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

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

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 \( e^{nt} \), where \( t \) 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 \( D \) be the set of distribution functions on \([0, 1]\). Define the metric \( \| \cdot \| \) on \( D \) by

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

This metrizes the weak topology on \( D \), 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 \( \mu \) be an initial distribution on \([0, 1]^{-N}\), and let \( \pi(\bar{x}, \cdot) :
\( \bar{x} \in [0, 1]^N \) be a family of transition probabilities on \([0, 1]\) with complete dependence on the past. There exists a process \( X_1, X_2, \ldots \) on \((\Omega, \mathcal{A}, P_\mu)\) that satisfies

\[
P_\mu(X_1 \in A_1, \ldots, X_n \in A_n) = \int \mu(dx) \int \pi(\bar{x}, dx_1) \int \pi(\bar{x}, x_1, dx_2) \cdots \int \pi(\bar{x}, x_1, \ldots, x_{n-1}, dx_n)
\]

where \( \bar{x} = (\ldots, x_2, x_1, x_0) \in [0, 1]^N \). 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

\[
u_y(x) = u(\ldots, x_i, x_0, y).
\]

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

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

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

\[
\pi_x(\bar{x}, \cdot) = \pi(\bar{x}, x_1, \ldots, x_n).
\]

Let \( c_0, c_1, c_2, \ldots \) 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 \( \lambda : D \to [0, \infty] \) 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

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

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

and the function \( \lambda \) vanishes at a unique point \( F_0 \in D \).
The fact that $\lambda$ vanishes at a unique point $F_0$ is established in Example 3.2 of [5]. This is based on the result that points where $\lambda$ 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 Erdös and Rényi [6].

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

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

as $n \to \infty$ ($P_\mu$ a.s.) for any initial distribution $\mu$ on $[0,1]^N$.

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

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

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

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

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 $\lfloor x \rfloor$ denote the greatest integer less than or equal to $x$. Now

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

$$\leq \sum_{i=1}^{d} P_\mu \left( \bigcap_{j=1}^{\lfloor \exp(nR)/n \rfloor} \left\{ \|F_{(j-1)n+1,n} - G_i\| \geq \varepsilon \right\} \right)$$

$$\leq \sum_{i=1}^{d} \left( \sup_{x} P_x \{ \|F_{1,n} - G_i\| \geq \varepsilon \} \right)^{\lfloor \exp(nR)/n \rfloor}$$

$$= \sum_{i=1}^{d} \left( 1 - \inf_{x} P_x \{ \|F_{1,n} - G_i\| < \varepsilon \} \right)^{\lfloor \exp(nR)/n \rfloor}.$$

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

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

using Theorem 2.1, since the function $\lambda$ on the segment of probability measures joining $F_0$ to $G_i$ is strictly less than $\lambda(G_i)$ by the convexity of $\lambda$ and the fact
that \( \lambda \) vanishes at the point \( F_0 \). But even when \( \lambda(G_i) = 0 \), we can still get the inequality
\[
\inf_x P_x(\|F_{1,n} - G_i\| < \varepsilon) \geq e^{-n(R-\delta)}, \quad i = 1, \ldots, d,
\]
for a \( \delta > 0 \) independent of \( i \) and for \( n \) sufficiently large, since \( R > 0 \). Thus
\[
P_\mu(A_n) \leq \sum_{i=1}^{d} (1 - e^{-n(R-\delta)})^{[\exp(nR)/n]} \leq d(1 - e^{-n(R-\delta)})^{[\exp(nR)/n]}.
\]
Therefore, \( \sum_{n=1}^{\infty} P_\mu(A_n) < \infty \), which proves the theorem. \( \Box \)

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 \( X_1, X_2, \ldots \) is an i.i.d. sequence in \([0, 1]\) with common distribution function \( F_0 \). For \( F \in \mathcal{D} \), 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 \( G \in \mathcal{D} \) satisfies \( K(G) < \infty \), then
\[
\min_{1 \leq m \leq \exp(nK(G))} \|F_{m,n} - G\| \to 0
\]
as \( n \to \infty \) (a.s.).

**Proof.** The proof is a simplified version of the previous proof. \( \Box \)

**References**