PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 1, Fasc. 2 (1980), p. 109-115

EMPIRICAL PROCESSES, VAPNIK-CHERVONENKIS CLASSES AND POISSON PROCESSES

BY

MARK DURST (Livermore, California)
AND RICHARD M. DUDLEY (Cambridge, Massachusetts)

Abstract. For background of this paper* see [2]. Given a probability space $(X, \mathscr{A}, P)$, let G_P be the Gaussian process with mean 0, indexed by $\mathscr{A}$, and such that

$$EG_P(A)G_P(B) = P(A\cap B)-P(A)P(B), \quad A, B \in \mathscr{A}.$$

(1) Let $\mathscr{C} \subset \mathscr{A}$ and suppose that, for all probability measures (laws) Q on $\mathscr{A}$, G_Q has a version with bounded sample functions on $\mathscr{C}$. (For example, suppose $\mathscr{C}$ is a "universal Donsker class".) Then, for some n, no set F of n elements has all its subsets of the form $C\cap F$, $C\in\mathscr{C}$, i.e. $\mathscr{C}$ is a Vapnik-Chervonenkis class. An example shows that limit theorems for empirical measures need not hold uniformly over a Vapnik-Chervonenkis class of measurable sets, unless further measurability is assumed.

(2) For a law P on $X = \{1, 2, \ldots\}$, the collection 2^X of all subsets is a Donsker class if and only if

$$\sum_m P(m)^{1/2} < \infty.$$

(3) For any probability space $(X, \mathscr{A}, P)$, suppose $\mathscr{C}$ is a P-Donsker class, $\mathscr{C} \subset \mathscr{A}$. Let T_a be a Poisson point process with intensity measure aP, $a > 0$. Then, as $a \to \infty$, $(T_a - aP)/a^{1/2}$ converges in law, with respect to uniform convergence on $\mathscr{C}$, to the Gaussian process W_P with mean 0 and $EW_P(A)W_P(B) = P(A\cap B)$, $A, B\in\mathscr{C}$.

1. Introduction. Let $(X, \mathscr{A}, P)$ be any probability space. Let G_P and W_P be the Gaussian processes, indexed by $\mathscr{A}$, with mean 0 and such that for all $A, B \in \mathscr{A}$

$$EW_P(A)W_P(B) = P(A\cap B) \quad \text{and} \quad EG_P(A)G_P(B) = P(A\cap B)-P(A)P(B).$$

* This research was partially supported by National Science Foundation Grant MCS-7904474.

Then for all $A \in \mathcal{A}$ we can write

$$W_P(A) = G_P(A) + P(A)H,$$

where $H := W_P(X)$ is a standard Gaussian variable independent of G_P.

Let $X_1, X_2, \ldots$ be independent and identically distributed with law P, and let P_n be the random empirical measure $n^{-1}(\delta_{X_1} + \ldots + \delta_{X_n})$. Let $\mathcal{C} \subset \mathcal{A}$. In [2], $\mathcal{C}$ was called a *P-Donsker class* if the convergence of laws $\mathcal{L}(n^{1/2}(P_n - P)) \to \mathcal{L}(G_P)$ holds with respect to uniform convergence on $\mathcal{C}$ in a suitable sense, together with some measurability conditions. Here we will need only the following Skorohod-Wichura form of convergence (see [2], p. 900-902):

1.1. If $\mathcal{C}$ is a P-Donsker class, then there is a probability space $(\Omega, \mathcal{B}, \Pr)$ and for $n = 1, 2, \ldots$ there are processes $(\omega, C) \to A_n(\omega, C)$, $\omega \in \Omega$, $C \in \mathcal{C}$, such that, for each fixed n, the laws of the processes $n^{1/2}(P_n - P)$ and A_n are the same and such that

$$\lim_{n \to \infty} \sup_{C \in \mathcal{C}} |A_n(\omega, C) - G_P(C)(\omega)| = 0 \text{ a.s.},$$

where G_P is defined on the probability space Ω. It follows that

$$\sup_{C \in \mathcal{C}} |G_P(C)(\omega)| < \infty \text{ a.s.}$$

Sections 2, 3 and 4 use the above, but are independent of one another.

2. Universal Donsker classes are Vapnik-Chervonenkis classes. For any set X let 2^X be the collection of all its subsets (power set). Let $\mathcal{C} \subset 2^X$. Then $\mathcal{C}$ is said to *shatter* a set $F \subset X$ if $2^F = \{F \cap C\colon C \in \mathcal{C}\}$. Also, $\mathcal{C}$ is called a *Vapnik-Chervonenkis class* if, for some finite n, no set F with n elements is shattered by $\mathcal{C}$.

2.1. THEOREM. *For any set X and collection $\mathcal{C}$ of subsets of X which is not a Vapnik-Chervonenkis class, there are a purely atomic probability measure P on X and a countable collection $\mathcal{D} \subset \mathcal{C}$ such that G_P is almost surely unbounded on $\mathcal{D}$.*

Proof. Since $\mathcal{C}$ shatters sets of all sizes, for each $n = 1, 2, \ldots$ there is a set F_n with 4^n elements, shattered by $\mathcal{C}$. Let

$$G_n := F_n \setminus \bigcup_{j<n} F_j.$$

Then the G_n are disjoint and have cardinality

$$\operatorname{card}(G_n) \geqslant 4^n - \sum_{j=1}^{n-1} 4^j = 4^n - (4^n - 4)/3 > 2^n,$$

with G_n shattered by $\mathcal{C}$. Take $E_n \subset G_n$ with $\operatorname{card}(E_n) = 2^n$. Then E_n remain disjoint and are shattered by $\mathcal{C}$.

Let $P(\{x\}) = 6/(\pi^2 n^2 \cdot 2^n)$ for each $x \in E_n$, and let $P = 0$ outside $\bigcup_{n=1}^{\infty} E_n$. Then P is a purely atomic probability measure on X.

Let $\mathscr{D}$ be a countable subset of $\mathscr{C}$ which shatters each of the E_n. Let us fix n. Then, for each $C \in \mathscr{D}$,

$$W_P(C) = W_P(C \cap E_n) + W_P(C \backslash E_n).$$

Thus for any K, $0 < K < \infty$, we have

$$\{\omega\colon |W_P(C)(\omega)| < K \text{ for all } C \in \mathscr{D}\} \subset \mathscr{E}_1 \cup \mathscr{E}_2,$$

where

$\mathscr{E}_1 := \{\omega\colon |W_P(B)(\omega)| \leqslant 2K \text{ for all } B \subset E_n\}$,

$\mathscr{E}_2 := \{\omega\colon$ for some $B \subset E_n$, $|W_P(B)(\omega)| > 2K$, and for all such B and all $C \in \mathscr{D}$ with $C \cap E_n = B$ we have $|W_P(C \backslash E_n)(\omega)| > K\}$.

Let

$$S_n := \sum_{x \in E_n} |W_P(\{x\})|.$$

Then since $\sup\{|W_P(B)|\colon B \subset E_n\} \geqslant S_n/2$, we have $\mathscr{E}_1 \subseteq \{S_n \leqslant 4K\}$. For each $x \in E_n$, $W_P(\{x\})$ is a normal random variable with mean 0 and variance $\sigma_n^2 := 6/(\pi^2 n^2 \cdot 2^n)$. Thus

$$\mathrm{E}\,|W_P(\{x\})| = (2/\pi)^{1/2} \sigma_n \quad \text{and} \quad \operatorname{var}\big(|W_P(\{x\})|\big) = \sigma_n^2 (1 - 2/\pi).$$

Then

$$\mathrm{E}\, S_n = 2^n (2/\pi)^{1/2} \sigma_n \quad \text{and} \quad \operatorname{var}(S_n) = \big(6/(\pi^2 n^2)\big)(1 - 2/\pi),$$

since W_P has independent values on disjoint sets. Hence, by Chebyshev's inequality, for large n we get

$$\begin{aligned} \Pr\{S_n \leqslant 4K\} &\leqslant \Pr\{|S_n - \mathrm{E}\, S_n| \geqslant \mathrm{E}\, S_n - 4K\} \leqslant n^{-2}/(4K - \mathrm{E}\, S_n)^2 \\ &\leqslant 1/\big(4Kn - 2^{n/2}(12/\pi^3)^{1/2}\big)^2 := f(n, K) \to 0 \quad \text{as } n \to \infty \end{aligned}$$

for any fixed K.

Now we consider the event $\mathscr{E}_2$. Let $tt(n) := 2^{2^n}$. Enumerate 2^{E_n} by $B(1), \ldots, B(tt(n))$, and let

$$M_1 := \{\omega\colon \big|W_P(B(1))(\omega)\big| > 2K\},$$

$$M_m := \{\omega \notin \bigcup_{j=1}^{m-1} M_j\colon \big|W_P(B(m))(\omega)\big| > 2K\}, \quad m \geqslant 2,$$

$$D_j := \{\omega \in M_j\colon \text{for all } C \in \mathscr{D} \text{ such that } C \cap E_n = B(j),\ |W_P(C \backslash E_n)(\omega)| > K\}.$$

By the independence of W_P on disjoint sets, we have

$$\begin{aligned} \Pr(D_j) = \Pr(M_j) \Pr\{&\text{for all } C \in \mathscr{D} \text{ such that } C \cap E_n = B(j), \\ &|W_P(C \backslash E_n)(\omega)| > K\} \leqslant \Pr(M_j) \cdot 2\Phi(-K), \end{aligned}$$

where Φ is the standard normal distribution function, since, for any fixed set A, $W_P(A)$ is normal with mean 0 and variance less than 1. Now,

$$\mathscr{E}_2 \subset \bigcup_{1 \leqslant j \leqslant tt(n)} D_j,$$

so that

$$\Pr(\mathscr{E}_2) \leqslant \sum_{1 \leqslant j \leqslant tt(n)} \Pr(M_j) \cdot 2\Phi(-K)$$
$$= 2\Phi(-K) \Pr\big(|W_P(B)| > 2K \text{ for some } B \subset E_n\big) \leqslant 2\Phi(-K).$$

It follows that

$$\Pr\big(|W_P(C)| < K \text{ for all } C \in \mathscr{D}\big) \leqslant f(n, K) + 2\Phi(-K).$$

Making K large enough, and then n large enough, completes the proof.

It follows that if $\mathscr{C}$ is a *universal Donsker class*, i.e. it is a P-Donsker class for all P on the σ-algebra $\mathscr{A} \supset \mathscr{C}$, then $\mathscr{C}$ is a Vapnik-Chervonenkis class. In [2], Section 7 and Correction, it is shown that every Vapnik-Chervonenkis class satisfying some measurability conditions is a universal Donsker class. The remaining problem is to find what measurability conditions are needed. The following example shows that some further measurability is necessary.

2.2. Proposition. *There exist a set X and a class $\mathscr{C}$ of countable subsets of X, which shatters no 2-element set, and a probability measure P such that almost surely*

$$\sup_{A \in \mathscr{C}} (P_n - P)(A) = 1 \quad \textit{for all } n.$$

Assuming the continuum hypothesis, we can take $X = [0, 1]$ and P to be Lebesgue measure.

Proof. Let $(X, \prec)$ be an uncountable well-ordered set such that all its initial segments $\{x\colon x \prec y\}$, $y \in X$, are countable. Let $\mathscr{C}$ be the collection of all these initial segments. Then $\mathscr{C}$ does not shatter any set with two elements. Let P be any probability measure on X which is 0 on countable sets and 1 on their complements. Given any finite set $\{X_1, \ldots, X_n\} \subset X$, there is a set A in $\mathscr{C}$ containing all the X_i, so $(P_n - P)(A) = 1$, which completes the proof.

Steele [3] assumes that all sets in $\mathscr{C}$ are measurable and that $\sup_{A \in \mathscr{C}} |(P_n - P)(A)|$ is measurable. These conditions are both satisfied in the example above. Thus it appears that further measurability conditions need to be added to some of the statements and proofs in [3].

3. When is 2^X P-Donsker for X countable? Let X be a countable set, say $X = \{1, 2, \ldots\}$, and let P be a law on X with $P\{m\} := p_m$, $m = 1, 2, \ldots$

3.1. THEOREM. *The collection 2^X of all subsets of X is a P-Donsker class if and only if*

$$(*)\qquad \sum_m p_m^{1/2} < \infty.$$

Proof. Suppose (*) holds. We have $E(\nu_n\{m\})^2 = p_m - p_m^2$ for all n and m, where $\nu_n := n^{1/2}(P_n - P)$. Thus $E|\nu_n\{m\}| \leqslant p_m^{1/2}$, and

$$\sup_n E \sum_{j \geqslant m} |\nu_n\{j\}| \to 0 \quad \text{as } m \to \infty.$$

So, for any $\varepsilon > 0$,

$$\sup_n \Pr\{\sum_{j \geqslant m} |\nu_n\{j\}| > \varepsilon\} \to 0 \quad \text{as } m \to \infty.$$

Thus condition (b) in Theorem 1.2 of [2] holds; as the other conditions also hold for $\mathscr{C} = 2^X$, X countable, it is a P-Donsker class.

On the other hand, if $\sum_m p_m^{1/2} = \infty$, then

$$\sum_m |G_P\{m\}| = \infty \text{ a.s.}$$

by Proposition 6.6 of [1], letting $b_m := p_m^{1/2}$, $\varphi_m = 1_{\{m\}}/p_m^{1/2}$, and recalling the relations $L(1_A) = W_P(A) = G_P(A) + P(A)W_P(X)$. Thus G_P has sample functions almost surely unbounded on 2^X (it is enough to consider the countable collection of finite sets). Consequently, 2^X is not a Donsker class, which completes the proof.

4. A limit theorem for Poisson processes. Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space. Then the *Poisson process* T_μ with *intensity measure* μ is indexed by the measurable sets A with $\mu(A) < \infty$; $T_\mu(A)$ is a Poisson variable with parameter $\mu(A)$, and T_μ has independent values on disjoint sets, being additive for (finitely many) disjoint sets. These conditions, as is known, consistently define a stochastic process.

For $0 < \mu(X) < \infty$, let $T(X) = n$ be a Poisson variable with parameter $\mu(X)$. Then let

$$T = \sum_{1 \leqslant i \leqslant n} \delta_{X_i},$$

where the X_i are independent and identically distributed with law $\mu/\mu(X)$, and independent of n. It is easily seen that this T is a Poisson process T_μ.

Now let P be a probability measure and $0 < \lambda < \infty$. Then, as $\lambda \to \infty$, $(T_{\lambda P} - \lambda P)/\lambda^{1/2}$ converges in law to W_P, at least on any finite collection of measurable sets. For $\mathscr{C} \subset \mathscr{A}$, we say that this *convergence in law holds with respect to uniform convergence on* $\mathscr{C}$ if there exists a probability space $(\Omega, \Pr)$ carrying a process W_P and processes S_λ, $0 < \lambda < \infty$, such that for each λ the process S_λ has the same law (as a process on $\mathscr{C}$) as $(T_{\lambda P}-\lambda P)/\lambda^{1/2}$, and such that

$$\lim_{\lambda\to\infty} \sup_{C\in\mathscr{C}} |(S_\lambda - W_P)(C)| = 0 \text{ a.s.}$$

The following result was proposed by E. B. Dynkin in a discussion in Oberwolfach, March 1979.

4.1. THEOREM. *For any probability measure P and P-Donsker class $\mathscr{C}$, $(T_{\lambda P}-\lambda P)/\lambda^{1/2}$ converges in law to W_P with respect to uniform convergence on $\mathscr{C}$.*

Proof. Take $X_1, X_2, \ldots$, independent with distribution P. For each λ, $0<\lambda<\infty$, let $n = n(\omega,\lambda)$ be a Poisson variable with parameter λ, independent of the X_i. Then we can write $T_{\lambda P} = n(\omega,\lambda) P_{n(\omega,\lambda)}$ (in law).

Now $(n(\omega,\lambda)-\lambda)/\lambda^{1/2}$ converges in law to a standard Gaussian variable as $\lambda\to\infty$. To replace this convergence by almost sure convergence of real random variables, we use the following standard procedure. For any probability distribution function F on $\boldsymbol{R}$ and for $0<y<1$, let

$$F^{-1}(y) := \inf\{x\colon F(x) \geqslant y\}.$$

Suppose laws μ_m on $\boldsymbol{R}$ with distribution functions F_m converge to a law μ_0. Then $F_m^{-1}(y)\to F_0^{-1}(y)$ whenever the interval $F_0^{-1}\{y\}$ contains at most one point. Thus $F_m^{-1}(y)\to F_0^{-1}(y)$ for all y, $0<y<1$, if F_0 has an everywhere positive density, e.g. if it is a non-degenerate normal distribution function. Thus if $\mu_\lambda\to\mu_0$ as $\lambda\to\infty$, where μ_λ has distribution function F_λ and μ_0 has an everywhere strictly positive density, then, for $0<y<1$, $F_\lambda^{-1}(y)\to F_0^{-1}(y)$ as $\lambda\to\infty$ (continuously).

Now, taking a new probability space if necessary, we may assume that, for all ω,

$$\lim_{\lambda\to\infty} (n(\omega,\lambda)-\lambda)/\lambda^{1/2} = H,$$

where H is a standard normal variable. Also, by 1.1, we can take $n^{1/2}(P_n-P) := \nu_n \to G_P$ uniformly on $\mathscr{C}$ almost surely as $n\to\infty$, where the $n(\omega,\lambda)$ and H are independent of P_n and G_P.

Now $(n(\omega,\lambda)-\lambda)/\lambda\to 0$ a.s., so $n(\omega,\lambda)/\lambda\to 1$ a.s. and $n(\omega,\lambda)\to\infty$ a.s. Thus $\nu_{n(\omega,\lambda)}\to G_P$ uniformly on $\mathscr{C}$ almost surely. So

$$\begin{aligned}(n(\omega,\lambda)&P_{n(\omega,\lambda)}-\lambda P)/\lambda^{1/2}\\ &= (n(\omega,\lambda)/\lambda)^{1/2}\nu_{n(\omega,\lambda)} + (n(\omega,\lambda)-\lambda)P/\lambda^{1/2} \to G_P+HP = W_P\end{aligned}$$

uniformly on $\mathscr{C}$ almost surely as $\lambda\to\infty$, which completes the proof.

Added in proof. Theorem 4.1 extends a result known in the classical one-dimensional case: cf. P. Gaenssler and W. Stute, Ann. Probability 7 (1979), p. 193-243, Theorem 2.6.2 on p. 230-231.

REFERENCES

[1] R. M. Dudley, *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes*, J. Functional Analysis 1 (1967), p. 290-330.

[2] – *Central limit theorems for empirical measures*, Ann. Probability 6 (1978), p. 899-929; *Correction*, ibidem 7 (1979), p. 909-911.

[3] J. M. Steele, *Empirical discrepancies and subadditive processes*, ibidem 6 (1978), p. 118-127.

[4] V. N. Vapnik and A. Ya. Chervonenkis, *On the uniform convergence of the relative frequencies of events to their probabilities*, Teor. Verojatnost. i Primenen. 16 (1971), p. 264-279 (in Russian); Theor. Probability Appl. 16 (1971), p. 264-280 (English translation).

Room 2-245, M.I.T.
Cambridge, MA 02139, U.S.A.

Mathematics-Statistics Section
Lawrence Livermore Laboratory
Livermore, CA 94550, U.S.A.

Received on 14. 11. 1979