

OBSERVING A PREFERRED DISTRIBUTION

I. H. DINWOODIE

(Communicated by Wei Y. Loh)

ABSTRACT. We give asymptotic results for the number of empirical laws that one must observe in order for one of them to be close to a certain distribution.

1. INTRODUCTION

Let X_1, X_2, \dots be a sequence of random variables in the interval $[0, 1]$ defined on a probability triple (Ω, \mathcal{A}, P) . Let $F_{m,n}$ be the empirical distribution function on $[0, 1]$ defined by

$$(1.1) \quad F_{m,n}(t) = \frac{1}{n} \sum_{i=m}^{n+m-1} 1_{[0,t]}(X_i),$$

and let F_n be the ordinary empirical distribution function given by $F_{1,n}$. We are interested in the following question: through what number of the $F_{m,n}$ for fixed n must one search to find any chosen distribution function G ? Roughly speaking, the answer is e^{nI} , where I is the large deviation rate for the empirical law at the chosen distribution. The results are applications of large deviation results for the empirical distribution. We will assume that the process $(X_n)_{n \geq 1}$ is a chain with complete dependence on the past in Theorem 2.2, and we will assume it is i.i.d. for Theorem 2.3. The i.i.d. case has been treated in greater generality by de Acosta and Kuelbs [1].

2. MAIN RESULTS

Let $L_{(1,1)}$ denote the set of functions on $[0, 1]$ which are Lipschitz for the constant 1 and bounded by 1. Let \mathbf{D} be the set of distribution functions on $[0, 1]$. Define the metric $\|\cdot\|$ on \mathbf{D} by

$$\|F - G\| = \sup_{f \in L_{(1,1)}} \left| \int f dF - \int f dG \right|.$$

This metrizes the weak topology on \mathbf{D} , the topology generated by finite families of bounded and continuous function on $[0, 1]$. Let $[0, 1]^{-N} = \{(x_{-i}) : x_{-i} \in [0, 1], i \geq 0\}$. Let μ be an initial distribution on $[0, 1]^{-N}$, and let $\{\pi(\bar{x}, \cdot) :$

Received by the editors March 31, 1992.

1991 *Mathematics Subject Classification.* Primary 62E20.

Key words and phrases. Empirical law, large deviations, empirical distribution.

$\bar{x} \in [0, 1]^{-N}$ be a family of transition probabilities on $[0, 1]$ with complete dependence on the past. There exists a process X_1, X_2, \dots on $(\Omega, \mathcal{A}, P_\mu)$ that satisfies

$$P_\mu(X_1 \in A_1, \dots, X_n \in A_n) = \int \mu(d\bar{x}) \int_{A_1} \pi(\bar{x}, dx_1) \int_{A_2} \pi(\bar{x}, x_1, dx_2) \cdots \int_{A_n} \pi(\bar{x}, x_1, \dots, x_{n-1}, dx_n)$$

where $\bar{x} = (\dots, x_{-2}, x_{-1}, x_0) \in [0, 1]^{-N}$. Chains of infinite order were introduced by Onicescu and Mihoc [8]. One of their examples was the Polya urn scheme, which is now considered a basic example of an exchangeable sequence. A general discussion with further applications is in [4]. Recently these processes have been studied in connection with symbolic dynamics (see [3]).

Let $C([0, 1]^{-N})$ denote the space of continuous functions on the compact product space $[0, 1]^{-N}$. The norm of a function $h \in C([0, 1]^{-N})$ will be the supremum norm of h , making the space a Banach space. If $y \in [0, 1]$ and $u \in C([0, 1]^{-N})$, we let $u_y \in C([0, 1]^{-N})$ be defined by

$$u_y(\bar{x}) = u(\dots, x_{-1}, x_0, y).$$

Suppose the family $\{\pi(\bar{x}, \cdot)\}$ has the property that the operator T on $C([0, 1]^{-N})$ given by

$$T(u)(\bar{x}) = \int_0^1 u_y(\bar{x})\pi(\bar{x}, dy)$$

indeed maps the space $C([0, 1]^{-N})$ into itself and is a compact operator. The operator T and related operators were studied by Ionescu Tulcea in [7].

For $x = (x_1, \dots, x_n) \in [0, 1]^n$ we will use the notation

$$\pi_x(\bar{x}, \cdot) = \pi(\bar{x}, x_1, \dots, x_n).$$

Let c_0, c_1, c_2, \dots be a sequence of numbers such that for all $\bar{x}, \bar{y} \in [0, 1]^{-N}$ one has

$$(2.1) \quad \pi(\bar{x}, \cdot) \leq c_0\pi(\bar{y}, \cdot), \quad \pi_x(\bar{x}, \cdot) \leq c_n\pi_x(\bar{y}, \cdot), \quad x \in [0, 1]^n,$$

and we suppose that $c^\infty = \prod_{n=0}^\infty c_n < \infty$. Assume as well that $\pi(\bar{x}, \cdot)$ assigns positive probability to open sets for $\bar{x} \in [0, 1]^{-N}$.

Let $U = \{u \in C([0, 1]^{-N}) : u > 0\}$, and let $\lambda: \mathbf{D} \rightarrow [0, \infty]$ be given by

$$(2.2) \quad \lambda(G) = \sup_{u \in U} \int_0^1 \inf_{\bar{x}} \log \left[\frac{u_y(\bar{x})}{(Tu)_y(\bar{x})} \right] dG(y).$$

Minor variations on the arguments of Dinwoodie [5] for the compact state space $S = [0, 1]$ give the following large deviation theorem uniform in the initial distribution. The uniformity in the starting point is not discussed explicitly in this reference but follows easily from the methodology.

Theorem 2.1. *The sequence of empirical distribution functions $\{F_n\}$ satisfies*

$$\limsup \frac{1}{n} \log \sup_{\bar{x}} P_{\bar{x}}\{F_n \in C\} \leq - \inf_{G \in C} \lambda(G), \quad C \subset \mathbf{D} \text{ closed,}$$

$$\liminf \frac{1}{n} \log \inf_{\bar{x}} P_{\bar{x}}\{F_n \in U\} \geq - \inf_{G \in U} \lambda(G), \quad U \subset \mathbf{D} \text{ open,}$$

and the function λ vanishes at a unique point $F_0 \in \mathbf{D}$.

The fact that λ vanishes at a unique point F_0 is established in Example 3.2 of [5]. This is based on the result that points where λ vanishes have a certain invariance property for T , and an ergodic theorem of Ionescu Tulcea [7] shows that this property is satisfied by only one distribution F_0 under (2.1). We will use the lower bound of Theorem 2.1 to derive the following theorem, which is closely related to the law of Erdős and Rényi [6].

Theorem 2.2. *Let $\{G_1, \dots, G_d\} \subset \mathbf{D}$ satisfy $R = \max_i \lambda(G_i) < \infty$. Then*

$$\max_{1 \leq i \leq d} \min_{1 \leq m \leq \exp(nR)} \|F_{m,n} - G_i\| \rightarrow 0$$

as $n \rightarrow \infty$ (P_μ a.s.) for any initial distribution μ on $[0, 1]^{-N}$.

Proof. We show that for each $\varepsilon > 0$,

$$P_\mu \left\{ \limsup_n \max_{1 \leq i \leq d} \min_{1 \leq m \leq \exp(nR)} \|F_{m,n} - G_i\| \geq \varepsilon \right\} = 1.$$

If $R = 0$, then $G_i = F_0$ for each $1 \leq i \leq d$, where $\lambda(F_0) = 0$. Then the statement above follows from the weak convergence of the empirical distribution function $F_{1,n} = F_n$ to F_0 . Assume henceforth that $R > 0$.

Let

$$A_n = \left\{ \max_{1 \leq i \leq d} \min_{1 \leq m \leq \exp(nR)} \|F_{m,n} - G_i\| \geq \varepsilon \right\} \in \mathcal{A}.$$

Then

$$\{A_n \text{ i.o.}\}^c = \bigcup_{n \geq 1} \bigcap_{m \geq n} \left\{ \max_{1 \leq i \leq d} \min_{1 \leq m \leq \exp(nR)} \|F_{m,n} - G_i\| < \varepsilon \right\}.$$

Thus it is enough to show that $P_\mu\{A_n \text{ i.o.}\} = 0$. By the Borel-Cantelli lemma it is enough to show that $\sum_{n=1}^{\infty} P_\mu(A_n) < \infty$. Let $[x]$ denote the greatest integer less than or equal to x . Now

$$\begin{aligned} P_\mu(A_n) &= P_\mu \left\{ \max_{1 \leq i \leq d} \min_{1 \leq m \leq \exp(nR)} \|F_{m,n} - G_i\| \geq \varepsilon \right\} \\ &\leq \sum_{i=1}^d P_\mu \left(\bigcap_{j=1}^{[\exp(nR)/n]} \{\|F_{(j-1)n+1,n} - G_i\| \geq \varepsilon\} \right) \\ &\leq \sum_{i=1}^d \left(\sup_{\bar{x}} P_{\bar{x}} \{\|F_{1,n} - G_i\| \geq \varepsilon\} \right)^{[\exp(nR)/n]} \\ &= \sum_{i=1}^d \left(1 - \inf_{\bar{x}} P_{\bar{x}} \{\|F_{1,n} - G_i\| < \varepsilon\} \right)^{[\exp(nR)/n]}. \end{aligned}$$

Now if $\lambda(G_i) > 0$, there is a $\delta > 0$ such that for n sufficiently large,

$$\inf_{\bar{x}} P_{\bar{x}} \{\|F_{1,n} - G_i\| < \varepsilon\} \geq e^{-n(\lambda(G_i) - \delta)} \geq e^{-n(R - \delta)}$$

using Theorem 2.1, since the function λ on the segment of probability measures joining F_0 to G_i is strictly less than $\lambda(G_i)$ by the convexity of λ and the fact

that λ vanishes at the point F_0 . But even when $\lambda(G_i) = 0$, we can still get the inequality

$$\inf_{\bar{x}} P_{\bar{x}}\{\|F_{1,n} - G_i\| < \varepsilon\} \geq e^{-n(R-\delta)}, \quad i = 1, \dots, d,$$

for a $\delta > 0$ independent of i and for n sufficiently large, since $R > 0$. Thus

$$P_{\mu}(A_n) \leq \sum_{i=1}^d (1 - e^{-n(R-\delta)})^{\lfloor \exp(nR)/n \rfloor} \leq d(1 - e^{-n(R-\delta)})^{\lfloor \exp(nR)/n \rfloor}.$$

Therefore, $\sum_{n=1}^{\infty} P_{\mu}(A_n) < \infty$, which proves the theorem. \square

We also state the following result for i.i.d. random variables, which could be deduced from the previous theorem but requires none of the technical hypotheses that figured in the treatment of dependent processes. The following result also follows essentially from Theorem 5 of [1], although there are a few technical points one must address to make the implication completely rigorous. The lower large deviation estimates are well-known i.i.d. results (see [2]). Suppose for the following theorem that X_1, X_2, \dots is an i.i.d. sequence in $[0, 1]$ with common distribution function F_0 . For $F \in \mathbf{D}$, define

$$K(F) = \begin{cases} \int_0^1 \log[dF/dF_0] dF & \text{if } F \ll F_0, \\ \infty & \text{otherwise.} \end{cases}$$

Theorem 2.3. *If $G \in \mathbf{D}$ satisfies $K(G) < \infty$, then*

$$\min_{1 \leq m \leq \exp(nK(G))} \|F_{m,n} - G\| \rightarrow 0$$

as $n \rightarrow \infty$ (a.s.).

Proof. The proof is a simplified version of the previous proof. \square

REFERENCES

1. A. de Acosta and J. Kuelbs, *Limit theorems for moving averages of independent random vectors*, Z. Wahrsch. Verw. Gebiete **64** (1983), 67–123.
2. R. R. Bahadur and S. L. Zabell, *Large deviations of the sample mean in general vector spaces*, Ann. Probab. **7** (1979), 587–621.
3. H. Berbee, *Chains with infinite connections: uniqueness and Markov representations*, Probab. Theory Related Fields **76** (1987), 243–253.
4. A. Blanc-Lapierre and R. Fortet, *Theory of random functions*, Gordon and Breach, New York, 1967.
5. I. H. Dinwoodie, *Occupation measures for chains of infinite order*, Stochastic Process Appl. (to appear).
6. P. Erdős and A. Rényi, *On a new law of large numbers*, J. Analyse Math. **23** (1970), 103–111.
7. C. T. Ionescu Tulcea, *On a class of operators occurring in the theory of chains of infinite order*, Canad. J. Math. **11** (1959), 112–121.
8. O. Onicescu and G. Mihoc, *Sur les chaînes de variables statistiques*, Bull. Sci. Math. **59** (1935), 174–192.