UNIVERSITY OF NORTH CAROLINA

Department of Statistics

Chapel Hill, N. C.

ON THE PROBABILITY OF LARGE DEVIATIONS OF FAMILIES OF SAMPLE MEANS

by

J. Sethuraman

June 1963

Grant No. AF - AFOSR - 62 - 169

Results concerning the exponential convergence to zero of the probability of large deviations of the sample mean have already been obtained in the literature, for instance, in Cramer [5], Chernoff [4], Bahadur and Rao [1], etc. We establish similar results about the exponential convergence to zero of the probability of large deviations of the sample distribution function, and more generally, of families of sample means.

This research was supported by the Mathematics Division of the Air Force Office of Scientific Research.

Institute of Statistics
Mimeo. Series No. 368

ON THE PROBABILITY OF LARGE DEVIATIONS OF

FAMILIES OF SAMPLE MEANS[1]

by

J. Sethuraman
University of North Carolina

1. General Summary:

Results concerning the exponential convergence to zero of the probability of large deviations of the sample mean have already been obtained in the literature for instance, in Cramer [5], Chernoff [4], Bahadur and Rao [1], etc. We establish similar results about the exponential convergence to zero of the probability of large deviations of the sample distribution function, and more generally, of families of sample means.

2. Introduction and Summary:

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

Let $f(x)$ be any real valued measurable function on $\mathcal{X}$ with $\int f(x) \quad \mu(dx) < \infty$. Then the strong law of large numbers due to Kolmogorov states that

$$(1) \quad P\left\{\omega: \frac{f(X_1(w)) + \ldots + f(X_n(\omega))}{n} - f(x)\,\mu(dx) \dashrightarrow 0\right\} = 1 .$$

[1]This research was supported by the Mathematics Division of the Air Force Office of Scientific Research.

Again, when $\int e^{tf(x)}\,\mu(dx) < \infty$ for all t, the above result has been improved upon as follows, by Cramer (1938), Chernoff (1952), Bahadur and Rao (1960), etc.

$$(2)\qquad \frac{1}{n}\log P\left\{\omega\,;\ \frac{f(X_1(\omega)) + \dots + f(X_n(\omega))}{n} - \int f(x)\,\mu(dx) \geq \epsilon\right\} \longrightarrow \log\ \rho(f,\epsilon)$$

where $0 \leq \rho(f,\epsilon) < 1$ and will be defined in (14). We interpret this result in words by saying that the probability of large deviations of the sample mean tends to zero exponentially.

Let $\mathcal{X} = R^k$, be the Euclidean space of k dimensions. Let $F(\underset{\sim}{x})$ and $F(n,\omega,\underset{\sim}{x})$ represent the distribution functions corresponding to $\mu(.)$ and $\mu(n,\omega,.)$ respectively. The Glivenko-Cantelli theorem states that

$$(3)\qquad P\left\{\omega:\ \sup_{\underset{\sim}{x}} \left|F(n,\omega,\underset{\sim}{x}) - F(\underset{\sim}{x})\right| \longrightarrow 0\right\} = 1$$

We now ask the question whether the probability of large deviations of the sample distribution function will tend to zero exponentially. We answer this question by the following theorem .

Theorem 1. For $\epsilon > 0$,

$$(4)\qquad \frac{1}{n}\ \log\ P\left\{\omega:\ \sup_{\underset{\sim}{x}} \left|F(n,\omega,\underset{\sim}{x}) - F(\underset{\sim}{x})\right| \geq \epsilon\right\} \longrightarrow \log\ \rho^*(f,\epsilon)$$

where $0 \leq \rho^*(F,\epsilon) < 1$ and $\rho^*(F,\epsilon)$ will be defined in (26).

Now we can interpret $F(n,\omega,\underset{\sim}{x})$ as the sample mean of a certain real valued function or as the sample probability measure of a certain subset of $\mathcal{X}$. Corresponding to each of these interpretations we can generalise the above theorem as follows. The proofs of all these theorems are given in section 4.

Theorem 2.

Let $\mathcal{F}$ be an equicontinuous[1] class of continuous functions from $\mathcal{X}$ to the real line. Let $g(x)$ be a continuous function such that $|f(x)| \leq g(x)$ for

[1] For a definition see section 3.

each f in $\mathcal{F}$ and such that $\int e^{tg(x)}\,\mu(dx) < \infty$ for all t. Then for $\epsilon > 0$,

$$\frac{1}{n}\log P\left\{\omega\colon \sup_{f\in\mathcal{F}}\left|\int f(x)\,\mu(n,\omega,dx) - \int f(x)\,\mu(dx)\right| \geq \epsilon\right.$$

$$\dashrightarrow \rho(\mathcal{F}, \epsilon)$$

where $\rho(\mathcal{F}, \epsilon)$ will be defined in (15).

Theorem 3.

Let $\mathcal{F}$ be a class of continuous functions from $\mathcal{X}$ into R^k that is compact under the uniform convergence on compacta[1] (u.c.c) topology. Let μf^{-1} be a non-atomic for each each f in $\mathcal{F}$. For each $\underset{\sim}{a}$ in R^k and f in $\mathcal{F}$ let $A(f,\underset{\sim}{a})$ be the set

$$\left\{x\colon\ f_1(x) \leq a_1\ ,\ \ldots\ f_k(x) \leq a_k\right\}.$$

Then for $\epsilon > 0$,

$$\frac{1}{n}\ \log\ P\left\{\omega\colon \sup_{f\in\mathcal{F}}\ \sup_{\underset{\sim}{a}}\ \left|\ \mu(n,\omega,A(f,\hat{a})) - \mu(A(f,\hat{a}))\right| > \epsilon\right\} \longrightarrow \log \rho^*(\epsilon)$$

where $0 < \rho^*(\epsilon) < 1$ and $\rho^*(\epsilon)$ will be defined in (26).

These results are allied to those of Sanov [11] and could be possibly deduced from them.. However elementary and straightforward proofs are presented here.

Section 3 deals with the necessary preliminaries and some well known lemmas. Section 4 gives the proofs of theorems 1,2 and 3 and section 5 contains some of their applications.

A result related to theorem 1, giving some upper bounds to the probability on the left hand side of (4) may be found in Kiefer and Wolfowitz [9]. Weaker

[1]For a definition see section 3.

forms of theorems 2,3,4 and 5 have been obtained by Rao [10]. A result extending theorem 4 but weaker in other respects may be found in Wolfowitz [14]. Theorem 6 extends the results of Blum [3].

3. Preliminaries:

A sequence of measures $\{\nu_n\}$ on $(\mathcal{X}, \mathcal{B})$ is said to converge weakly to ν, $\nu_n \Rightarrow \nu$ in symbols, if for every bounded continuous function h on $\mathcal{X}$,

$$\int h(x)\ \nu_n(dx) \longrightarrow \int h(x)\ \nu(dx) \ .$$

In such a case, for any $\delta > 0$, there is a compact set K in $\mathcal{X}$ such that

$$\nu_n(K) \geq 1 - \delta \quad \text{for} \quad n = 1,2,\ldots \ .$$

and

$\nu(K) \geq 1 - \delta$. This result may be found in Billingsley [2].

A family of functions $\mathcal{F}$ is said to be equicontinuous, if for every $\delta > 0$, there is an r such that

$$|f(x) - f(x')| < \delta \quad \text{for all } f \text{ in } \mathcal{F} \text{ whenever } d(x,x') < r$$

where d is the metric on $\mathcal{X}$. A sequence of functions $\{f_n\}$ on $\mathcal{X}$ is said to converge uniformly on compacta (u.c.c) to a functions f, if $f_n(x)$ converges to $f(x)$ for each x and the convergence is uniform on each compact set in $\mathcal{X}$. A theorem due to Ascoli, for instance see Kelley [8], Chap. 7, states that an equicontinuous family of functions $\mathcal{F}$ bounded by a function g is conditionally compact under the u.c.c. topology. In this case for any $\delta > 0$ and compact set K in $\mathcal{X}$, we can find a finite collection of functions $\{f_1, \ldots f_m\}$ in $\mathcal{F}$ such that for any function f in $\mathcal{F}$, there is an index i such that

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

Let

(5) $$\Omega_0 = \left\{ \omega : \mu(n, \omega, .) \Rightarrow \mu(.) \right\} .$$

An interesting result due to Varadarajan [12] which is a generalization of the Glivenko-Cantelli theorem states that

(6) $$P(\Omega_0) = 1 .$$

A conclusion that we draw from this is that for each $\delta > 0$ and ω in Ω_0, there is a compact set K_ω in $\mathcal{X}$ such that

(7) $$\begin{cases} \mu(n, \omega, K_\omega) \geq 1 - \delta & n = 1,2,\dots \\ \text{and} \quad \mu(K_\omega) \geq 1 - \delta . & \end{cases}$$

Let $g(x)$ be any measurable function with $\int g(x) \; \mu(dx) < \infty$.

Let

(8) $$\Omega_1 = \left\{ \omega : \int g(x) \; \mu(n, \omega, dx) \dashrightarrow \int g(x) \; \mu(dx) \right\} .$$

The relation (1) can now be rewritten as

(9) $$P(\Omega_1) = 1.$$

Another conclusion that we can draw from this is the following. For each $\delta > 0$ and $\omega \in \Omega_1$, there is a compact set K_ω such that

(10) $$\int_{\mathcal{X} - K_\omega} g(x) \; \mu(n, \omega, dx) < \delta , \quad n = 1, \dots$$

$$\text{and} \quad \int_{\mathcal{X} - K_\omega} g(x) \; \mu(dx) < \delta .$$

For any function f we define $E(f)$ by the relation

(11) $$E(f) = \int f(x) \; \mu(dx) .$$

Let $f(x)$ be a function for which $E(e^{tf}) < \infty$ for all t. Define for $\epsilon > 0$,

$$(12)\quad \begin{cases} H_1(f,t,\epsilon) = E(\exp(tf - tE(f) - t\epsilon)) \\ \text{and} \\ H_2(f,t,\epsilon) = E(\exp(tf - tE(f) + t\epsilon)) \,. \end{cases}$$

Define

$$(13)\quad \begin{cases} \rho_1(f,\epsilon) = \inf\limits_{t \geq 0} H_1(f,t,\epsilon) \\ \text{and} \\ \rho_2(f,\epsilon) = \inf\limits_{t \leq 0} H_2(f,t,\epsilon) \end{cases}$$

Define

$$(14)\qquad \rho(f,\epsilon) = \max(\rho_1(f,\epsilon), \rho_2(f,\epsilon)) \,.$$

For any class of functions $\mathcal{F}$ such that $E(e^{tf}) < \infty$ for all t whenever f is in $\mathcal{F}$, define

$$(15)\qquad \rho(\mathcal{F},\epsilon) = \sup_{f \in \mathcal{F}} \rho(f,\epsilon) \,.$$

The following are well known and are given here only for completeness.

Lemma 1.

If f is any function such that $E(e^{tf}) < \infty$ for all $t \geq 0$, and $\epsilon > 0$, then one of the following two is true.

(i) $P\{f - E(f) > \epsilon\} > 0$ and

$$(16)\qquad \min_{0 \leq t \leq T} H_1(f,t,\epsilon) = \rho_1(f,t,\epsilon)$$

where T is finite

(ii) $P\{f - E(f) \leq \epsilon\} = 1$ and

$$\rho_1(f,t,\epsilon) = P\left\{ f - E(f) = \epsilon \right\}$$

and given any $\Theta > 0$, there is a T such that

(17) $$\min_{0 \leq t \leq T} H_1(f,t,\epsilon) \leq \rho_1(f,\epsilon) + \Theta .$$

Proof: (i) may be found in Wald [13] pp 158. (ii) follows from similar analysis. That either (i) or (ii) must be true, is obvious.

Lemma 2. If f is any function for which $E(e^{tf}) < \infty$ for all t, and $\epsilon > 0$ then given any $\Theta > 0$, there is a T such that

(18) $$\min_{0 \leq t \leq T} H_1(f,t,\epsilon) < \rho_1(f,t,\epsilon) + \Theta$$

and

$$\min_{-T \leq t \leq 0} H_2(t,t,\epsilon) \leq \rho_2(f,t,\epsilon) + \Theta .$$

Proof: This follows from lemma 1 and a similar result for $H_2(f,t,\epsilon)$.

Lemma 3. If $\mathcal{F}$ is a class of continuous functions that is conditionally under the w.c.c. topology then the function $\rho(\mathcal{F}, \epsilon)$ is continuous from the left at each $\epsilon > 0$.

Proof:

Let $\delta > 0$. We note that

$$H_1(t, t, \epsilon-\delta) \geq H_1(f,t,\epsilon) \quad \text{for} \quad t \geq 0$$

and $$H_2(f, t, \epsilon-\delta) \geq H_2(f,t,\epsilon) \quad \text{for} \quad t \leq 0 .$$

Hence

$$\rho(f, \epsilon - \delta) \geq \rho(f,\epsilon) \quad \text{for each} \quad f.$$

Thus

$$\rho(\mathcal{F}, \epsilon-\delta) \geq \rho(\mathcal{F},\epsilon).$$

Hence

(19) $$\liminf_{\delta \to 0+} \rho(\mathcal{F}, \epsilon - \delta) \geq \rho(\mathcal{F}, \epsilon) .$$

Now, let ϵ_n be a sequence of numbers tending to ϵ. For any $\theta > 0$, we can find a sequence $\{f_n\}$ of functions in $\mathcal{F}$ such that

$$(20) \qquad \rho(f_n,\epsilon_n) \geq \rho(\mathcal{F}, \epsilon_n) - \theta \ , \ n = 1,2,\ldots \ .$$

Since $\mathcal{F}$ is compact, there is a subsequence of $\{f_n\}$, say $\{f_n\}$ itself, for simplicity, that converges to f uniformly on compact sets. Hence

$$H_1(f_n, t, \epsilon_n) \longrightarrow H_1(f,t,\epsilon)$$

uniformly on bounded intervals to t. Further

$$\rho_1(f_n, \epsilon_n) \leq \min_{0 \leq t \leq T} H_1(f_n, t, \epsilon_n)$$

and

$$\rho_2(f_n, \epsilon_n) \leq \min_{0 \leq t \leq T} H_2(f_n, t, \epsilon_n)$$

for each finite T. Again given $\theta > 0$, we can find a T to satisfy relation (18) of lemma 2. We thus obtain

$$(21) \qquad \limsup \ \rho(f_n,\epsilon_n) \leq \rho(f,\epsilon) + \theta \ ,$$

for arbitrary θ. Combining (21) and (20) with the fact that $\rho(f,\epsilon) \leq \rho(\mathcal{F},\epsilon)$ we obtain

$$(22) \qquad \limsup \ \rho(\mathcal{F}, \epsilon_n) \leq \rho(\mathcal{F}, \epsilon) \ .$$

(19) and (22) establish lemma 3.

We now require some properties of $\rho(\mathcal{F}, \epsilon)$ where $\mathcal{F}$ is restricted to certain classes of functions. Let f be the indicator function of a set and let $P\{f = 1\} = p$. Then $\rho(f,\epsilon)$ depends only on p and is written as $\rho^*(p,\epsilon)$. We can now reqrite result (2) as follows. If $X_1(\omega), X_2(\omega),\ldots$ is a sequence of independent and identical binomial random variables with mean p, then

$$(23) \qquad \frac{1}{n} \log P \left\{ \omega : \left| \frac{X_1(\omega) + \ldots + X_n(\omega)}{n} - p \right| \geq \epsilon \right\} \longrightarrow \log \rho^*(p,\epsilon).$$

We can find an explicit expression for $\rho^*(p,\epsilon)$ from (14) as follows.

$$(24) \qquad \rho^*(p,\epsilon) = \max(\rho_1^*(p,\epsilon),\ \rho_2^*(p,\epsilon))$$

where

$$(25) \qquad \rho_1^*(p,\epsilon) = \begin{cases} \left(\frac{p}{p+\epsilon}\right)^{p+\epsilon} \left(\frac{1-p}{1-p-\epsilon}\right)^{1-p-\epsilon} & \ldots\ 0 \leq p \leq 1-\epsilon \\ 0 & 1-\epsilon-p \leq 1 \end{cases}$$

and

$$\rho_2^*(p,\epsilon) = \begin{cases} \left(\frac{1-p}{1-p+\epsilon}\right)^{1-p+\epsilon} \left(\frac{p}{p-\epsilon}\right)^{p-\epsilon} & \ldots\ \epsilon \leq p \leq 1 \\ 0 & 0 \leq p < \epsilon \end{cases}.$$

Let $F(\underset{\sim}{x})$ be any distribution function in R^k. We define

$$(26) \qquad \begin{cases} \rho^*(F,\ \epsilon) = \sup\limits_{\underset{\sim}{x}} \rho^*(F(\underset{\sim}{x}),\ \epsilon) \\ \text{and} \quad \rho^*(\epsilon) = \sup\limits_{0 \leq p \leq 1} \rho^*(p,\epsilon) \end{cases}.$$

We now state a few results concerning the function $\rho^*(p,\epsilon)$.

Lemma 4. (Hoeffding)

If $\epsilon > 0$ and $0 < p < 1/2$, then

$$(27) \qquad \rho^*(p,\epsilon) \leq \exp\left[- \frac{2\,\epsilon^2}{1-2p} \log \frac{1-p}{p} \right]$$

and hence,

$$(28) \qquad \lim_{p \to 0} \rho^*(p,\epsilon) = 0 .$$

Proof: This result may be found in Hoeffding (1963, relation (2.4).)

Lemma 5.

The function $\rho^*(F,\epsilon)$ is left continuous at each $\epsilon > 0$.

Proof: The proof is elementary and proceeds through steps analogous to lemmas 1, 2 and 3 and will therefore be omitted.

We now proceed to a few well-known lemmas which are connected with uniform convergence of distributions.

Lemma 6.

If $\{f_n\}$ is a sequence of continuous functions converging to a function f uniformly on compact subsets then

$$\mu f_n^{-1} \Rightarrow \mu f^{-1}$$

and, moreover, if μf^{-1} is non-atomic, then

$$(29) \qquad \sup_a \left| \mu(A(f_n,a)) - \mu(A(f,a)) \right| \longrightarrow 0$$

where $A(f,a)$ is the set

$$(30) \qquad \left\{ x:\ f(x) \leq a \right\}.$$

Proof: The first part of the lemma is well known. The second part is equally well known and is usually referred to as Polya's theorem.

Lemma 7. Let $\mathcal{F}$ be a class of continuous functions that is compact under the u.c.c. topology and μf^{-1} is non-atomic for each f in $\mathcal{F}$. Then as $\delta \longrightarrow 0$

$$(31) \qquad \sup_{f \in \mathcal{F}} \ \sup_{a \in R^1} \left| \mu(A(f+\delta,a)) - \mu(A(f,a)) \right| \longrightarrow 0$$

where $f + \delta$ is the function $f(x) + \delta$ and $A(f,a)$ is as defined in (30).

Proof: The lemma is easily proved by contradiction using lemma 6.

Lemma 8. Let $\mathcal{X} = R^k$ and $\mathcal{L}$ be the class of linear functions from $\mathcal{X}$ into R' of norm unity. Then as $\delta \to 0$

$$(32) \qquad \sup_{L \in \mathcal{L}} \sup_{a} \left| \mu(A(L+\delta, a)) - \mu(A(L,a)) \right| \to 0 .$$

Proof: This is immediate from lemma 7 since $\mathcal{L}$ is compact under the u.c.c. topology.

Whenever $\mathcal{X} = R^k$ and L in $\mathcal{L}$ we will sometimes call $A(L,a)$ a half space. $\mathcal{H}_m$ will denote the class of all sets that are formed by the intersection of m half-spaces. $|\underset{\sim}{x}|$ represents the norm of $\underset{\sim}{x}$ in the Euclidean space. $S(\underset{\sim}{a}, r)$ is the closed sphere $\{\underset{\sim}{x}: |\underset{\sim}{x}-\underset{\sim}{a}| \leq r\}$. For any subset D in $\mathcal{X}$, $S(D, r) = \{\underset{\sim}{x} : \underset{\sim}{y} \in D, |\underset{\sim}{x}-\underset{\sim}{y}| \leq r\}$.

We will also need some properties of convex sets in R^k. For the general properties of convex sets, the treatise of Eggleston (1958) is an excellent reference, but Section 4.1 of a recent paper by Rao [10] covers the material adequare for our purposes and we shall reproduce that section with slight alterations.

Let C be a convex set with non-empty interior. Let $\underset{\sim}{a}$ be an inner point of C. Then the guage function, (g.f.) $H(\underset{\sim}{x})$, of C with respect to $\underset{\sim}{a}$ is defined as follows:

$$H(\underset{\sim}{x}) = \inf \left\{ \lambda: \underset{\sim}{a} + \frac{\underset{\sim}{x}-\underset{\sim}{a}}{\lambda} \in C \right\}.$$

It is clear that if $H(\underset{\sim}{x})$ is the gauge function of C, then C^o = interior of $C = \{\underset{\sim}{x} : H(\underset{\sim}{x}) < 1\}$ and $\bar{C}$ = closure of $C = \{\underset{\sim}{x} : H(\underset{\sim}{x}) \leq 1\}$ and the boundary of $C = \{\underset{\sim}{x}: H(\underset{\sim}{x}) = 1\}$. The following properties of a gauge function are well known and can be immediately deduced from its definition

(i) If $H(\underset{\sim}{x})$ is the g.f. of C with respect to the origin then (a) $H(\underset{\sim}{x}) \geq 0$, and $H(\underset{\sim}{x}) = 0$ if and only if $\underset{\sim}{x} = 0$, (b) $H(c\underset{\sim}{x}) = cH(\underset{\sim}{x})$, for all $c > 0$, and (c) $H(\underset{\sim}{x}+\underset{\sim}{y}) \leq H(\underset{\sim}{x}) + H(\underset{\sim}{y})$, for each pair $\underset{\sim}{x}, \underset{\sim}{y}$. Conversely any function $H(\underset{\sim}{x})$ with the above properties is the gauge function of the convex set $\{\underset{\sim}{x}: H(\underset{\sim}{x}) \leq 1\}$.

(ii) Suppose that $H(\underset{\sim}{x})$ is the gauge function of C with respect to the origin and that $S(\underset{\sim}{0},r) \subset C$. Then for each $\underset{\sim}{x}$, $H(\underset{\sim}{x}) \leq \frac{|x|}{r}$.

For any convex set C, the inradius of C, denoted by $r(C)$, is defined to be

$$r(C) = \sup\left\{ r\colon S(\underset{\sim}{a}, r) \subset C, \text{ for some } \underset{\sim}{a} \text{ in } C \right\}.$$

Then it follows from definition that $r(C) = 0$ if and only if C has empty interior and that $r(C) < \infty$, if C is bounded.

(iii) If C is a convex set with non-empty interior then there is a gauge function $H(\underset{\sim}{x})$ associated with C for which

$$\left| H(\underset{\sim}{x}) - H(\underset{\sim}{y}) \right| \leq \left| \underset{\sim}{x} - \underset{\sim}{y} \right| / r(C)$$

for all x, y in R^k. We shall call this the primary gauge function of C.

(iv) If C is a convex set with inradius $< \alpha$ then there is a linear function L such that

$$C \subset \left\{ x\colon\ a \leq L(x) \leq a + 2\alpha \right\}.$$

Let K be any bounded set in R^k. $\mathcal{C}(K)$ denotes the class of all closed convex sets contained in K. for $\alpha > 0$, $\mathcal{C}(K, 1, \alpha)$ denotes the class of all closed convex sets contained in K and of inradius $\geq \alpha$. $\mathcal{C}(K, 2, \alpha) = \mathcal{C}(K) - \mathcal{C}(K, 1, \alpha)$. The final results about convex sets that we shall use in section 4 can now be stated as two lemmas.

Lemma 9.

The class $\mathcal{H}$ of primary gauge functions, H, of the elements in $\mathcal{C}(K, 1, \alpha)$ is compact under the u.c.c. topology. Further if the boundary of every convex set has μ-measure zero then μH^{-1} is non-atomic for each H in $\mathcal{H}$.

Proof: The proof is immediate from property (iii)

Lemma 10. As $\alpha \dashrightarrow 0$,

$$\sup_{C \subset \mathcal{C}(K,2,\alpha)} \mu(S(C,\alpha)) \dashrightarrow 0 \ .$$

Proof: We note from (iv) that

$$S(C,\alpha) \subset \left\{ x : a - \alpha \leq L(x) \leq a + 3\alpha \right\} .$$

The lemma follows from lemma 8 .

Again let $\mathcal{X} = R^k$. Let $\mathcal{Q}_1$ be the class of sets A each of which possesses the following property. If $\underset{\sim}{x} = (x_1, \ldots, x_k) \in A$ and $\underset{\sim}{y} = (y_1, \ldots y_k)$ is such that $y_i < x_i$, $i = 1, \ldots k$, then $\underset{\sim}{y} \in A$. Let $\mathcal{Q}_j$, $j = 2, 3, \ldots, 2^k$, be the $2^k - 1$ classes of sets which can be obtained by reversing, one at a time, the k inequalities occuring in the definition of $\mathcal{Q}_1$. Let $\mathcal{Q} = \bigcup_{j=1}^{2^k} \mathcal{Q}_j$. For two classes of sets $\mathcal{D}$ and $\mathcal{E}$ let $\mathcal{D} \otimes \mathcal{E}$ denote the class os sets $D \cap E$ where D is in $\mathcal{D}$ and E in $\mathcal{E}$. Let

$$(33) \qquad \mathcal{Q}^* = \mathcal{Q} \otimes \mathcal{Q} \otimes \ \ldots \ \otimes \mathcal{Q}$$

the 'product' containing a finite number, say m, terms.

Lemma 11. (Blum) Let μ be absolutely continuous w.r.t. the Lebesgue measure. Given any $\delta > 0$, there is a finite class of sets $\mathcal{B}(\delta)$ such that for each A in $\mathcal{Q}$ there exist two elements B_1 and B_2 in $\mathcal{B}(\delta)$ such that

$$B_1 \subset A \subset B_2 \quad \text{and} \quad \mu(B_2 - B_1) < \delta \ .$$

Proof: The proof of this lemma follows from elementary considerations of the properties of the sets in $\mathcal{Q}$ and may be found in Blum [3] .

Lemma 12.

Let μ be absolutely continuous with respect to the Lebesgue measure. Then for each $\delta > 0$, there is a class of sets $\mathcal{B}^*(\delta)$ such that for each A in $\mathcal{Q}^*$ there exist sets B_1 and B_2 in $\mathcal{B}^*(\delta)$ such that

$$B_1 \subset A \subset B_2 \quad \text{and} \quad \mu(B_2 - B_1) < \delta .$$

Proof: This follows immediately from lemma 11 and the definition of $\mathcal{Q}^*$.

4. Proofs of Theorems 1, 2 and 3:

Proof of theorem 1.

Let $G_1(x), \ldots G_k(x)$ be the right-continuous marginal distribution functiosn of $F(\underset{\sim}{x})$. For any $\delta > 0$, we can choose k sets of $(m+1)$ points

$$(x_{10}, x_{11}, \ldots\ldots x_{1m}), \quad \ldots\ldots \quad (x_{ko}, x_{k1}, \ldots\ldots x_{km})$$

corresponding to the k marginals so that

$$(34) \quad \begin{cases} -\infty = x_{io} < x_{i1} < \ldots < x_{im} = +\infty \\ \\ G(x_{is} - 0) - G(x_{is-1}) < \delta \end{cases}$$

$i = 1, \ldots k$ and $s = 1, \ldots m$.

Any point $\underset{\sim}{x} = (x_1, \ldots x_k)$ in R^k will satisfy the relations

$$(35) \qquad x_{i(s_i - 1)} < x_i \leq x_{i\, s_i} \qquad i = 1, \ldots k$$

for some integers $s_1, \ldots, s_k$, $1 \leq s_i \leq m$, $i = 1, \ldots k$. A straight-forward analysis then shows that

$$(36) \qquad \sup_{\underset{\sim}{x}} \left| F(n, \omega, \underset{\sim}{x}) - F(\underset{\sim}{x}) \right|$$

$$\leq \sup_{\underset{\sim}{x}^* \in \mathcal{X}^*} \left| F(n, \omega, \underset{\sim}{x}^*) - F(x^*) \right| + \delta$$

where $\mathcal{X}^* = \left\{ \underset{\sim}{x}^* : \underset{\sim}{x}^* = (x_{1i_1}, \ldots, x_{ki_k}) \text{ for some integers } i_1, \ldots i_k \right.$ with $\left. 0 \le i_s \le m \;, \; s = 1, \ldots k \;. \right\}$.

Also from (23) we have

$$(37) \qquad \frac{1}{n} \log P \left\{ \omega : \left| F(n, \omega, \underset{\sim}{x}^*) - F(\underset{\sim}{x}^*) \right| \ge \epsilon - \delta \right\}$$

$$\longrightarrow \log \rho^*(F(\underset{\sim}{x}^*), \epsilon - \delta) \le \log \rho^*(F, \epsilon - \delta) \,.$$

(36) and (37) yield

$$(38) \qquad \limsup \frac{1}{n} \log P\left\{ \omega : \sup_{\underset{\sim}{x}} \left| F(n, \omega, \underset{\sim}{x}) - F(\underset{\sim}{x}) \right| \ge \epsilon \right\} \le \log \rho^*(F, \epsilon - \delta).$$

Since $\delta > 0$ is arbitrary and $\rho^*(F,\epsilon)$ is left continuous from lemma 5 we have

$$(39) \qquad \limsup \frac{1}{n} \log P\left\{ \omega : \sup_{\underset{\sim}{x}} \left| F(n_1, \omega, \underset{\sim}{x}) - F(\underset{\sim}{x}) \right| \ge \epsilon \right\} \le \log \rho^*(F, \epsilon).$$

Also

$$P\left\{ \omega : \sup_{\underset{\sim}{x}} \left| F(n, \omega, \underset{\sim}{x}) - F(\underset{\sim}{x}) \right| \ge \epsilon \right\} \ge P\left\{ \omega : \left| F(n, \omega, \underset{\sim}{x}) - F(\underset{\sim}{x}) \right| \ge \epsilon \right\}$$

for any $\underset{\sim}{x}$ in R^k. Thus using (23) and noting that $\underset{\sim}{x}$ is arbitrary we obtain

$$(40) \qquad \liminf \frac{1}{n} \log P\left\{ \omega : \sup_{\underset{\sim}{x}} F(n, \omega, \underset{\sim}{x}) - F(\underset{\sim}{x}) \right| \ge \epsilon \right\} \ge \log \rho^*(F,\epsilon) \,.$$

(39) and (40) prove the theorem.

Proof of theorem 2.

Define Ω_1 as in (8). We have $P(\Omega_1) = 1$ from (9). For each f in $\mathcal{F}$, let

$$(41) \qquad \Omega(f,\epsilon) = \left\{ \omega : \omega \in \Omega_1 \text{ and } \left| \int f(x)\mu(n,\omega,dx) - \int f(x)\mu(dx) \right| \ge \epsilon \right\}.$$

It is enough to show that

$$\frac{1}{n} \log P\left\{ \bigcup_{f \in \mathcal{F}} \Omega(f, \epsilon) \right\} \longrightarrow \log \rho(\mathcal{F}, \epsilon).$$

Choose and fix a $\delta > 0$. For any ω in Ω_1, there is a compact set K_ω in $\mathcal{X}$ satisfying relation (10), namely,

$$(10) \qquad \int_{\mathcal{X} - K_\omega} f(x)\, \mu(n,\omega,dx) < \delta \quad n = 1, 2, \ldots$$

and

$$\int_{\mathcal{X} - K_\omega} f(x)\, \mu(dx) < \delta, \text{ for all } f \text{ in } \mathcal{F}.$$

Further since $\mathcal{F}$ is conditionally compact we have a finite collection $\{f_1, \ldots f_m\}$ of functions in $\mathcal{F}$ such that if f is in $\mathcal{F}$ there is an index i such that

$$(42) \qquad \sup_{x \in K_\omega} \left| f(x) - f_i(x) \right| < \delta .$$

Relations (42) and (10) yeild the inequality

$$(43) \qquad \sup_{f \in \mathcal{F}} \left| \int f(x)\, \mu(n, \omega, dx) - \int f(x)\, \mu(dx) \right|$$

$$\leq \sup_{1 \leq i \leq m} \left| \int f_i(x)\, \mu(n, \omega, dx) - \int f_i(x)\, \mu(dx) \right| + 6\delta$$

for each ω in Ω_1 .

Thus

$$(44) \qquad \bigcup_{f \in \mathcal{F}} \Omega(f,\epsilon) \subset \bigcup_{i=1}^{m} \Omega(f_i, \epsilon - 6\delta) .$$

Relation (2) can be rewritten as

$$(45) \qquad \frac{1}{n} \log P(\Omega(f,\epsilon)) \longrightarrow \log \rho(f,\epsilon) .$$

From (44) and (45) we have

$$(46) \qquad \limsup \frac{1}{n} \log P\left\{ \bigcup_{f \in \mathcal{F}} \Omega(f,\epsilon) \right\} \leq \max_{1 \leq i \leq m} \log \rho(f_i, \epsilon - 6\delta)$$

$$\leq \log \rho(\mathcal{F}, \epsilon - 6\delta) \ .$$

Since $\delta > 0$ is arbitrary, and $\rho(\mathcal{F}, \epsilon)$ is left continuous from lemma 3 we have

$$(47) \qquad \limsup \frac{1}{n} \log P\left\{ \bigcup_{f \in \mathcal{F}} \Omega(f,\epsilon) \right\} \leq \log \rho(\mathcal{F}, \epsilon) \ .$$

Again, $\bigcup_{f \in \mathcal{F}} \Omega(f,\epsilon) \supset \Omega(f,\epsilon)$ for each f in $\mathcal{F}$. Using (45), we have

$$(48) \qquad \liminf \frac{1}{n} \log P\left\{ \bigcup_{f \in \mathcal{F}} \Omega(f,\epsilon) \right\} \geq \log \rho(\mathcal{F}, \epsilon) \ .$$

The theorem now follows from (47) and (48).

Proof of theorem 3.

Define Ω_0 as in (5). From (6) we have $P(\Omega_0) = 1$. Choose and fix a $\delta > 0$. Then for each ω in Ω_0, there is a compact set K_ω in $\mathcal{X}$ satisfying relation (7), namely,

$$(7) \begin{cases} \mu(n, \omega, K_\omega) \geq 1 - \delta \ , \quad n = 1, 2, \ldots \\ \text{and} \\ \mu(K_\omega) \geq 1 - \delta \ . \end{cases}$$

Let $A(f, \underset{\sim}{a}) = \left\{ x: \ f_1(x) \leq a_1, \ldots f_k(x) \leq a_k \right\}$. Since $\mathcal{F}$ is compact under the u.c.c. topology, there is a finite collection $\left\{ f_1, \ldots f_m \right\}$ of members of $\mathcal{F}$ such that for each f in $\mathcal{F}$ there is an index i such that

$$\sup_{x \in K_\omega} \left| f(x) - f_i(x) \right| < \delta \ .$$

Using this fact and relation (7) we have

$$\mu(n,\omega, A(f_i+\delta, \underset{\sim}{a})) - \mu(A(f_i-\delta, \underset{\sim}{a})) - 4\delta$$

$$\leq \mu(n,\omega, A(f, \underset{\sim}{a})) - \mu(A(f, \underset{\sim}{a}))$$

$$\leq \mu(m,\omega, A(f_i-\delta, \underset{\sim}{a})) - \mu(A(f_i+\delta, \underset{\sim}{a})) + 4\delta$$

for each ω in Ω_0.

Now choose and fix any $\Theta > 0$. Using lemma 7 we choose δ so that

$$(49) \qquad \sup_{f\in \mathcal{F}} \sup_{\underset{\sim}{a}} \left| \mu(n,\omega, A(f, \underset{\sim}{a})) - \mu(A(f,\underset{\sim}{a}))\right|$$

$$\leq \sup_{1\leq i\leq m} \max_{1\leq j\leq 2} \sup_{\underset{\sim}{a}} \Big| \mu(n,\omega, A(f_i + \phi_j\delta, \underset{\sim}{a})) - \mu(A(f_i+\phi_j\delta, \underset{\sim}{a}))\Big| + 4\delta + 2\Theta .$$

where $\phi_1 = 1$ and $\phi_2 = -1$.

If f is any function from $\mathcal{X}$ into R^k with non-atomic induced distribution, we have from theorem 1 that,

$$(50) \qquad \frac{1}{n} \log P\left\{\omega: \sup_{\underset{\sim}{a}} \left|\mu(n,\omega,A(f,\underset{\sim}{a})) - \mu(A(f,\underset{\sim}{a}))\right| \geq \epsilon\right\} \longrightarrow \log \rho^*(\epsilon)$$

where $\rho^*(\epsilon)$ is as defined by (26).

Since $f_i + \phi_j\delta$ has non-atomic induced distribution we have

$$(51) \quad \limsup \frac{1}{n} \log P\left\{\omega: \sup_{f\in} \sup_{\underset{\sim}{a}} \mu(n,\omega,A(f,\underset{\sim}{a}))-\mu(A(f,\underset{\sim}{a})) \geq \epsilon\right\}$$

$$\leq \log \rho^*(\epsilon - 4\delta - 2\Theta) .$$

Since $\rho^*(\epsilon)$ is left continuous by lemma 5 we can replace the right hand side of (51) by $\log \rho^*(\epsilon)$.

Now

$$\sup_{f \in \mathcal{F}} \sup_{\underset{\sim}{a}} \left| \mu(n,\omega,A(f,a)) - \mu(A(f,\underset{\sim}{a})) \right|$$

$$\geq \sup_{\underset{\sim}{a}} \left| \mu(n,\omega, A(f, \underset{\sim}{a})) - \mu(A(f,\underset{\sim}{a})) \right|$$

for each f in $\mathcal{F}$. This together with relation (50) gives

$$(52) \qquad \liminf \frac{1}{n} \log P\left\{\omega: \sup_{f \in \mathcal{F}} \sup_{\underset{\sim}{a}} \left| \mu(n, ,A(f,\underset{\sim}{a})) - \mu(A(f,\underset{\sim}{a})) \right| \geq \varepsilon \right\}$$

$$\geq \log \rho^*(\epsilon) .$$

The theorem follows from (51) and (52).

5. Some Applications of the Results of section 4:

We now state and prove three further theorems which are applications of the above and are of immediate applicability to practical problems. From now on we shall assume that $\mathcal{X} = R^k$.

Theorem 4.

For each linear function L let μL^{-1} be non-atomic.

$$(53) \qquad \frac{1}{n} \log P\left\{\omega: \sup_{A \in \mathcal{H}_m} \left| \mu(n,\omega, A) - \mu(A) \right| \geq \epsilon \right\} \dashrightarrow \log \rho^*(\epsilon).$$

where $\mathcal{H}_m$ is the collection of all sets formed by the intersection of m half-spaces.

Proof: This theorem is immediate from Theorem 3 and the observation that $\mathcal{L}$ = all linear functions of norm 1 is compact under the u.c.c. topology.

Theorem 5.

Let $\mathcal{C}$ be the class of all closed convex sets. Let $\mu(\text{bd } C) = 0$ for each set $C \in \mathcal{C}$, where bd C = boundary of C. Then

$$(54) \qquad \frac{1}{n} \log P\left\{\omega : \sup_{C \subset \mathcal{C}} \left| \mu(n, \omega, C) - \mu(C) \right| \geq \epsilon \right\} \longrightarrow \log \rho^*(\epsilon) \quad .$$

Proof:

Choose and fix a $\delta > 0$. We can find a bounded closed convex set K_1 in such that

$$(55) \qquad \mu(K_1) \geq 1 - \delta \quad .$$

We divide $\mathcal{C}$ into three classes, for some α, with $1 > \alpha > 0$, which we will choose later.

$$\mathcal{C} = \mathcal{C}(K_1, 1, \alpha) \cup \mathcal{C}(K_1, 2, \alpha) \cup \mathcal{C}^*$$

where $\mathcal{C}^*$ = class of all convex sets C with $C \cap K_1'$ not empty,

and K_1' is the complement of K_1.

Let $K = S(K_1, 1)$. Then trivially

$$(56) \qquad \sup_{C \subset \mathcal{C}(K_1,1,\alpha)} \left| \mu(n, \omega, C) - \mu(C) \right| \leq \sup_{C \subset \mathcal{C}(K, 1, \alpha)} \left| \mu(n, \omega, C) - \mu(C) \right| \quad .$$

Let $C \subset \mathcal{C}(K_1, 2, \alpha)$. Then α, with $1 > \alpha > 0$, can be chosen from lemma 10 so that

$$\mu(C) < \delta \quad .$$

Thus

(57) $$-\delta \leq \mu(n,\omega, C) - \mu(C) \leq \mu(n,\omega, S(C,\alpha)) - \mu(S(S,\alpha)) + \delta .$$

Now $S(C,\alpha)$ is convex, has inradius $\geq \alpha$ and is contained in K and so is a member of $\mathcal{C}(K, 1, \alpha)$. Thus

(58) $$\sup_{C \subset (K_1,2,\alpha)} \left| \mu(n,\omega, C) - \mu(C) \right| \leq \sup_{C \subset (K,1,\alpha)} \left| \mu(n,\omega, C) - \mu(C) \right| .$$

Combining (56) and (58) we have

(59) $$\sup_{C \subset \mathcal{C}(K_1)} \left| \mu(n,\omega, C) - \mu(C) \right| \leq \sup_{C \subset (K,1,\alpha)} \left| \mu(n,\omega, C) - \mu(C) \right| + \delta .$$

Let $C \subset \mathcal{C}^*$. Then

(60) $$\begin{aligned} \mu(n,\omega, K_1 \cap C) &- \mu(K_1 \cap C) - \delta \\ &\leq \mu(n,\omega, C) - \mu(C) \\ &\leq \mu(n,\omega, K_1 \cap C) - \mu(K_1 \cap C) \\ &+ \mu(n,\omega, K_1') - \mu(K_1') + \delta . \end{aligned}$$

Here $C \cap K_1$ is convex and is in $\mathcal{C}(K)$. Thus combining (59) and (60) we have

(61) $$\sup_{C \subset \mathcal{C}} \left| \mu(n,\omega, C) - \mu(c) \right|$$

$$\leq \sup_{C \subset \mathcal{C}(K,1,\alpha)} \left| \mu(n,\omega, C) - \mu(C) \right| + \left| \mu(n,\omega, K_1') - \mu(K_1') \right| + \delta .$$

Consider $\mathcal{H}$ the class of primary gauge functions of elements of $\mathcal{C}(K,1,\alpha)$. This class is compact under the u.c.c. topology from lemma 9. Each $C \subset \mathcal{C}(K,1,\alpha)$ is the set $\left\{ \underset{\sim}{x}: H(\underset{\sim}{x}) \leq 1 \right\}$ for some $H(\underset{\sim}{x})$ in $\mathcal{H}$ and from the conditions of the theorem 5, H has non-atomic induced distribution. Thus from theorem 3 we have

$$(62) \quad \limsup \frac{1}{n} \log P\left\{\omega: \sup_{C \subset \mathcal{C}(K,1,\alpha)} \left| \mu(n,\omega,C) - \mu(C) \right| \geq \beta \right\} \leq \log \rho^*(\beta) .$$

Again, from relation (23) we have

$$(63) \quad \frac{1}{n} \log P\left\{\omega: \left|\mu(n,\omega, K_1') - \mu(K_1')\right| \geq \theta \right\} \overline{\longrightarrow} \log \rho^* (\mu(K_1'), \theta) .$$

From (61), (62) and (63), for each fixed θ, with $\epsilon - \delta > \theta > 0$, we have

$$(64) \quad \limsup \frac{1}{n} \log P\left\{\omega: \sup_{C \subset \mathcal{C}} \left| \mu(n,\omega, C) - \mu(C) \right| \geq \epsilon \right\}$$
$$\leq \log \max (\rho^*(\epsilon - \delta - \theta), \rho^*(\mu(K_1'), \theta)) .$$

From lemma 4 we know that for each $\theta > 0$, $\rho^*(p,\theta) \overline{\longrightarrow} 0$ as $p \overline{\longrightarrow} 0$. Hence for sufficiently small δ

$$\rho^*(\epsilon - \delta - \theta) > \rho^*(\mu(K_1'), \theta) .$$

Again since $\rho^*(\epsilon)$ is left continuous from lemma 5 we can replace the right hand side of (64) by $\log \rho^*(\epsilon)$. The corresponding inequality for the lim inf is trivial. Hence the theorem is proved.

Theorem 6.

Let $\mu(.)$ be absolutely continuous with respect to the Lebesgue measure. Let $\mathcal{a}^*$ be as defined in (33). Then

Proof: The proof of this theorem follows immediately by the application of lemma 12.

6.. Acknowledgements.

The author wishes to thank Professor C. R. Rao of the Indian Statistical Institute for suggesting the problem and for the thoughtful conversations, and to Professor W. Hoeffding for several helpful guidances and remarks.

7. References.

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

[2] Billingsley, P. (1956). The invariance principle for dependent random variables, Trans. Math. Soc., 83, 250-268.

[3] Blum, J. R. (1955). On the convergence of empirical distribution functions, Ann. Math. Stat., 26, 527-529.

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

[5] Cramer, H. (1938). Sur un nouveau théorème limite de la théorie des probabilités, Actualités Scientifiques et Industrielles, 736.

[6] Eggleston, H. G. (1958). Convexity, Cambridge tracts in mathematics and mathematical physics, No. 47. Cambridge Univ. Press.

[7] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, Jour. Amer. Stat. Assoc., 58, 13-30.

[8] Kelley, J. L. (1955). General Topology, Van Nostrand, New York.

[9] Kiefer, J. and Wolfowitz, J. (1958). On the deviations of the empiric distribution functions of vector chance variables, Trans. Amer. Math. Soc., 87, 173-186.

[10] Rao, R. Ranga (1962). Relations between weak and uniform convergence of measures with applications, <u>Ann. Math. Stat.</u>, <u>33</u>, 659-680.

[11] Sanov, I. N. (1957). On the probability of large deviations, <u>Mat. Sbornik</u>, <u>42</u> (<u>84</u>), 11-44 (in Russian). Also (1961) <u>Selected Translations in Math. Stat. and Prob.</u>, <u>1</u>, 213-244.

[12] Varadarajan, V. S. V. (1958). On the convergence of probability distributions, <u>Sankhyā</u> , <u>19</u>, 23-26.

[13] Wald, A. (1947). <u>Sequential Analysis</u>, John Wiley Press.

[14] Wolfowitz, J. (1960). Convergence of the empiric distribution function on half-spaces, <u>Contributions to Prob. and Stat.</u>, <u>Essays in Honor of Hotelling</u>, Stanford Univ. Press, 504-507.