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
and

$$\beta(P, Q) = \|P - Q\|_\infty = \sup \left\{ \int f \, d(P - Q) : \|f\|_\infty \leq 1 \right\}$$

with \(\|f\|_\infty = \sup_x |f(x)|, \|f\|_1 = \sup_{x,y} |f(x) - f(y)|/d(x, y), \) and \(\|f\|_\infty = \|f\|_\infty + \|f\|_L.$$  

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

In the present case of possibly differing \(P_n\)'s, the measures \(P_n\) 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 \(\{P_n\}_{n=1}^\infty\) be tight.

**Theorem 1.** If \(\{P_n\}_{n=1}^\infty\) is tight, then \(\rho(P_n, P) \to 0\) a.s. and \(\beta(P_n, P) \to 0\) a.s. as \(n \to \infty\).

**Proof.** Since \(\rho\) and \(\beta\) are equivalent metrics, it suffices to show that \(\beta(P_n, P) \to 0\) a.s. For \(f\) bounded and continuous

$$\int_S f \, d(P_n - P) = \frac{1}{n} \sum_{i=1}^n (f(X_i) - Ef(X_i)) \to 0 \text{ a.s.}$$  \hspace{1cm} (1)

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

$$P\left\{ \int_S f \, d(P_n - P) \to 0 \text{ as } n \to \infty \text{ for each } f \in \mathcal{F} \right\} = 1.$$  \hspace{1cm} (2)

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

$$\sup_{x \in K} |f(x) - f_j(x)| \leq 3\varepsilon.$$  \hspace{1cm} (3)

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

$$P_n(K^c) \geq \int_S g \, dP_n \Rightarrow P_n(K^c) \geq 1 - \varepsilon$$  \hspace{1cm} (4)

and

$$P_n(K^c) \geq \int_S g \, dP_n = \int_S g \, d(P_n - P) + \int_S g \, dP_n \geq -\varepsilon + 1 - \varepsilon = 1 - 2\varepsilon$$  \hspace{1cm} (5)
for $n$ sufficiently large and for all $\omega$ in a set with probability 1 by (1). Therefore, using $\|f\|_\infty \leq 1$, (2), (3), (4) and (5), we get

\[
\left| \int_S f \, d(P_n - \bar{P}_n) \right| = \left| \left( \int_{K'} + \int_{(K')^c} \right) f \, d(P_n - \bar{P}_n) \right|
\]

\[
= \left| \int_{K'} (f - f_1 + f_1) \, d(P_n - \bar{P}_n) + \int_{(K')^c} f \, d(P_n - \bar{P}_n) \right|
\]

\[
\leq 2 \cdot 3 \varepsilon + \left| \int_{K'} f_1 \, d(P_n - \bar{P}_n) \right| + P_n((K')^c) + \bar{P}_n((K')^c) \quad \text{(by (3))}
\]

\[
\leq 9 \varepsilon + \left| \int_{S} f_1 \, d(P_n - \bar{P}_n) \right| \quad \text{(by (4) and (5))}
\]

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

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

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

When $S = \mathbb{R}^1$, $F_n(x) = \mathbb{P}_n(-\infty, x)$, and $\bar{F}_n(x) = \mathbb{P}_n(-\infty, x)$, Shorack [7, Th. 1, p. 9] has shown that $\|\mathbb{P}_n - \bar{P}_n\|_\infty \to 0$ as $n \to \infty$ for arbitrary triangular arrays of row-independent r.v.'s. Dudley [3] has examined the rate of convergence to zero of $\mathbb{E} \beta(\mathbb{P}_n, \bar{P}_n)$ and $\mathbb{E} \rho(\mathbb{P}_n, \bar{P}_n)$ 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 $S$, $S^\prime$, is compact (so $\{\bar{P}_n\}$ is tight as a sequence of measures on $S^\prime$). 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 $\{P_n\}$ 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 ($\bar{P}_n \to \delta_1$, with $0 \not\in S$), but $S^\prime = [0, 1]$ is compact and $\text{BL}(S, d)$ is naturally isometric to $\text{BL}(S^\prime, d)$ where Theorem 1 applies. Hence $\beta(\mathbb{P}_n, \bar{P}_n) = \|P_n - \bar{P}_n\|_{\text{BL}(S, d)} = \|P_n - \bar{P}_n\|_{\text{BL}(S^\prime, d)} \to 0$ a.s.
Example 2. Let $S = \mathbb{R}^1$ and let $X_n = n$, $n \geq 1$. Then $\{\tilde{P}_n\}$, the sequence of uniform measures on $\{1, \ldots, n\}$, is not tight, but $\mathbb{P}_n = \tilde{P}_n$ with probability one, and hence $\beta(\mathbb{P}_n, \tilde{P}_n) = 0$ 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 $n \geq 1$, $\tilde{P}_n$ is the uniform measure on $\bigcup_{i=1}^{n-1} (2i-2, 2i-1)$, and it is easily seen that $\{\tilde{P}_n\}$ is not tight. To show that $\beta(\mathbb{P}_n, \tilde{P}_n)$ 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 $f(2n-\frac{1}{2}, \omega) = \frac{1}{2}$ for all $n \geq 1$, $f(x) = 0$ for $x \in X_1$, and let $f$ be linear between these points. Then $\|f\|_L \leq \frac{1}{2}$ and $\|f\|_\infty \leq \frac{1}{2}$ so $\|f\|_L \leq 1$. Note that $2n-1 - X_n = U_n$ are i.i.d. Uniform(0, 1) rv's, and that $X_n - (2n-2) = V_n$ are also i.i.d. Uniform(0, 1) rv's. Since $\int f \, d\mathbb{P}_n = 0$, an elementary computation shows that

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

$$\to \frac{1}{2} \mathbb{E} \left[ \frac{U^2}{U + \frac{1}{2}} \right] = \frac{1}{2} \log 3 > 0$$

and hence $\lim_{n \to \infty} \beta(\mathbb{P}_n, \tilde{P}_n) > 0$. With $\mathbb{P}_n - \tilde{P}_n$ with probability one.

Acknowledgment

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

References

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
Prohorov metric
Bounded-dual-Lipschitz metric
convergence

1. The theorem

Let \((S, d)\) be a separable metric space; let \(\mathcal{P}(S)\) 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 \(P_n \in \mathcal{P}(S)\). For \(x \in S\) let \(\delta_x\) be the unit mass at \(x\). For \(n \geq 1\) define the `empirical measure' \(\bar{P}_n\) by

\[
\bar{P}_n = \frac{(\delta_{X_1} + \cdots + \delta_{X_n})}{n}
\]

and the `average measure' \(\overline{P}_n\) by

\[
\overline{P}_n = \frac{(P_1 + \cdots + P_n)}{n}.
\]

Let \(\rho\) and \(\beta\) denote the Prohorov and dual-bounded-Lipschitz metrics on \(\mathcal{P}(S)\) respectively: thus for \(P, Q \in \mathcal{P}(S)\),

\[
\rho(P, Q) = \inf\{\varepsilon > 0 : P(A) \leq Q(A') + \varepsilon \} \quad \text{and} \quad \beta(P, Q) = \inf\{\varepsilon > 0 : P(A) \leq Q(A') + \varepsilon \}
\]

where

\[
A' = \{y \in S : d(x, y) < \varepsilon \text{ for some } x \in A\}.
\]

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

0304–4149/81/0000–0000/$02.50 © North-Holland Publishing Company