_n = \int g\,\mathrm{d}(\mathbb{P}_n - \bar{P}_n) + \int g\,\mathrm{d}\bar{P}_n \\ &> -\varepsilon + 1 - \varepsilon = 1 - 2\varepsilon\end{aligned} \tag{5}$$

for n sufficiently large and for all ω in a set with probability 1 by (1). Therefore, using $\|f\|_\infty \leq 1$, (2), (3), (4) and (5), we get

$$\left|\int_S f\,\mathrm{d}(\mathbb{P}_n - \bar{P}_n)\right| = \left|\left(\int_{K^\varepsilon} + \int_{(K^\varepsilon)^c}\right) f\,\mathrm{d}(\mathbb{P}_n - \bar{P}_n)\right|$$

$$= \left|\int_{K^\varepsilon} (f - f_j + f_j)\,\mathrm{d}(\mathbb{P}_n - \bar{P}_n) + \int_{(K^\varepsilon)^c} f\,\mathrm{d}(\mathbb{P}_n - \bar{P}_n)\right|$$

$$\leq 2\cdot 3\varepsilon + \left|\int_{K^\varepsilon} f_j\,\mathrm{d}(\mathbb{P}_n - \bar{P}_n)\right| + \mathbb{P}_n((K^\varepsilon)^c) + \bar{P}_n((K^\varepsilon)^c) \quad \text{(by (3))}$$

$$\leq 9\varepsilon + \left|\int_S f_j\,\mathrm{d}(\mathbb{P}_n - \bar{P}_n)\right| \quad \text{(by (4) and (5))}$$

$$\leq 10\varepsilon \quad \text{for } n \geq N(\varepsilon, \omega) \quad \text{(by (2))}.$$

Letting $\varepsilon \downarrow 0$ (through a countable set) completes the proof.

If $\bar{P}_n$ converges weakly to some $P \in \mathcal{P}(S)$ then $\{\bar{P}_n\}$ is tight ([6], [4, Th. 10.3]). But, of course, $\{\bar{P}_n\}$ may be tight and not weakly convergent. If $\{P_n\}$ is tight, then $\{\bar{P}_n\}$ is tight.

When $S = \mathbb{R}^1$, $\mathbb{F}_n(x) \equiv \mathbb{P}_n(-\infty, x]$, and $\bar{F}_n(x) \equiv \bar{P}_n(-\infty, x]$, Shorack [7, Th. 1, p. 9] has shown that $\|\mathbb{F}_n - \bar{F}_n\|_\infty \to 0$ as $n \to \infty$ for arbitrary triangular arrays of row-independent rv's. Dudley [3] has examined the rate of convergence to zero of $\mathbf{E}\beta(\mathbb{P}_n, \bar{P}_n)$ and $\mathbf{E}\rho(\mathbb{P}_n, \bar{P}_n)$ in the case $P_n = P$ for all $n \geq 1$.

2. Remarks and examples

In each of the following three examples the sequence of average measures $\{\bar{P}_n\}$ is not tight. In Example 1 $\{\bar{P}_n\}$ is not tight because S is not complete, but yet $\beta(\mathbb{P}_n, \bar{P}_n) \to 0$ a.s. since S is totally bounded, and hence the closure of S, S^-, is compact (so $\{\bar{P}_n\}$ is tight as a sequence of measures on S^-). In Example 2 S is not totally bounded, but the measures P_n are degenerate and hence $\beta(\mathbb{P}_n, \bar{P}_n) = 0$ a.s. for all $n \geq 1$, even though $\{\bar{P}_n\}$ is not tight. Finally, in Example 3 $\{\bar{P}_n\}$ is not tight and $\liminf_{n\to\infty} \beta(\mathbb{P}_n, \bar{P}_n) > 0$ with probability one. Thus although tightness of the sequences $\{\bar{P}_n\}$ is a useful sufficient condition for a.s. convergence of $\beta(\mathbb{P}_n, \bar{P}_n)$ or $\rho(\mathbb{P}_n, \bar{P}_n)$ to zero, it is not necessary. The examples suggest that a necessary and sufficient condition will probably involve some sort of 'degenerateness at infinity' of the sequence $\{P_n\}$.

Example 1. Let $S = (0, 1]$, and suppose that $X_n \sim \text{Uniform}(2^{-n}, 2^{-(n+1)})$ are independent for $n \geq 1$. Then $\{\bar{P}_n\}$ is not tight ($\bar{P}_n \to \delta_0$, with $0 \notin S$), but $S^- = [0, 1]$ is compact and $\mathrm{BL}(S, d)$ is naturally isometric to $\mathrm{BL}(S^-, d)$ where Theorem 1 applies. Hence $\beta(\mathbb{P}_n, \bar{P}_n) \equiv \|\mathbb{P}_n - \bar{P}_n\|^*_{\mathrm{BL}(S,d)} \equiv \|\mathbb{P}_n - \bar{P}_n\|^*_{\mathrm{BL}(S^-,d)} \to 0$ a.s.

Example 2. Let $S=\mathbb{R}^1$ and let $X_n\equiv n$, $n\geq 1$. Then $\{\bar{P}_n\}$, the sequence of uniform measures on $\{1,\ldots,n\}$, is not tight, but $\mathbb{P}_n=\bar{P}_n$ with probability one, and hence $\beta(\mathbb{P}_n,\bar{P}_n)=0$ a.s. for all $n\geq 1$.

Example 3. Let $S=\mathbb{R}^1$ and suppose that $X_n\sim\text{Uniform}(2n-2,\ 2n-1)$ are independent for $n\geq 1$. Then for each $n\geq 1$, $\bar{P}_n$ is the uniform measure on $\bigcup_{i=1}^n(2i-2,2i-1)$, and it is easily seen that $\{\bar{P}_n\}$ is not tight. To show that $\beta(\mathbb{P}_n,\bar{P}_n)$ does not converge to zero we proceed as follows: given $X_1(\omega), X_2(\omega),\ldots,$ define $f(x)=f(x,\omega)$ by $f(X_n(\omega),\omega)=0$ and $f(2n-\frac{1}{2},\omega)=\frac{1}{3}$ for all $n\geq 1$, $f(x)=0$ for $x\leq X_1$, and let f be linear between these points. Then $\|f\|_{\mathrm{L}}\leq\frac{2}{3}$ and $\|f\|_\infty\leq\frac{1}{3}$ so $\|f\|_{\mathrm{BL}}\leq 1$. Note that $2n-1-X_n\equiv U_n$ are i.i.d. Uniform(0, 1) rv's, and that $X_n-(2n-2)\equiv V_n$ are also i.i.d. Uniform(0, 1) rv's. Since $\int f\,\mathrm{d}\mathbb{P}_n=0$, an elementary computation shows that

$$\left|\int_S f\,\mathrm{d}(\mathbb{P}_n-\bar{P}_n)\right|=\frac{1}{6n}\sum_{i=1}^n\left\{\frac{U_i^2}{U_i+\frac{1}{2}}+\frac{V_i^2}{V_i+\frac{1}{2}}\right\}$$

$$\underset{\text{a.s.}}{\to}\tfrac{1}{3}\mathbf{E}\left\{\frac{U^2}{U+\frac{1}{2}}\right\}=\tfrac{1}{12}\log 3>0$$

and hence $\liminf_{n\to\infty}\beta(\mathbb{P}_n,\bar{P}_n)\geq\liminf_{n\to\infty}|\int f\,\mathrm{d}(\mathbb{P}_n-\bar{P}_n)|=\frac{1}{2}\log 3>0$ with probability one.

Acknowledgment

I owe thanks to Joe Glover and Ronald Pyke for helpful discussions.

References

[1] R.M. Dudley, Convergence of Baire measures, Studia Math. 27 (1966) 251–268.
[2] R.M. Dudley, Distances of probability measures and random variables, Ann. Math. Statist. 39 (1968) 1563–1572.
[3] R.M. Dudley, The speed of mean Glivenko–Cantelli convergence, Ann. Math. Statist. 40 (1969), 40–50.
[4] R.M. Dudley, Probabilities and metrics; Convergence of laws on metric spaces with a view to statistical testing, Lecture Notes Series No. 45, Mathematics Institute, Aarhus University, Aarhus, Denmark (1976).
[5] R. Fortet and E. Mourier, Convergence de la répartition empirique vers la répartition théorique, Ann. Sci. École Norm. Sup. 70 (1953) 266–285.
[6] L. LeCam, Convergence in distribution of stochastic processes, Univ. of Calif. Publications in Statistics 2 (11) (1957) 207–236.
[7] G.R. Shorack, The weighted empirical process of row independent random variables with arbitrary d.f.'s, Statistica Neerlandica 33 (1979) 169–189.
[8] V.S. Varadarajan, On the convergence of sample probability distributions, Sankhyā 19 (1958) 23–26.