ON THE PROBABILITY OF LARGE DEVIATIONS OF THE MEAN FOR RANDOM VARIABLES IN $D[0, 1]$[1]

By J. Sethuraman[2]

Michigan State University and Stanford University

1. Introduction. Let (Ω, S, P) be a probability measure space. Let $X_1(\omega)$, $X_2(\omega)$, $\cdots$ be a sequence of B-measurable random variables in $\mathfrak{X}$ which are independently and identically distributed with common distribution $\mu(\cdot)$. Here $\mathfrak{X}$ is a separable complete metric space and B is the class of all Borel subsets. Let $\mu(n, \omega, \cdot)$ be the empirical probability measure of $X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$, namely, the probability measure that assigns masses $1/n$ at each of the points $X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$.

Let $f(x)$ be any measurable real-valued function on $\mathfrak{X}$ with $\int \exp(sf(x))\mu(dx) < \infty$ for all s. It has been shown by Cramér [3], Chernoff [2], Bahadur and Ranga Rao [1], etc. that

$$(1) \quad (1/n) \log P\{\omega: |\int f(x)\mu(n, \omega, dx) - \int f(x)\mu(dx)| \geqq \epsilon\} \to \log \rho(f, \epsilon)$$

where, if $E(g) = \int g(x)\mu(dx)$ for any function g,

$$(2) \quad \begin{aligned}\rho(f, \epsilon) = \max \{&\inf_{s \geqq 0} \exp[-s\epsilon - sE(f)]E(e^{sf}),\\ &\inf_{s \leqq 0} \exp[s\epsilon - sE(f)]E(e^{sf})\}.\end{aligned}$$

Now let $\mathfrak{X}$ be a separable Banach space and let

(i) $\int \exp(s\|x\|)\mu(dx) < \infty$ for all s and

(ii) $\int x^*(x)\mu(dx) = 0$ for each continuous linear functional x^* on $\mathfrak{X}$. Sethuraman [6] (Theorem 7) has shown that

$$(3) \quad 1/n \log P\{\omega: \|(1/n)(X_1(\omega) + \cdots + X_n(\omega))\| \geqq \epsilon\} \to \log \rho(\mathfrak{X}_1^*, \epsilon)$$

where $\mathfrak{X}_1^* = \{x^*: \|x^*\| = 1\}$ and for any collection, $\mathfrak{F}$, of functions $\rho(\mathfrak{F}, \epsilon) = \sup_{f \in \mathfrak{F}} \rho(f, \epsilon)$.

Now, let $\mathfrak{X}$ be the space $D[0, 1]$ of all real valued functions $x(t)$ on $[0, 1]$ with the properties

(i) $x(t - 0)$ and $x(t + 0)$ exist for $0 < t < 1$ and $x(t) = x(t + 0)$

(ii) $x(t)$ is continuous at $t = 0$ and $t = 1$.

We endow this space with the J_1-topology of Skorohod [7] and it becomes a separable complete metric space. (See Section 2 for more details.) Let $\|x\| = \sup_{0 \leqq t \leqq 1} |x(t)|$. The main result of this note, proved in Section 3, can now be stated.

Received 12 March 1964; revised 24 July 1964.

[1] Reproduction in whole or in part is permitted for any purpose of the United States Government. Prepared with the support of the Army Research Office, Grant No. DA-ARO (D)-31-124-G485.

[2] On leave from the Indian Statistical Institute, Calcutta.

THEOREM 1. *Let* $\int \exp(s\|x\|)\mu(dx) < \infty$ *for all* s. *Then*

$$(4)\quad (1/n)\log P\{\omega: \sup_{0\le t\le 1}|\int x(t)\mu(n,\omega,dx) - \int x(t)\mu(dx)| \ge \epsilon\} \to \log \rho(\mathcal{G},\epsilon)$$

where $\mathcal{G} = \{g: g = g_t, 0 \le t \le 1 \text{ or } g = g_{\underline{t}}, 0 < t < 1\}$ *where*

$$g_t(x) = x(t), \qquad 0 \le t \le 1$$

$$g_{\underline{t}}(x) = x(t-0), \qquad 0 < t < 1.$$

We remark that (1), (3), and (4) are strengthenings of the corresponding results on the strong law of large numbers (SLLN) under the added assumption of the existence of the moment generating function. The SLLN for real-valued and Banach-valued random variables are well known and need no references. The SLLN for $D[0, 1]$-valued random variables was given by Ranga Rao [5].

2. Preliminaries. The J_1-topology of Skorohod on $\mathfrak{X} = D[0, 1]$ is defined as follows:

A sequence x_n converges to x if there is a sequence $\{\lambda_n\}$, of 1-1, continuous transformations from [0, 1] to [0, 1] such that

$$\sup_{0\le t\le 1}[|x_n(\lambda_n(t)) - x(t)| + |\lambda_n(t) - t|] \to 0.$$

Kolmogorov [4] and Skorohod [7] have shown the following:

(i) With this topology $\mathfrak{X}$ becomes a separable complete metric space.

(ii) A set K in $\mathfrak{X}$ is compact if and only if for each $\theta > 0$ there is a $\delta = \delta(K, \theta)$ such that

$$(5)\quad [\sup_{0<t<\delta}|x(t) - x(0)| + \sup_{1-\delta<t<1}|x(t) - x(1)| + \sup_{a<t_1<t_2<t_3<a+\delta}\{\min(|x(t_1) - x(t_2)|, |x(t_2) - x(t_3)|)\}] < \theta$$

and $|x(t)| < M$, $t \in [0, 1]$ for all x in K.

This implies that if $0 = t_0 < t_1 < \cdots < t_k = 1$ is a partition, $\mathfrak{J}$, of [0, 1] with $\max_i (t_{i+1} - t_i) < \delta(K, \theta)$ then either $|x(t) - x(t_i)| < \theta$ or $|x(t) - x(t_{i+1})| < \theta$ whenever $t_i \le t < t_{i+1}$ and for all x in K. In any case for each t in [0, 1] there is an integer i such that

$$(6)\quad |x(t) - x(t_i)| \le |x(t_i) - x(t_{i+1})| + \theta$$

for all x in K.

The plan of the proof of Theorem 1 is as follows. From (6) we note that for any compact set $K \subset \mathfrak{X}$ and $\theta > 0$, there is a $\delta(K, \theta)$ such that if $\mathfrak{J} = (t_0, t_1, \cdots, t_k)$ is a partition of [0, 1] such that $\max_i (t_{i+1} - t_i) < \delta(K, \theta)$ then

$$\begin{aligned}\sup_t|\int x(t)\mu(n,\omega,dx) - \int x(t)\mu(dx)| &\le \max_i|\int x(t_i)\mu(n,\omega,dx) - \int x(t_i)\mu(dx)| \\ &+ \max_i \int |x(t_i) - x(t_{i+1})|(\mu(n,\omega,dx) + \mu(dx)) \\ &+ 4\int_{\mathfrak{X}-K}\|x\|(\mu(n,\omega,dx) + \mu(dx)) + 2\theta.\end{aligned}$$

The purpose of Lemmas 3 and 4 is to show that $\mathfrak{I}$ can be chosen more carefully so that the contribution of the second term on the right hand side is small. Lemma 2 shows that K can be chosen so that the contribution of the third term can be neglected. Lemma 6 is to show that the arbitrary but positive θ can be made to tend to 0 without affecting the results. A few elementary considerations outlined in Lemma 5 help to evaluate the contribution of the first term precisely. This would complete the proof.

LEMMA 1. *Let ν_1 and ν_2 be two probability measures on $(\mathfrak{X}, B)$ with $\int \|x\|\nu_1(dx)$ and $\int \|x\|\nu_2(dx)$ finite. For any compact set $K \subset \mathfrak{X}$ and $\theta > 0$ there is a number $\delta(K, \theta)$ such that if $\mathfrak{I} = (t_0, t_1, \cdots, t_k)$ is a partition of* $[0, 1]$ *with* $\max_i (t_{i+1} - t_i) < \delta(K, \theta)$, *then*

$$\begin{aligned}\sup_{0 \leq t \leq 1} \left|\int x(t)\nu_1(dx) - \int x(t)\nu_2(dx)\right| &\leq \max_i \left|\int x(t_i)\nu_1(dx) - \int x(t_i)\nu_2(dx)\right| \\ &+ \max_i \int |x(t_i) - x(t_{i+1})|(\nu_1(dx) + \nu_2(dx)) \\ &+ 4\int_{K'} \|x\|(\nu_1(dx) + \nu_2(dx)) + 2\theta\end{aligned}$$

where K' is the complement of K.

PROOF. The lemma follows immediately from (6).

LEMMA 2. *Let*

$$\begin{aligned} f_m(x) &= \|x\| \qquad \text{if } x \in K_m' \\ &= 0 \qquad \text{if } x \in K_m \end{aligned}$$

where $\{K_m\}$, $m = 1, 2, \cdots$, is an increasing sequence of compact sets whose union is a support for μ. Then

(a) $(1/n) \log P\{\omega: \int f_m(x)\mu(n, \omega, dx) \geq \gamma\} \to \log \rho^*(f_m, \gamma)$

where

$$\rho^*(f_m, \gamma) = \inf_{s \geq 0} \exp(-s\gamma)E(e^{sf_m}), \tag{7}$$

(b) $\lim_{m\to\infty} \rho^*(f_m, \gamma) = 0$ *for each* γ

and

$$\text{(c)} \qquad \lim_{m\to\infty} E(f_m) = 0.$$

PROOF. The first part of this lemma is just another version of (1). (b) follows from the fact that $E(e^{sf_m}) \to 1$ for each s. (c) is trivial.

LEMMA 3. *Given any η and $s > 0$ there exists a partition $\mathfrak{I}_\eta(s)$, of* $[0, 1]$ *such that if $\mathfrak{I} = (t_0, t_1, \cdots, t_k)$ is any subpartition of $\mathfrak{I}_\eta(s)$, then*

$$\text{(a)} \qquad \max_i \left|\int \exp[s|x(t_i) - x(t_{i+1})|]\mu(dx) - 1\right| \leq \eta$$

and

$$\text{(b)} \qquad \max_i \int |x(t_i) - x(t_{i+1})|\mu(dx) < \eta.$$

PROOF. Part (b) of this lemma has been established by Ranga Rao [5] (Lemma 2). Part (a) is proved in an analogous fashion.

The following, which is an immediate consequence of Lemma 3, is stated without proof.

LEMMA 4. *Let* $f_{u,v}(x) = |x(u) - x(v)|$. *Given* η *and* $\gamma > 0$ *there exists a partition* $\mathfrak{J}_{\eta,\gamma}$ *of* $[0, 1]$ *such that if* $\mathfrak{J} = (t_0, t_1, \cdots, t_k)$ *is a subpartition of* $\mathfrak{J}_{\eta,\gamma}$, *then*

$$\text{(a)} \qquad \max_i \rho^*(f_{t_i,t_{i+1}}, \gamma) \leqq \eta$$

and

$$\text{(b)} \qquad \max_i E(f_{t_i,t_{i+1}}) \leqq \eta$$

where $\rho^*(f, \gamma)$ *is as defined in* (7) *in Lemma* 2.

The following lemma is trivial.

LEMMA 5. *Let* $\{X_{n,i}\}$, $i = 1, 2, \cdots, k$ *be* k *sequences of random variables with*

$$\lim_n (1/n) \log P\{|X_{n,i}| \geqq \epsilon\} = \log \rho_i(\epsilon), \qquad i = 1, 2, \cdots, k.$$

Then, for any $\theta_2, \theta_3, \cdots, \theta_k > 0$ *with* $\sum_2^k \theta_i < \epsilon$

$$\text{(a)} \quad \limsup (1/n) \log P\{|X_{n,1} + X_{n,2} + \cdots + X_{n,k}| \geqq \epsilon\} \leqq \log [\max \{\rho_1(\epsilon - \textstyle\sum_2^k \theta_i), \rho_2(\theta_2), \cdots, \rho_k(\theta_k)\}]$$

and

$$\text{(b)} \quad \limsup (1/n) \log P\{\max_i |X_{n,i}| \geqq \epsilon\} \leqq \log [\max_i \rho_i(\epsilon)].$$

An important lemma that we shall need is the following:

LEMMA 6. *Let* $\mathfrak{F}$ *be a class of measurable functions on* $\mathfrak{X}$ *with the properties*

(a) *every sequence in* $\mathfrak{F}$ *has a subsequence that converges to a function* f *in* $\mathfrak{F}$ *almost everywhere* $[\mu]$,

(b) *there is a function* g *such that* $|f(x)| \leqq g(x)$ *for all* f *in* $\mathfrak{F}$ *and* $E(\exp(sg)) < \infty$ *for all* s. *Then* $\rho(\mathfrak{F}, \epsilon)$ *is continuous from the left at each* $\epsilon > 0$.

PROOF. This assertion is essentially Lemma 3 of Sethuraman [6].

3. Proof of Theorem 1. Let $\{\theta_m\}$ and $\{\eta_m\}$ be two sequences of positive numbers tending to zero; γ_1, γ_2 be two positive numbers with $\gamma = \gamma_1 + \gamma_2 < \epsilon$ and $\{K_m\}$ be an increasing sequence of compact subsets of $\mathfrak{X}$ whose union is the support of μ. Let $\mathfrak{J}_m = (t_{m0}, t_{m1}, \cdots, t_{mk_m})$ be a subpartition of the partition $\mathfrak{J}_{\eta_m,\gamma_2}$ of $[0, 1]$ with $\max_i (t_{m(i+1)} - t_{mi}) < \delta(K_m, \theta_m)$ where $\mathfrak{J}_{\eta_m,\gamma_2}$ and $\delta(K_m, \theta_m)$ are as defined in Lemmas 4 and 1, respectively. We then have from Lemma 1

$$\begin{aligned} Z_n(\omega) &= \sup_{0 \leqq t \leqq 1} \left|\int x(t)\mu(n, \omega, dx) - \int x(t)\mu(dx)\right| \\ &\leqq \max_i \left|\int x(t_{mi})\mu(n, \omega, dx) - \int x(t_{mi})\mu(dx)\right| \\ &\quad + \max_i \int |x(t_{m(i+1)}) - x(t_{m,i})|(\mu(n, \omega, dx) + \mu(dx)) \\ &\quad + 4\int_{K_m'} \|x\|(\mu(n, \omega, dx) + \mu(dx)) + 2\theta_m \\ &= Z_{n,1,m}(\omega) + Z_{n,2,m}(\omega) + Z_{n,3,m}(\omega) + 2\theta_m, \text{ (say)}. \end{aligned} \tag{8}$$

Now,

$$\limsup_n (1/n) \log P\{\omega: Z_{n,3,m}(\omega) \geqq \gamma_3\} \leqq \log \rho_{3,m}(\gamma_3)$$

where $\lim_m \rho_{3,m}(\gamma_3) = 0$ according to Lemma 2,

$$\limsup_n (1/n) \log P\{\omega: Z_{n,2,m}(\omega) \geqq \gamma_2\} \leqq \log \rho_{2,m}(\gamma_2)$$

where $\lim_m \rho_{2,m}(\gamma_2) = 0$ according to Lemmas 4 and 5, and

$$\limsup_n (1/n) \log P\{\omega: Z_{n,1,m}(\omega) \geqq \theta\} = \log \rho_{1,m}(\theta)$$

where

$$\rho_{1,m}(\theta) \leqq \max_i \rho(g_{i_{m_i}}, \theta) \leqq \rho(\mathcal{G}, \theta)$$

according to (1) and Lemma 5. Using Lemma 5 once again

$$\limsup_n (1/n) \log P\{\omega: Z_n(\omega) \geqq \epsilon\} \leqq \log \rho(\mathcal{G}, \epsilon - \gamma).$$

Using Lemma 6, we have

$$\limsup_n (1/n) \log P\{\omega: Z_n(\omega) \geqq \epsilon\} \leqq \log \rho(\mathcal{G}, \epsilon). \tag{9}$$

And, from the obvious inequality

$$Z_n(\omega) \geqq |\int g(x)\mu(n, \omega, dx) - \int g(x)\mu(dx)|$$

for each g in $\mathcal{G}$, we also have

$$\liminf_n (1/n) \log P\{\omega: Z_n(\omega) \geqq \epsilon\} \geqq \log \rho(\mathcal{G}, \epsilon). \tag{10}$$

Theorem 1 now follows from (9) and (10).

4. Some remarks.

(i) Let $F(t)$ be any distribution function on $[0, 1]$. Let $Y_1(\omega), Y_2(\omega), \cdots$ be a sequence of independent random variables with the common distribution function $F(t)$. Define

$$X_i(t, \omega) = 0 \qquad t < Y_i(\omega)$$
$$= 1 \qquad t \geqq Y_i(\omega).$$

It is easy to show that $X_i(\cdot, \omega)$ is a random variable on $(\mathfrak{X}, B)$. We note that

$$\sup_{0 \leqq t \leqq 1} |\int x(t)\mu(n, \omega, dx) - \int x(t)\mu(dx)| = \sup_{0 \leqq t \leqq 1} |F_n(t) - F(t)|$$

where $F_n(t)$ is the empirical distribution function of $Y_1(\omega), \cdots, Y_n(\omega)$. We therefore deduce from Theorem 1, that

$$(1/n) \log P\{|F_n(t) - F(t)| \geqq \epsilon\} \to \log \rho(\mathcal{G}, \epsilon),$$

a result previously shown in a more general form in Sethuraman ([6], Theorem 3).

(ii) Let $\mathfrak{X} = C[0, 1]$ the space of continuous functions on $[0, 1]$. Both (3) and (4) can be applied to $\mathfrak{X}$ since it is a Banach space and also a subset of $D[0, 1]$. The limits of

$$(1/n) \log P\{\omega: \sup_{0 \leq t \leq 1} |\int x(t)\mu(n, \omega, dx) - \int x(t)\mu(dx)| \geq \epsilon\}$$

so obtained would be log $\rho(\mathfrak{X}_1^*, \epsilon)$ and $\rho(\mathcal{G}, \epsilon)$ from (3) and (4), respectively, which of course must be equal. Of these, the second would be easier to compute since it involves taking a supremum over a smaller set than for the first.

REFERENCES

[1] Bahadur, R. R. and Ranga Rao, R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[2] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[3] Cramér, H. (1938). Sur un noveau théorème limite de la théorie des probabilités. *Actualités Sci. Ind.* **736**.

[4] Kolmogorov, A. N. (1956). On Skorohod convergence. *Teor. Veroyatnost. i Primenen.* **1** 239–247 (also *Theor. Prob. Appl.* **1** 215–222).

[5] Ranga Rao, R. (1963). The law of large numbers for $D[0, 1]$ valued random variables. *Teor. Veroyatnost. i Primenen.* **8** 75–80 (also *Theor. Prob. Appl.* **8** 70–74).

[6] Sethuraman, J. (1964). On the probability of large deviations of families of sample means. *Ann. Math. Statist.* **35** 1304–1316.

[7] Skorohod, A. V. (1956). Limit theorems for stochastic process. *Teor. Veroyatnost. i Primenen.* **1** 289–319 (also *Theor. Prob. Appl.* **1** 261–290).