Statistica Sinica 2(1992), 1-15

UNIFORM CONVERGENCE OF PROBABILITY MEASURES ON CLASSES OF FUNCTIONS

P. J. Bickel[1] and P. W. Millar[2]

University of California, Berkeley

Abstract: Let P_n, P be probabilities, and $\mathbf{F}$, $\mathbf{F}^*$ be collections of real functions. Simple conditions are derived under which the simple convergence of $\int f(x)P_n(dx)$ to $\int f(x)P(dx)$ for every f in $\mathbf{F}^*$ implies uniform convergence over $\mathbf{F}$: $\sup_{f\in\mathbf{F}} |\int f(x)P_n(dx) - \int f(x)P(dx)|$ converges to 0. Several examples are discussed, some historical and some new.

Key words and phrases: Weak convergence of probabilities, uniform convergence of probabilities, Pólya class, Pólya's theorem, Glivenko-Cantelli theorem, dual Lipschitz norm, bracketing, Vapnik-Cervonenkis class, convex sets, uniformity class, delta-tight.

0. Foreword

Let P be a probability on a metric space X endowed with an appropriate sigma field. Let $\mathbf{F}$ be a fixed collection of real functions on X such that $|f(x)| \leq c_0 < \infty$ for all f in $\mathbf{F}$. Let $\mathbf{F}^*$ be an auxiliary collection of functions. Both $\mathbf{F}$ and $\mathbf{F}^*$ may depend on P. We consider the question: under what conditions on $\mathbf{F}$ and P does

$$\int f(x)P_n(dx) \to \int f(x)P(dx) \text{ for every } f \in \mathbf{F}^* \tag{0.1}$$

imply

$$\sup\{|\int f(x)P_n(dx) - \int f(x)P(dx)| : f \in \mathbf{F}\} \to 0 \tag{0.2}$$

for every sequence of probability measures $\{P_n\}$ satisfying (0.1)? For reasons given in Section 1, we call such a class $\mathbf{F}$ a Pólya class for P.

(1) Research partially supported by ONR N00014-89-J-1563

(2) Research partially supported by NSA MDA904-88-C-3068

Uniformity results of this sort have a long history; they are also a useful part of the tool kit of every asymptotic statistician. Some specific applications are discussed in Section 1.

In this paper we give two simple conditions for $\mathbf{F}$ to be a Pólya class for P (see, Section 2). Illustrations of the use of these criteria appear in Section 4, where we give simple proofs of uniform convergence in a number of examples, including a recent theorem of Beran, LeCam and Millar. (In some of these examples, X is not separable, but the P_n are δ-tight, so that the support of limiting measures is separable.)

1. Background

The utility of these uniformity results resides in the possibility of deducing from a simple pointwise convergence of measures, (plus possibly a smoothness condition) a far stronger form of convergence. We first introduce a standard notation before proceeding to the illustrations.

If P is a probability on X, define Pf to be the expected value of f under P:

$$Pf \equiv \int f(x)P(dx) \equiv P(f). \tag{1.1}$$

If X is the real line, define

$$\begin{aligned} P(t) &= P\{(-\infty, t]\} \\ P(t-) &= P\{(-\infty, t)\}. \end{aligned} \tag{1.2}$$

Using this notation we see that $\mathbf{F}$ is a Pólya class for P if

$$P_n f \to Pf \quad \forall f \in \mathbf{F}^* \tag{1.3}$$

implies

$$\sup\{|P_n f - Pf| : f \in \mathbf{F}\} \to 0. \tag{1.4}$$

This compact notation, which emphasizes the interpretation of P as a linear functional, has been used for decades in functional analysis and in general Markov theory. Recently it has been extensively used in the literature on empirical processes.

A. Historical examples

(A.1) **Pólya's theorem (1920).** Let P_n, P be probabilities on the line. Suppose (i) there is a dense set D on the line such that

$$P_n(t) \to P(t),\ t \in D$$

(ii) $P(t)$ is continuous as a function of t.
Then Pólya showed (Math. Zeitschr. Vol. 8, Satz I, 1920) that

$$\sup_{t \in R^1} |P_n(t) - P(t)| \to 0.$$

This result is the earliest theorem that we know, of our general framework. Here $\mathbf{F}^*$ consists of indicators of intervals $(-\infty, t], t \in D$ while $\mathbf{F}$ consists of indicators of intervals $(-\infty, t], t \in R$. Thus continuity, plus pointwise convergence in a "small" set of points, is sufficient to imply uniform convergence over a "larger" collection of points.

It is from this seminal theorem that we introduce the term "Pólya class".

(A.2) **Dual Lipschitz Norm.** Let P_n, P be probabilities on, say, a separable metric space X with metric d. Let $\mathbf{F}^*$ consist of all bounded, uniformly continuous, real functions on X. Let $\mathbf{F}$ denote the collection of functions f satisfying
(i) $\sup_x |f(x)| \leq c_0$
(ii) $\sup_x \{|f(x) - f(y)|/d(x,y)\} \leq c_0$.
A standard weak convergence result then asserts that

$$\text{if } P_n f \to Pf \quad \forall f \in \mathbf{F}^*, \text{ then } \sup_{f \in \mathbf{F}} |P_n f - Pf| \to 0;$$

i.e., weak convergence implies convergence in the dual Lipschitz norm. If $X = R^k$, then the class $\mathbf{F}^*$ may be replaced by a smaller class whose linear span is dense in the sup norm on $\mathbf{F}^*$, e.g. the exponentials $\exp\{itx\}$, $t \in R^k$. The relevance of this particular form of uniform convergence to the theory of robust statistical inference was emphasized by Huber (1981). See also (B.3) below.

(A.3) **The Billingsley-Topsoe theorem (1967).** Let X be a separable metric space and assume
(i) $P_n f \to Pf$
for every bounded uniformly continuous function f. Building on work of Ranga Rao (1962) and others, they showed that a collection $\mathbf{F}$ is a Pólya class iff

$$\lim_{\epsilon \downarrow 0} \sup_{\mathbf{F}} P[x : w_f(x, \epsilon) > \delta] = 0$$

for all $\delta > 0$, where $w_f(A)$, the oscillation of f on A, is defined by

$$w_f(A) = \sup\{|f(x) - f(y)| : x, y \in A\}$$

and

$$w_f(x, \epsilon) = w_f(B(x : \epsilon))$$

where $B(x : \epsilon)$ is the ϵ ball about x. This result gives an elegant solution to the problem (when X is separable). However, the conditions given are not easy to check. Indeed, the desired uniform convergence is replaced by a secondary uniformity problem which is often harder than the original. Possibly because of these considerations, this fascinating result appears to have been only rarely applied.

B. Statistical examples

Statistical application of the type of result described in (1.3), (1.4) often involves replacing $\{P_n\}$ there by a sequence of *random* measures and even P itself by a random measure. It turns out that in some asymptotic arguments, these random measures can be replaced by non random measures, via judicious use of available almost sure convergence results, such as Wichura (1970). In such cases one may then apply the basic uniformity theorems for ordinary measures.

(B.1) **Glivenko-Cantelli theorem.** Let $X_1, X_2, \ldots$ be independent, identically distributed real random variables having common distribution P. Let $\widehat{P}_n$ be the empirical measure of this sample. The law of large numbers implies at once that, for each t,

$$\widehat{P}_n(t) \to P(t) \text{ a.s.}$$

The Glivenko-Cantelli theorem, of course, asserts that the above simple result can be parlayed into uniform convergence;

$$\sup_t |\widehat{P}_n(t) - P(t)| \to 0 \text{ a.s.}$$

The result is reminiscent of Pólyas, but without the continuity assumption – this being finessed by the particular random sequence $\{\widehat{P}_n\}$.

(B.2) **The BLMV theorem.** This result is discussed in Section 4; here we just describe quickly its statistical meaning. Let $\widehat{P}_n$, $\widehat{Q}_n$ be random measures on an abstract, not necessarily separable, metric space. Under suitable hypotheses, the result concludes that if $\widehat{P}_n f - \widehat{Q}_n f$ converges to zero (in probability) for bounded, continuous (and appropriately measurable) functions f, that then the convergence to zero holds in the dual Lipschitz norm (c.f. (B.2) above). In a number of statistical applications, $\widehat{Q}_n$ is a random measure based on an estimate of a parameter in a particular statistical model, while $\widehat{P}_n$ is the empirical measure constructed from an appropriate bootstrap sample. The result is therefore an abstract underpinning for the common procedure of estimating, by bootstrap methods, the sampling distribution of a particular statistic. Concrete illustration of its use may be found in Beran and Millar (1987), for example.

(B.3) Consistent estimation of parameters in stationary ergodic processes. Suppose $X_1, \ldots, X_n, \ldots$ is a real stationary ergodic process with distribution P. Let $\widehat{P}_{nj}$ be the empirical joint distribution of $\{X_{i+k}, 1 \leq k \leq j\}$, $1 \leq i \leq n$. By the ergodic theorem $\widehat{P}_{nj}(f) \to P_j(f)$ a.s., where P_j is the distribution of $(X_1, \ldots, X_j)$, and this convergence holds for every bounded uniformly continuous function f on R^j. By a familiar argument, one may take the a.s. exceptional set independent of f. Then one can employ (A.2) to deduce a.s. uniform convergence on the set $\mathbf{F}$ given in (A.2). From this one can then easily deduce consistency properties and qualitative robustness for "robust serial correlations" and similar parameters (cf., Papantoni-Kazakos and Gray (1979)).

2. Conditions for F to be a Pólya class for P

To ease the exposition we shall take $\mathbf{F}^* = \mathbf{F}$ in what follows. The general case simply requires that $\mathbf{F}^*$ satisfy the specified conditions. The uniformity conclusion holds for $\mathbf{F}$.

(i) *Bracketing.* As usual, f^L and f^U *bracket* f if $f^L \leq f \leq f^U$.

$$\begin{aligned} &\text{Given } \mathbf{F}^*, \text{ let}, H_B(\epsilon, P) = \inf\{m : \text{There exist } f_i^L \leq f_i^U, 1 \leq i \leq m \\ &\text{all belonging to } \mathbf{F}^* \text{ such that } P(f_i^U - f_i^L) \leq \epsilon \text{ and } \forall f \in \mathbf{F} \\ &\text{there exists } i(f) \text{ such that } f_i^L,\ f_i^U \text{ bracket } f\}. \end{aligned} \tag{2.1}$$

Proposition 2.1. $\mathbf{F}$ *is a Pólya class if*

$$H_B(\epsilon, P) < \infty \text{ for all } \epsilon > 0.$$

Proof. Using bracketing note that

$$\begin{aligned} |P_n(f) - P(f)| \leq\ & |P_n(f^U_{i(f)}) - P_n(f^L_{i(f)})| \\ & + |P_n(f^L_{i(f)}) - P(f^L_{i(f)})| + |P(f^L_{i(f)}) - P(f)| \end{aligned} \tag{2.2}$$

where $f_1, \ldots, f_m$ achieve the inf in (2.1) and $m = H_B(\epsilon, P)$. Therefore,

$$\begin{aligned} \sup_{\mathbf{F}} |P_n(f) - P(f)| \leq\ & \max_i |P_n(f_i^U) - P(f_i^U)| \\ & + 2\max_i |P_n(f_i^L) - P(f_i^L)| + 2\epsilon \to 2\epsilon \text{ as } n \to \infty. \end{aligned} \tag{2.3}$$

The proposition follows since $\epsilon > 0$ is arbitrary.

(ii) *Metric related criteria*

Proposition 2.2. *Assume there is a metric d on $\mathbf{F}$ such that*

$$\mathbf{F} \text{ is } d \text{ compact.} \tag{2.4}$$

$$\text{If } d(f_n, f) \to 0 \text{ then } P_n f_n \to Pf \text{ and } Pf_n \to Pf. \tag{2.5}$$

Then $\sup_{f \in \mathbf{F}} |P_n f - Pf| \to 0$.

Proof. Suppose the conclusion fails. Then there is $\delta > 0$ and f_n such that

$$|P_{n'} f_{n'} - Pf_{n'}| \geq \delta \text{ for a sequence } n'.$$

Since $\mathbf{F}$ is compact, there is a subsequence $\{f_{n''}\}$ of $\{f_{n'}\}$ such that $d(f_{n''}, f) \to 0$ for some f. Because of (2.5) and the triangle inequality, this is a contradiction:

$$|P_{n''} f_{n''} - Pf_{n''}| \leq |P_{n''} f_{n''} - Pf| + |Pf_{n''} - Pf| \to 0.$$

Two extensions of this criterion are useful. We say $\mathbf{F}_\epsilon$ is an ϵ *approximating class* to $\mathbf{F}$ for P if $\forall f \in \mathbf{F}$ there exist $f_\epsilon \in \mathbf{F}_\epsilon$ such that $|P_n(f_\epsilon - f)| \leq \epsilon$, $|P(f_\epsilon - f)| \leq \epsilon$, whenever $P_n f \to Pf$ for all $f \in \mathbf{F}$.

Remark 2.1. If

$$\{P_n\} \text{ are tight} \tag{2.6}$$

a simple ϵ approximating class is given by $\mathbf{F}_{K_\epsilon}$ where

$$\mathbf{F}_K \equiv \{f I_K : f \in \mathbf{F}\}$$

for K compact (I_K denoting the indicator of K), and

$$P_n(K_\epsilon) \geq 1 - \frac{\epsilon}{c_0} \text{ all } n, P(K_\epsilon) \geq 1 - \frac{\epsilon}{c_0}.$$

Evidently, if $\mathbf{F}$ has for each $\epsilon > 0$ an approximating class $\mathbf{F}_\epsilon$ which satisfies the conditions of either Proposition 2.1 or 2.2 and

$$P_n f \to Pf \quad \forall f \in \mathbf{F}_\epsilon,\ \epsilon > 0,$$

then $\mathbf{F}$ is a Pólya class.

It is also convenient to have a closely related nonmetric criterion.

Proposition 2.3. *Let* (2.6) *hold and* $\mathbf{F}_{K_\epsilon}$ *be as above. Suppose that for every sequence* $\{f_n\}$, $f_n \in \mathbf{F}_{K_\epsilon}$ *there exists a subsequence* $\{f_{n'}\}$ *and* $f \in \mathbf{F}_{K_\epsilon}$ *such that whenever* $x_{n'} \to x$ *in* K_ϵ *it follows that*

$$f_{n'}(x_{n'}) \to f(x), \ f(x_{n'}) \to f(x) \tag{2.7}$$

except for $x \in N$ such that $P[N] = 0$. Then $\mathbf{F}$ *is a Pólya class.*

Proof. It is enough to consider $\mathbf{F} = \mathbf{F}_{K_\epsilon}$. Since K_ϵ is separable, by the Dudley-Skorokhod-Wichura theorem there exist measurable X_n, $X : [0,1] \to K_\epsilon$ where $[0,1]$ is endowed with the uniform distribution such that the distribution of X_n is P_n, and that of X is P and

$$X_n(w) \to X(w) \text{ for all } w.$$

Suppose the proposition is false. Then there exist $f_n \in \mathbf{F}$, $|P_n f_n - P(f_n)| \geq \epsilon$, all n. If $f_{n'}$ is the subsequence converging to f which is assumed to exist, then,

$$\begin{aligned} f_{n'}(X_{n'}(w)) &\to f(X(w)) \\ f(X_{n'}(w)) &\to f(X(w)) \end{aligned}$$

except on $\{w : X(w) \in N\}$. But, by hypothesis that set has measure 0. By dominated convergence,

$$\begin{aligned} P_{n'} f_{n'} &\to Pf \\ P_{n'} f &\to Pf \end{aligned}$$

and we have a contradiction.

Note: The above argument shows that this proposition, with $\mathbf{F} = \mathbf{F}_{K_\epsilon}$, is a special case of Proposition 2.2, if d there is taken to be the metric of convergence in P probability.

3. Stability Properties of Pólya Classes

Let $\mathbf{G}$ be a compact subset of $C[-c_0, c_0]^k$, the continuous bounded functions on $[-c_0, c_0]^k$ endowed with the sup norm.

Proposition 3.1. *Let* $\mathbf{F_G} = \{g(f_1, \cdots, f_k) : f_1, \cdots, f_k \in \mathbf{F}, g \in \mathbf{G}\}$, *and* $\mathbf{F}, P$ *satisfy the conditions of Proposition* 2.2. *Then* $\mathbf{F_G}$ *is a Pólya class.*

Proof. Metrize $\mathbf{F_G}$ by,

$$\tilde{d}(g_1(f_1, \cdots, f_k),\ g_2(f_1^*, \cdots, f_k^*)) = \sum_{i=1}^{k} d(f_i, f_i^*) + \|g_1 - g_2\|.$$

Apply Proposition 2.2.

Special cases:

Algebraic stability: Let k be an integer, $k = k_0 k_1$.

Let $g(x) = \sum_{i=1}^{k_1} \alpha_i \Pi_{j=1}^{k_0} x_{ij}$

$$\text{where} \quad x = \{x_{ij}\}\ 1 \leq i \leq k_1,\ 1 \leq j \leq k_0$$

and $|\alpha_i| \leq M, 1 \leq i \leq k$. More generally by taking $\mathbf{G} = \{g = \sum \alpha_{r_1,\cdots,r_k} \prod_{i=1}^k x_i^{r_i} : r_i \geq 0, \sum_{i=1}^k r_i = m$ and $|\alpha_{r_1,\dots,r_k}| \leq M < \infty$ for all $r_1, \dots, r_k\}$ we obtain the result that the set $\mathbf{F_G}$ of all polynomials of degrees m in k variables $f_1, \dots, f_k$ with bounded coefficients is a Pólya class whenever $\mathbf{F}$, P satisfy the conditions of Proposition 2.2.

Lattice operations:

The Pólya property is similarly inherited if $\mathbf{F_G} = \{f_1 \otimes f_2 : f_1, f_2 \in \mathbf{F}\}$ where $\otimes$ is max or min.

These operations can clearly be iterated a *finite* number of times. In particular, note that the Pólya property is preserved under a *finite* number of field operations.

4. Applications

Throughout this section we assume

$$P_n \Rightarrow P \tag{4.1}$$

Example 4.1. *Sufficiency of the Billingsley-Topsoe conditions.* (This is a paraphrase of their argument.) Let $\{A_n\}$ be a partition of X into sets of diameter $\leq \epsilon$ which are continuity sets of P. Then $\mathbf{F}_\epsilon = \{f = \sum c_k 1_{A_k} : |c_k| \leq c_0\}$ is a Pólya class. To see this note that, by Proposition 2.1, $\mathbf{F}_{m\epsilon} = \{f = \sum_{k=1}^m c_k 1_{A_k} : |c_k| \leq c_0\}$ is a Pólya class. Further, since $\sum_{j=1}^m P_n(A_j) \to \sum_{j=1}^m P(A_j)$ for any $\gamma > 0$, there exists $m(\gamma)$ such that $\sum_{j=m(\gamma)+1}^\infty P_n(A_j) \leq \gamma$, $\sum_{j=m(\gamma)+1}^\infty P(A_j) \leq \gamma$ for all n. Evidently $\mathbf{F}_{m(\gamma)\epsilon}$ is a γ approximating class to $\mathbf{F}_\epsilon$ and the claim follows. Now, if $f \in \mathbf{F}$, there exist $f_\epsilon \in \mathbf{F}_\epsilon$ such that

$$\begin{aligned} P_n|f - f_\epsilon| &\leq \delta + c_0 \sum \{P_n(A_m) : w_f(A_m) > \delta\} \\ &\leq \delta + c_0 |P_n(\sum 1(w_p(A_k) > \delta) 1_{A_k}) \\ &\quad - P(\sum 1(w_f(A_k) > \delta) 1_{A_k})| + P[x : w_p(x, \epsilon) > \delta]. \end{aligned}$$

Hence,

$$\sup_{\mathbf{F}} |P_n f - Pf| \leq (1 + c_0) \sup_{\mathbf{F}_\epsilon} |P_n f - Pf|$$
$$+ \delta + \sup_{\mathbf{F}} P[x : w_f(x, \epsilon) > \delta].$$

Since δ is arbitrary and $\mathbf{F}_\epsilon$ is a Pólya class the result follows.

Example 4.2. *Convex sets in Euclidean space.* This example was first treated by Ranga Rao (1962). Take $X = R^p$ and let $\mathbf{C} = \{$ convex sets on $X\}$. Let d denote the Hausdorff metric on $\mathbf{C}$.

Let $\mathbf{C}_0$ denote a subcollection of elements of $\mathbf{C}$. Let K_δ denote the cube in R^d centered at the origin, sides parallel to the axes, and having edges of length $[1/\delta]$. Let $\mathbf{C}^*_{0,\delta}$ denote the closure (for the metric d) of the collection of sets $\mathbf{C}_0 \cap K_\delta$. Assume

(4.2) Every set in $\mathbf{C}^*_{0,\delta}$ is a P-continuity set for all $\delta > 0$.

Proposition 4.1. *Let* $\mathbf{F}$ *denote the collection of indicator functions of sets in* $\mathbf{C}_0$. *Assume* (4.1) *and* (4.2)*: Then* $\mathbf{F}$ *is a Pólya class.*

Proof. $\mathbf{C}^*_{0,\delta}$ is, for each δ, compact for the Hausdorff metric by Blaschke's theorem (Blaschke (1916) §18). It is easy to see that $d(C_n, C) \to 0$ (in the Hausdorff metric d) implies that $1_{C_n}(x) \to 1_C(x)$ except possibly for $x \in \partial C$. Conditions (2.4) and (2.5) therefore apply to $\mathbf{F}_\delta = \{1_C : C \in \mathbf{C}^*_{0,\delta}\}$ which then is a Pólya class by Proposition 2.2. Now apply Proposition 2.3.

Remark 4.1. It is essential to put the P-continuity hypothesis on $\mathbf{C}^*_{0,\delta}$ rather than just on $\mathbf{C}_0 \cap K_\delta$. If $\mathbf{C}_0$ consists of proper ellipses in R^2, then $\mathbf{C}^*_{0,\delta}$ will contain straight lines, and so (4.2) has to be verified over a strictly larger collection than $\mathbf{C}_0 \cap K_\delta$.

Remark 4.2. Let $\mathbf{C}_0$ satisfy the conditions of Proposition 4.1. Then, by Proposition 3.1, the class of sets obtained by finite field operations on $\mathbf{C}_0$, will be a Pólya class. This result therefore covers the usual simple statistical operations on Euclidean spaces.

Remark 4.3. Suppose $\mathbf{C}_0 = \mathbf{C}$. The hypotheses of Proposition 4.1 will then be satisfied provided P has a density with respect to Lebesgue measure. The collection $\mathbf{C}$ is, of course, not a Vapnik-Cervonenkis class.

Example 4.3. *Cylinder sets.* Let B be an infinite dimensional Banach space. The generalization of Example 4.2 to B will fail in general. Indeed, if $\mathbf{C}_0$ is the collection of half-spaces in B, then $\mathbf{C}_0$ is, in general, not a Pólya class. This fact can be deduced, for example, from well-known necessary conditions for the general Glivenko-Cantelli theorem. On the other hand, certain interesting results are still possible.

Let B^* denote the dual of B; for $m \in B^*$ and $x \in B$, let $\langle m, x\rangle = m(x)$ denote the usual duality relationship. For an integer $p \geq 1$, let $K \equiv K_{\delta_0}$ denote a fixed cube in R^p, as described in the previous example. Let

> $\mathbf{C}^e$ denote the restriction to K of a collection of convex sets, possibly perturbed by finite field operations (cf. Remark 4.2).

Put the Hausdorff metric d on $\mathbf{C}^e$; then $\mathbf{C}^e$ is precompact. For convenience we assume $\mathbf{C}^e$ is already compact; this entails no loss.

Let D^* denote a subset of B^*. Assume that

$$D^* \text{ is weak star compact.} \tag{4.3}$$

For $p \geq 1$ chosen above, define $\mathbf{A}$, a collection of sets $A \subset B$ by;

$$A \in \mathbf{A} \text{ iff } A = \{x \in B : (\langle m_1, x\rangle, \ldots, \langle m_p, x\rangle) \in C\} \tag{4.4}$$

where $m_i \in D^*$, $1 \leq i \leq p$, $C \in \mathbf{C}^e$. Assume, for the collection $\mathbf{A}$:

$$\text{Each } A \in \mathbf{A} \text{ is a } P\text{-continuity set.} \tag{4.5}$$

Proposition 4.2. *Let* $\mathbf{F}$ *denote the collection of indicator functions of sets in* $\mathbf{A}$. *Assume* (4.1), (4.2), (4.3), (4.5). *Then* $\mathbf{F}$ *is a Pólya class.*

Proof. Since $P_n \Rightarrow P$ it is enough (by applying Remark 2.1 and Proposition 2.3) to replace B by a compact K which is necessarily separable. Now if $A \longleftrightarrow (C_A, m_{1A}, \ldots, m_{pA})$ with an obvious notation let

$$d(1_A, 1_B) = d_H(C_A, C_B) + \sum_{i=1}^{p} d_W(m_{iA}, m_{iB})$$

when d_H is the Hausdorff metric and d_W metrizes the weak topology induced by K on B^* for which D^* is also necessarily compact. Apply Proposition 2.2.

Illustration 4.1. Half-spaces. Fix D^* above as a weak * compact subset of B^*. Let $\mathbf{C}^e$ consist of the sets $\{[a, b] : a \in R, b \in R, -c \leq a \leq b \leq c\}$ for some fixed positive c. Then $\mathbf{A}$ is defined to consist of the sets

$$\{x \in B \cdot \langle m, x\rangle \in [a, b], m \in D^*, [a, b] \in \mathbf{C}\}.$$

This gives a collection of "strips" for which the Pólya property holds.

A fancier collection—consisting e.g. of "wedges"—can then be obtained by the finite field operations.

Illustration 4.2. Convergence on $C[0,1]$. In this example $X \equiv C[0,1] = B$, and there are processes X_n, X such that X_n is $C[0,1]$ valued, and $X_n \Rightarrow X$. Statistical interest would focus on the special case that X is Brownian motion or Brownian Bridge. Here the dual of B consists of finite signed measures on $[0,1]$; thus take, for illustrative purposes, $D^* = \{$ all measures μ_T, where μ_T is unit mass at $t, 0 \leq t \leq 1\}$. This set is weak star compact. For the collection $\mathbf{C}^e$ of convex sets in R^1, let us take the intervals $[-s,s]$, $s > 0$; this collection is not restricted to a compact K of R (as specified in Theorem 4.2), but thanks to the properties of the sup norm, this inconvenience does not matter. Proposition 4.2 therefore implies, for the special choices of D^*, $\mathbf{C}^e$, that:

$$\sup_{\substack{0 \leq t \leq 1 \\ u \in R}} |P\{|X_n(t)| \leq u\} - P\{|X(t)| \leq u\}| \to 0. \tag{4.6}$$

Since the unit ball in B^* is * compact, and since this ball consists of signed measures μ such that $\|\mu\| \leq 1$ ($\| \ \|$ = variation norm), the result (4.6) extends to

$$\sup_{\substack{\|\mu\| \leq 1 \\ u \in R+}} |P\{|\int X_n(t)\mu(dt)| \leq u\} - P\{|\int X(t)\mu(dt)| \leq u\}| \to 0. \tag{4.7}$$

Extensions of Illustration 4.2. The result continues to hold if $D[0,1]$ replaces $C[0,1]$, as long as $X \in C[0,1]$. One can also get a similar result if the norm is the μ ess sup norm on bounded functions as long as μ is absolutely continuous with respect to P.

Simple variants of Example 4.2. Again take $X = B$, a Banach space. The cylinder set classes worked in Example 4.2 because they could be neatly parametrized by a compact set. Obviously, other sets can be so indexed.

Illustration 4.3. Let $\beta(c,r) \equiv \{x \in B : |x - c| \leq r\}$ be the ball in B, centered at $c \in B$, and having radius r. Let $\mathbf{F}$ consist of all $\beta(c,r)$ such that (a) $r \leq r_0$ and (b) c belongs to a (norm) compact set N in B. Then (subject to the usual continuity-set conditions), $\mathbf{F}$ is a Pólya class. This class can be augmented by finite field operations. The addition of classes such as those in Example 4.2 should be particularly interesting.

Illustration 4.4. This example illustrates the fact that the collection of sets need not be indexed by a *compact* set. Let B_1 denote the unit ball in B, an infinite dimensional Banach space. Then there are balls $B(x_i, 1/4)$ which are disjoint and which are entirely within B_1 (see Millar (1988), for example). These balls are indexed by $\{x_i\}$, and obviously the set $\{x_i, 1 \le i < \infty\}$ is not compact. Nevertheless, the Pólya property continues to hold for $\{B(x_i, 1/4), 1 \le i\}$.

Example 4.4. *A theorem of Beran, LeCam, Millar, Varadarajan.* Example 4.1 was derived as an application of Proposition 2.1. All others, so far, were applications of Propositions 2.2, 2.3 but could have been derived, sometimes under stronger conditions, from Proposition 2.1. We conclude with an application of a slight extension of Proposition 2.1 which gives directly and fairly simply the following recent result of Beran, Le Cam, Millar (1987) generalizing a result of Varadarajan (1958).

Theorem. *Let $\{P_n\}$ be a sequence of probability measures on a metric space D which is δ tight. That is, $\forall \epsilon > 0$, there exists K compact, $\delta_n \downarrow 0$, $P_n(K^{\delta_n}) > 1-\epsilon$, $\forall n$, where $K^\delta = \{x : d(x, K) < \delta\}$. Let $\widehat{P}_n$ be the empirical distribution of $\{X_{ni}\}$, $1 \le i \le j_n$* i.i.d. *random variables drawn from P_n,*

$$P_n(A) = \frac{1}{j_n} \sum_{i=1}^{j_n} 1_A(X_{ni}).$$

Let $\mathbf{F} = \{f : \|f\|_{\mathrm{BL}} \le C_0\}$ where $\|f\|_{\mathrm{BL}} = \sup_{x,y}\{|f(x) - f(y)|/d(x,y)\}$ and C_0 is fixed say, $C_0 = 1$. Then,

$$\sup_{f \in \mathbf{F}} |\widehat{P}_n f - P_n f| \to 0 \text{ in } P_n^{j_n} \text{ probability.} \tag{4.8}$$

Remark 4.4. We can interpret (4.8) as

$$\rho(\widehat{P}_n, P_n) \to 0 \text{ in } P_n^{j_n} \text{ probability} \tag{4.9}$$

where ρ is the BL metric metrizing weak convergence. A further straightforward extension motivated by bootstrap applications is given by Beran et al. Indeed, let P_n itself be random, say $P_n(\cdot)$ is a map from a probability space $(\Omega, \mathbf{a}, P)$ to $L_\infty(\mathbf{F})$ endowed with the σ field induced by the open balls in $L_\infty(\mathbf{F})$ and suppose that,

$$P[\{P_n\} \text{ is } \delta \text{ tight}] = 1.$$

Then (4.9) can be extended to

$$\rho(\widehat{P}_n, P_n) \to 0 \text{ in } P \text{ probability.} \tag{4.10}$$

Statement (4.10) can be deduced from the theorem by using the Skorokhod-Dudley-Wichura a.s. convergence theorem.

Proof of (4.8).

Step 1. To show $\sup_{f \in \mathbf{F}} |\widehat{P}_n f - P_n f| \to 0$ it suffices to show

$$\sup_{f \in \mathbf{F}_n} |\widehat{P}_n f - P_n f| \to 0$$

where $\mathbf{F}_n = \mathbf{F}_{K^{\delta_n}}$. This follows from δ tightness since

$$\text{(a) } P_n(K^{\delta_n}) \geq 1 - \epsilon$$

$$\text{(b) } P_n\{\widehat{P}_n(K^{\delta_n}) \leq 1 - \epsilon^{1/4}\} \leq \frac{1}{j_n \epsilon^{1/4}}.$$

Part (b) is shown by noting that $j_n \widehat{P}_n(K^{\delta_n})$ is binomial $B(j_n, p_n)$ where $p_n = P_n(K^{\delta_n})$, and then using Chebychev.

Step 2. Note that $\mathbf{F}_K$ is compact by Arzela-Ascoli. Then, $\mathbf{F}_n = \mathbf{F}_{K^{\delta_n}}$ is bracketed within δ_n (see below) because of the form of the BL norm.

More precisely, fix δ_n and $f \in \mathbf{F}_{K^{\delta_n}}$. Let $f_1, \ldots, f_{k_n}$ be a δ_n-grid on $\mathbf{F}_K$ (where the metric is the sup norm). If $f \in \mathbf{F}_n = \mathbf{F}_{K^{\delta_n}}$ then there is i such that

$$\sup_{x \in K} |f(x) - f_i(x)| < \delta_n.$$

If $x \in K^{\delta_n}$, then there exists, $x' \in K$ with $d(x, x') < \delta_n$. Thus

$$|f(x) - f(x')| \leq d(x', x) < \delta_n \text{ since } \|f\|_{\mathrm{BL}} \leq 1. \tag{4.11}$$

This implies

$$f(x) \leq f(x') + \delta_n \leq f_i(x') + 2\delta_n \leq f_i(x) + 3\delta_n$$

by (4.11) and f_i being BL.
Thus if $f_1, \ldots, f_{k_n}$ is a δ_n-grid for $\mathbf{F}_n$ in sup norm, $f_i \pm 3\delta_n$ brackets $\mathbf{F}_{\delta_n}$ within δ_n.

Step 3. Let $\mathbf{G}_n$ denote the collection of functions that bracket $\mathbf{F}_n = \mathbf{F}_{K^{\delta_n}}$ – a finite set given in Step 2. Since P_n is δ-tight, every subsequence has a weakly convergent subsequence, so we can assume WLOG that

$$P_n \Longrightarrow P$$

$$\text{i.e. } \sup_f |P_n f - Pf| \to 0.$$

Thus it suffices to show by the argument of Proposition 2.1, that

$$\sup_{f \in \mathbf{G}_n} |\widehat{P}_n f - Pf| \xrightarrow{P_n} 0.$$

Step 4. **Lemma 4.1.** *Let $\{X_t^n, t \in I\}, \{X_t, t \in I\}$ be real stochastic processes, where I is countable. Suppose the fidis (finite dimensional distributions) of X^n converge to those of X. Then there exist $\{\tilde{X}_t^n, t \in I\}$, $\{\tilde{X}_t, t \in I\}$ with the same fidis as X_n, X defined on $[0,1]$ endowed with Lebesgue measure such that*

$$\tilde{X}_t^n(w) \to \tilde{X}_t(w) \text{ for all } t \in I$$

except for w in a set of Lebesgue measure 0.

Proof. $X^n \equiv \{X_t^n, t \in I\}$ is R^I valued (with the topology of co-ordinatewise convergence). R^I is then a separable, complete metric space.

The fidi convergence implies $X^n \Rightarrow X$ as elements of R^I. Apply Skorokhod.

Step 5. Apply the lemma to the stochastic process

$$X_t^n = \widehat{P}_n(t) - P(t), \; t \in I \equiv \mathbf{G} \equiv U_n \mathbf{G}_n.$$

The fidis converge here, so by Lemma 4.1 we can assume by Skorokhod that $\forall w$

$$\widehat{P}_n(w, f) - Pf \xrightarrow{P_n} 0 \quad \forall f \in \mathbf{G}.$$

Evidently, we may, without loss of generality, take the $\mathbf{G}_n$ nested. Then if $\delta_{m(\epsilon)} \leq \frac{\epsilon}{6}$, $\mathbf{G}_{m(\epsilon)}$ is an ϵ approximating class to $\mathbf{G}$ for the sequence $\{\widehat{P}_n(w, \cdot)\}, P$. By Proposition 2.3 we can conclude that

$$\sup_{\mathbf{G}} |\widehat{P}_n(w, f) - P(f)| \to 0 \text{ a.e. } w$$

and hence that (4.12) holds. The theorem follows.

References

Beran, R. J., LeCam, L. and Millar, P. W. (1987). Convergence of stochastic empirical measures. *J. Multivariate Anal.* **23**, 159-168.

Beran, R. and Millar, P. W. (1987). Stochastic estimation and testing. *Ann. Statist.* **15**, 1131-1154.

Billingsley, P. and Topsoe, F. (1967). Uniformity in weak convergence. *Z. Warsch. verw. Gebiete* **7**, 1-16.

Blaschke, W. (1916). Kreis and Kugel. *Leipzig.*

Huber, P. J. (1981). *Robust Statistics.* John Wiley, New York.

Millar, P. W. (1988). Gamma-Funnels in the domain of a probability. Technical Report 177, Department of Statistics, University of California.

Papantoni-Kazakos, P. and Gray, R. M. (1979). Robustness of estimators on stationary observations. *Ann. Probab.* **7**, 989-1002.

Pólya, G. (1920). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und Momenten problem. *Math. Zeitschrift,* **8**.

Ranga Rao, R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33**, 659-680.

Varadarajan, V. S. (1958). Weak convergence of measures on separable metric spaces. *Sankhyā Ser. A* **19**, 15-22.

Wichura, M. J. (1970). On the construction of almost uniformly convergent random variables with given weakly convergent image laws. *Ann. Math. Statist.* **41**, 284-291.

Department of Statistics, University of California, Berkeley, CA 94720, U.S.A.

Department of Statistics, University of California, Berkeley, CA 94720, U.S.A.

(Received January 1991; accepted August 1991)