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 \((X, \mathcal{A}, P)\), let \(G_P\) be the Gaussian process with mean 0, indexed by \(\mathcal{A}\), and such that

\[
E(G_P(A)G_P(B)) = P(A \cap B) - P(A)P(B), \quad A, B \in \mathcal{A}.
\]

(1) Let \(\mathcal{C} \subseteq \mathcal{A}\) and suppose that, for all probability measures \(Q\) on \(\mathcal{A}\), \(G_Q\) has a version with bounded sample functions on \(\mathcal{C}\). (For example, suppose \(\mathcal{C}\) is a "universal Donsker class".) Then, for some \(n\), no set \(F\) of \(n\) elements has all its subsets of the form \(C \cap F, C \in \mathcal{C}\), i.e. \(\mathcal{C}\) 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 \((X, \mathcal{A}, P)\), suppose \(\mathcal{C}\) is a \(P\)-Donsker class, \(\mathcal{C} \subseteq \mathcal{A}\). 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 \(\mathcal{C}\), to the Gaussian process \(W_P\) with mean 0 and \(EW_P(A)W_P(B) = P(A \cap B), A, B \in \mathcal{A}\).

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

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

* 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 $\mathcal{E} \subset \mathcal{A}$. In [2], $\mathcal{E}$ 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 $\mathcal{E}$ 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 $\mathcal{E}$ is a $P$-Donsker class, then there is a probability space $(\Omega, \mathcal{B}, \Pr)$ and for $n = 1, 2, \ldots$ there are processes $(\omega, C) \to A_n(\omega, C)$, $\omega \in \Omega$, $C \in \mathcal{E}$, 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

$$\limsup_{n \to \infty} |A_n(\omega, C) - G_p(C)(\omega)| = 0 \text{ a.s.},$$

where $G_p$ is defined on the probability space $\Omega$. It follows that

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

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 $2^F = \{ F \cap C : C \in \mathcal{C} \}$. 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

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

with $G_n$ shattered by $\mathcal{C}$. Take $E_n \subset G_n$ with $\text{card}(E_n) = 2^n$. 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 \( \mathcal{D} \) be a countable atomic subset of \( \mathcal{C} \) which shatters each of the \( E_n \).

Let us fix \( n \). Then, for each \( C \in \mathcal{D} \),

\[
W_p(C) = W_p(C \cap E_n) + W_p(C \setminus E_n).
\]

Thus for any \( K, 0 < K < \infty \), we have

\[
\{ \omega : | W_p(C)(\omega) | < K \text{ for all } C \in \mathcal{D} \} \subseteq \mathcal{E}_1 \cup \mathcal{E}_2,
\]

where

\[
\mathcal{E}_1 := \{ \omega : | W_p(B) (\omega) | \leq 2K \text{ for all } B \subseteq E_n \},
\]

\[
\mathcal{E}_2 := \{ \omega : \text{for some } B \subseteq E_n, | W_p(B)(\omega) | > 2K, \text{ and for all such } B \text{ and all } C \in \mathcal{D} \text{ with } C \cap E_n = B \text{ we have } | W_p(C \setminus E_n)(\omega) | > K \}.
\]

Let

\[
S_n := \sum_{x \in E_n} | W_p(\{ x \}) |.
\]

Then since sup \( \{ | W_p(B) | : B \subseteq E_n \} \geq S_n/2 \), we have \( \mathcal{E}_1 \subseteq \{ S_n \leq 4K \} \). For each \( x \in E_n \), \( W_p(\{ x \}) \) is a normal random variable with mean 0 and variance

\[
\sigma^2 := 6/(\pi^2 n^2 \cdot 2^n). \quad \text{Thus}
\]

\[
E|W_p(\{ x \})| = (2/\pi)^{1/2} \sigma_n \quad \text{and} \quad \text{var} (|W_p(\{ x \})|) = \sigma_n^2 (1 - 2/\pi).
\]

Then

\[
ES_n = 2^n (2/\pi)^{1/2} \sigma_n \quad \text{and} \quad \text{var} (S_n) = (6/(\pi^2 n^2)) (1 - 2/\pi),
\]

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

\[
\Pr \{ S_n \leq 4K \} \leq \Pr \{ |S_n - ES_n| \geq ES_n - 4K \} \leq n^{-2}/(4K - ES_n)^2 \leq 1/(4Kn - 2^{n/2} (12/\pi^3)^{1/2})^2 := f(n, K) \to 0 \quad \text{as } n \to \infty
\]

for any fixed \( K \).

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

\[
M_1 := \{ \omega : | W_p(B(1))(\omega) | > 2K \},
\]

\[
M_m := \{ \omega \notin \bigcup_{j=1}^{m-1} M_j : | W_p(B(m))(\omega) | > 2K \}, \quad m \geq 2,
\]

\[
D_j := \{ \omega \in M_j : \text{for all } C \in \mathcal{D} \text{ such that } C \cap E_n = B(j), | W_p(C \setminus E_n)(\omega) | > K \}.
\]

By the independence of \( W_p \) on disjoint sets, we have

\[
\Pr (D_j) = \Pr (M_j) \Pr \{ \text{for all } C \in \mathcal{D} \text{ such that } C \cap E_n = B(j), | W_p(C \setminus E_n)(\omega) | > K \} \leq \Pr (M_j) \cdot 2\Phi(-K),
\]
where \( \Phi \) 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,

\[
\mathcal{E}_2 \subset \bigcup_{1 \leq j \leq t(n)} D_j,
\]

so that

\[
\Pr(\mathcal{E}_2) \leq \sum_{1 \leq j \leq t(n)} \Pr(M_j) \cdot 2\Phi(-K) = 2\Phi(-K) \Pr(\|W_P(B)\| > 2K \text{ for some } B \in E_n) \leq 2\Phi(-K).
\]

It follows that

\[
\Pr(|W_P(C)| < K \text{ for all } C \in \mathcal{D}) \leq f(n, K) + 2\Phi(-K).
\]

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

It follows that if \( \mathcal{C} \) is a universal Donsker class, i.e. it is a \( P \)-Donsker class for all \( P \) on the \( \sigma \)-algebra \( \mathcal{A} \rightarrow \mathcal{C} \), then \( \mathcal{C} \) 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 \( \mathcal{C} \) of countable subsets of \( X \), which shatters no 2-element set, and a probability measure \( P \) such that almost surely

\[
\sup_{A \in \mathcal{C}} (P_n - P)(A) = 1 \quad \text{for all } n.
\]

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

Proof. Let \((X, <)\) be an uncountable well-ordered set such that all its initial segments \( \{x: x < y, y \in X\} \), are countable. Let \( \mathcal{C} \) be the collection of all these initial segments. Then \( \mathcal{C} \) 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 \( \mathcal{C} \) containing all the \( X_i \), so \( (P_n - P)(A) = 1 \), which completes the proof.

Steele [3] assumes that all sets in \( \mathcal{C} \) are measurable and that \( \sup_{A \in \mathcal{C}} |(P_n - P)(A)| \) 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
\[ \sum_m p_m^{1/2} < \infty. \]

**Proof.** Suppose (*) holds. We have $E(v_n \{m\})^2 = p_n - p_m^2$ for all $n$ and $m$, where $v_n := n^{1/2} (P_n - P)$. Thus $E|v_n \{m\}| \leq p_m^{1/2}$, and
\[ \sup_n E \sum_{j \geq m} |v_n \{j\}| \to 0 \quad \text{as} \quad m \to \infty. \]

So, for any $\varepsilon > 0$,
\[ \sup_n \Pr \left\{ \sum_{j \geq m} |v_n \{j\}| > \varepsilon \right\} \to 0 \quad \text{as} \quad m \to \infty. \]

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

On the other hand, if $\sum_m p_m^{1/2} = \infty$, then
\[ \sum_m |G_p \{m\}| = \infty \quad \text{a.s.} \]

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 $L(I_A) = W_p (A) = G_p (A) + P(A) W_p (X)$. 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 $(X, \mathfrak{A}, \mu)$ be a $\sigma$-finite measure space. Then the Poisson process $T_\mu$ with intensity measure $\mu$ is indexed by the measurable sets $A$ with $\mu(A) < \infty$; $T_\mu(A)$ is a Poisson variable with parameter $\mu(A)$, and $T_\mu$ 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
\[ T = \sum_{1 \leq i \leq n} \delta_{X_i}, \]
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 $T_\mu$.

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 $\mathcal{C} \subset \mathfrak{A}$, we say that this convergence in law holds with respect to uniform convergence on $\mathcal{C}$ if there exists a probability space $(\Omega, \Pr)$ carrying a process $W_p$ and processes $S_\lambda$, $0 < \lambda < \infty$, such that for
each $\lambda$ the process $S_\lambda$ has the same law (as a process on $\mathcal{C}$) as $(T_{\lambda P} - \lambda P)/\lambda^{1/2}$, and such that
\[
\limsup_{k \to \infty} |(S_{\lambda k} - W_P)(C)| = 0 \text{ a.s.}
\]

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 $\mathcal{C}$, 
$(T_{\lambda P} - \lambda P)/\lambda^{1/2}$ converges in law to $W_P$ with respect to uniform convergence on $\mathcal{C}$.

Proof. Take $X_1, X_2, \ldots$, independent with distribution $P$. For each $\lambda$, $0 < \lambda < \infty$, let $n = n(\omega, \lambda)$ be a Poisson variable with parameter $\lambda$, 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 $\mathbb{R}$ and for $0 < y < 1$, let
\[
F^{-1}(y) := \inf \{ x : F(x) \geq y \}.
\]

Suppose laws $\mu_n$ on $\mathbb{R}$ with distribution functions $F_m$ converge to a law $\mu_0$. 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 $\mu_\lambda$ has distribution function $F_\lambda$ and $\mu_0$ 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 $\omega$,
\[
\lim_{\lambda \to \infty} (n(\omega, \lambda) - \lambda)/\lambda^{1/2} = H,
\]
where $H$ is a standard normal variable. Also, by 1.1, we can take $n^{1/2}(P_n - P) := \psi_n \to G_P$ uniformly on $\mathcal{C}$ 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 $\psi_n(\omega, \lambda) \to G_P$ uniformly on $\mathcal{C}$ almost surely. So
\[
(n(\omega, \lambda) P_{n(\omega, \lambda)} - \lambda P)/\lambda^{1/2} = (n(\omega, \lambda)/\lambda)^{1/2} \psi_n(\omega, \lambda) + (n(\omega, \lambda) - \lambda) P/\lambda^{1/2} \to G_P + HP = W_P
\]
uniformly on $\mathcal{C}$ almost surely as $\lambda \to \infty$, which completes the proof.
Empirical processes


REFERENCES


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