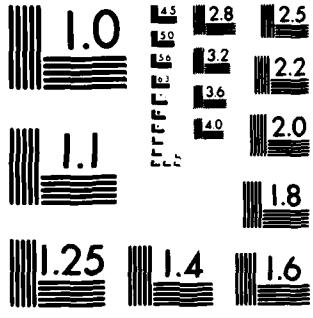
AD-A171 452 THE LARGE DEVIATION PRINCIPLE FOR THE SAMPLE AVERAGE PROCESS AND FUNCTION..(U) FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS J LYNCH ET AL. JUN 86 1/1

UNCLASSIFIED FSU-TR-M-730 ARO-20063.5-MA F/G 12/1 NL

END
DATE FILMED
10-86
DTIC



Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

(2)

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY: Unclassified	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): USARO D-89	5. MONITORING ORGANIZATION REPORT NUMBER(S): ARO 20063.5-MA; ARO 23448.2-MA; ARO 23699.2-MA
6a. NAME OF PERFORMING ORGANIZATION: Florida State University; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: U.S. Army Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code): Department of Statistics, Tallahassee, FL 32303	7b. ADDRESS (City, State and ZIP Code): Post Office Box 12211, Research Triangle Park, NC 27709
8a. NAME OF FUNDING/SPONSORING ORGANIZATION; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: DAAG29-84-K-0007; DAAL03-86-K-0094; DAAL03-86-K-0010
8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS. — PROGRAM ELEMENT NO.; PROJECT NO.; TASK NO.; WORK UNIT NO.

11. TITLE (Include Security Classification): The Large Deviation Principle for the Sample Average Process and Functional Erdös-Rényi Laws

12. PERSONAL AUTHOR(S): James Lynch and Jayaram Sethuraman

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM _____ TO _____	June 1986	16

16. SUPPLEMENTARY NOTATION: The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES		
FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Key Words: Large deviations, rate function, functional laws

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The large deviation principle is established for the sample average process. The principle is then used to obtain functional counterparts of the Erdős-Rényi type laws of Erdős and Rényi (1970) and Shepp (1964).

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
J. Sethuraman	(904) 644-3218	

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

SECURITY CLASSIFICATION OF THIS PAGE

ARO 20063.5-MA

ARO 23448.2-MA

ARO 23699.2-MA

The Large Deviation Principle for the Sample Average Process and Functional Erdös-Rényi Laws

by

James Lynch[1] and Jayaram Sethuraman[2]

June, 1986

FSU Technical Report No. M-730
USARO Technical Report No. D-99

[1]Department of Statistics, University of South Carolina and Pennsylvania State University. Research supported by the U.S. Army Research Office Nos. DAAG29-84-K-0007 and DAAL03-86-K-0010.

[2]Department of Statistics, Florida State University. Research supported by the U.S. Army Research Office Grant No. DAAL03-86-K-0094.

Key Words: Large deviations, rate function, functional laws.

AMS 1980 Subject Classification Numbers: Primary 60F10, 60F17
Secondary 60B05, 60F15, 60G50

The Large Deviation Principle for the Sample Average Process
and Functional Erdös-Rényi Laws

Abstract

The large deviation principle is established for the sample average process. The principle is then used to obtain functional counterparts of the Erdös-Rényi type laws of Erdös and Rényi (1970) and Shepp (1964).

1. Introduction. Let $X_1, X_2, \ldots$ be i.i.d. random variables with $E(X_1) = 0$ and $\sigma^2 = E(X_1^2) < \infty$. Let $P(t)$, $t \geq 0$, denote the random polygon where $P(n) = S_n = X_1 + \ldots + X_n$ with $P(0) = S_0 = 0$ and is linear between the integers.

The classical functional laws examine the asymptotic behavior of segments of the polygon originating at the origin upon various renormalizations of the space and time axes. In particular, let $D_n(t) = P(tn)/\sigma\sqrt{n}$ and $S_n(t) = P(tn)/\sqrt{\sigma n \log\log n/2}$. Let $J(a) = a^2/2$ which is the large deviation rate (Section 2) for the normal distribution. Let $I(f) = \int_0^1 J(\dot{f})dt$ if f is absolutely continuous ($\dot{f}$ denotes the Radon-Nikodym of f) and $= \infty$ if not. Let $\Gamma_a = \{f: I(f) \leq a\}$. Then, the invariance principles of Donsker (see Billingsley (1968) and Strassen (1964)) state that $\{D_n(t): 0 \leq t \leq 1\} \to \{W(t): 0 \leq t \leq 1\}$ in distribution, $W(\cdot)$ the standard Brownian Motion, and that the set of cluster points of $\{S_n(t): 0 \leq t \leq 1\}$ is Γ_a a.s.. Here $D_n(\cdot)$ and $S_n(\cdot)$ are both viewed as random functions on $C[0,1]$ with the usual norm topology.

The two functional laws proved here examine the asymptotic behavior of segments of $P(t)$ that move off toward infinity, i.e., moving averages. These laws are the functional counterparts of the Erdös-Rényi (E-R) type laws of Erdös and Rényi (1970) and of Shepp (1964) just as the Donsker and the Strassen Invariance Principles are the functional analogues of the central limit theorem and the law of the iterated logarithm. Note, though, that these E-R type functional laws are not invariance principles since the large deviation rate of X_1, which uniquely determines the distribution of X_1, can be calculated from the conclusions of these laws.

To state these results we need the following. For m and n positive integers and $a > 0$, let $\Delta_m(s) = P(m+s) - P(m)$ and let $\Delta_{m,n,a}(s) = \Delta_m(s\lfloor a^{-1}\log n\rfloor)/\lfloor a^{-1}\log n\rfloor$ where $[x]$ indicates the integer part of x. We futher assume that the m.g.f. of X_1 is finite in some open interval about 0.

Let J denote the large deviation rate for X_1 (see 2.10). Let $\Gamma_a^J = \{b: J(b) \le a\}$. Then Γ_a^J is an interval $[b_1,b_2]$ with $-\infty < b_1 < b_2 < \infty$ since $E(X_1) = 0$ and the m.g.f. is finite in an open interval about 0.

With $\Delta_{m,n,a} = \Delta_{m,n,a}(1)$ we now state the classical E-R type laws in a manner that contrasts them with their functional counterparts. The Erdös-Rényi Law, as stated below, is stronger than the original statement and is along the lines of that given in Theorem 5 of Acosta and Kuelbs (1983) for moving averages of i.i.d. random variables taking values in a Banach space.

Shepp's Law. The set of cluster points of $\{\Delta_{n,n,a}\}$ is $\Gamma_a^J = [b_1,b_2]$. In particular $\underline{\lim}\, \Delta_{n,n,a} = b_1$ and $\overline{\lim}\, \Delta_{n,n,a} = b_2$ a.s..

For a set A and, $\varepsilon > 0$, let $A_\varepsilon = \{b: |b-a| < \varepsilon \text{ for some } a \in A\}$.

The Erdös-Rényi Law. Let $\varepsilon > 0$. Then with probability 1,

$$\{\Delta_{m,n,a} : m \le n\} \subset (\Gamma_a^J)_\varepsilon \text{ eventually}$$

and

$$\Gamma_a^J \subset \{\Delta_{m,n,a} : m \le n\}_\varepsilon \text{ eventually.}$$

In particular, the set of cluster points of the triangular array $\{\Delta_{m,n,a} : m \le n\}$ is $\Gamma_a^J = [b_1,b_2]$.

Let $I(\cdot)$ denote the rate function for the sample average process. This is defined in terms of $J(\cdot)$ in Section 3. Let $\Gamma_a^I = \Gamma_a = \{f: I(f) \le a\}$ which is "compact" (Theorem 3.1, Definitions 2.1 and 2.2) and convex (definition of I from the convexity of J).

In Section 4 we establish the following functional counterparts of Shepp's and of Erdös and Rényi's laws:

(1.1) The set of cluster points of $\{\Delta_{n,n,a}(\cdot)\}$ is Γ_a.

(1.2) Let $\varepsilon > 0$. With probability 1,

$$\{\Delta_{m,n,a}(\cdot) : m \le n\} \subset \Gamma_{a,\varepsilon} \text{ eventually}$$

and

$$\Gamma_a \subset \{\Delta_{m,n,a}(\cdot) : m \le n\}_\varepsilon .$$

In particular, the set of cluster points of the triangular array $\{\Delta_{m,n,a}(\cdot) : m \le n\}$ is Γ_a.

Remark 1.1. Since Γ_a is "compact", we have from (1.1) that the set of cluster points of $H(\Delta_n)$ is $H(\Gamma_a)$ and $\overline{\lim}\, H(\Delta_{n,n,a}(\cdot)) = \sup_{f\in\Gamma_a} H(f) < \infty$ for any "continuous" function H. In particular, if $H(f) = f(1)$, then the set of cluster points of $\{\Delta_{n,n,a}\}$ is $\{f(1) : f\in\Gamma_a\}$ and $\overline{\lim}\, \Delta_{n,n,a} = \sup_{f\in\Gamma_a} f(1)$ which is a restatement of Shepp's law. A similar remark can be made concerning the Erdös-Rényi law.

The original Erdös-Rényi Law, which states that $\lim \max_{m<n} \Delta_{m,n,a} = b_2$ and $\lim \min_{m<n} \Delta_{m,n,a} = b_1$ a.s, does not seem to be a direct consequence of the functional law (1.2).

Remark 1.2. The issues of what space in which $\Delta(\cdot)$ takes its values and with what topology have been ignored here. These will be addressed in Sections 3 and 4.

The key to proving (1.1) and 1.2) is to establish the "large deviation principle" (LDP) for $X_n(t) = P(nt/n)$. Roughly speaking, the principle says that $P(X_n(\cdot)\in f+\underline{\underline{df}}) \doteq e^{-nI(f)}df$. This provides suitable approximations for

certain probability statements about the sample average process, $\bar{X}_n(\cdot)$. Once such approximations are established, the proofs of (1.1) and (1.2) are essentially those of Shepp and of Erdös and Rényi with modifications taken from Acosta and Kuebb. This is done in Section 4. A review concerning the LDP is given in Section 2 while the LDP is established for $\bar{X}_n(\cdot)$ in Section 3.

2. The Large Deviation Principle. Let X be a topological space and $\mathcal{F}$ be the Borel σ-field in X. Let $\{P_n\}$ be a family of probability measures on $(X,\mathcal{F})$. Here we review some general ideas regarding large deviations for $\{P_n\}$ while in the next section we are interested in results specifically for P_n, the probability measure induced by the sample average process $\bar{X}_n(\cdot)$.

The following definitions which are slight variants of Varadhan (1984) allow us to state many large deviation results in a concise form.

Definition 2.1. A function $I(\cdot)$ on X is said to be a regular rate function if

(2.1) $0 \le I(x) \le \infty$,

(2.2) $I(\cdot)$ is lower semi-continuous (lsc) and

(2.3) for each $c < \infty$, $\Gamma_c = \{x: I(x) \le c\}$ is compact.

For any subset A of X, define

(2.4) $$I(A) = \inf_{x \in A} I(x).$$

Definition 2.2. The measures $\{P_n\}$ satisfy the large deviation principle (LDP or LD principle) with rate function $I(\cdot)$ if

(2.5) $I(\cdot)$ is a regular rate function,

(2.6) for each closed set F,

$$\overline{\lim} \frac{1}{n} \log P_n(F) \leq - I(F), \text{ and}$$

(2.7) for each open set G,

$$\underline{\lim} \frac{1}{n} \log P_n(G) \geq - I(G).$$

Definition 2.3. The measures $\{P_n\}$ satisfy the weak large deviation principle (WLDP or the weak LD principle) with rate function $I(\cdot)$ if (2.5) and (2.7) of Definition 2.2 together with (2.8) below are satisfied:

(2.8) for each compact set K, $\overline{\lim} \frac{1}{n} \log P_n(K) \leq - I(K)$.

Definition 2.4. The measures $\{P_n\}$ are large deviation tight (LD tight) if, for each $M > \infty$, there exists a compact set K_M such that

(2.9) $$\overline{\lim} \frac{1}{n} \log P_n(K_M^c) \leq - M.$$

The following lemma whose proof is simple and left to the reader shows the usefulness of LD tightness.

Lemma 2.5. Let $\{P_n\}$ be LD tight and satisfy the WLDP. Then it satisfies the LDP.

Many interesting applications in large deviations occur when X is a Polish space, that is a separable complete metric space. This is the setting in Section 3. In this context, two important and immediate derivatives of the LDP are the contraction principle and the asymptotic expression for certain intergrals. This latter result is not used in the sequel but is stated below for completeness. For proofs see Varadhan (1966, 1984).

Let $\{P_n\}$ satisfy the LDP with rate function I(x). Let h be a continuous map from X into another Polish space Y, and let $Q_n = P_n h^{-1}$.

Contraction Principle. The measures $\{Q_n\}$ satisfy the LDP with rate function

$$K(y) = \inf_{x:h(x)=y} I(x).$$

Asymptotic expression for certain integrals. Let F be a bounded real valued continuous function on X. Then

$$\frac{1}{n} \log \int \exp(\lambda F(x)) \, dP_n(x) \to \sup [F(x)-I(x)].$$

The LDP, along with the above two results, has the flavor of weak convergence of probability measures (Theorems 2.1 and 5.1 of Billingsley, 1968). The following lemma is the analogue of the converse part of Prohorov's Theorem (Billingsley, 1968, Theorem 6.2) and with Lemma 2.5 shows that for Polish spaces the LDP is equivalent to the WLDP and LD tightness. The proof is similar to Billingsley's proof of Prohorov's Theorem and can be found in Lynch and Sethuraman (1984).

Lemma 2.6. If $\{P_n\}$ is a sequence of probability measures which satisfies the LDP, then $\{P_n\}$ is LD tight.

The earliest example of the LDP is when P_n is the probability measure induced by the average of n i.i.d. observations. It is summarized in the following theorem which is variously referred to as Cramer's Theorem and Chernoff's Theorem.

Let $X_1, X_2, \ldots$ be i.i.d. with m.g.f. $\phi(\theta) = E(e^{\theta X_1})$. Let $\psi(\theta) = \log\phi(\theta)$ denote the cumulant generating function of X_1 and define

$$(2.10) \qquad J_X(a) = J(a) = \sup_{\theta} \{\theta a - \psi(\theta)\}.$$

Let P_n denote the probability measure induced by $\bar{X}_n = (X_1+\ldots+X_n)/n$. Then,

Theorem 2.7. (Cramer, 1938; Chernoff, 1952). The probability measures $\{P_n\}$ satisfy the LDP with rate function $J(\cdot)$.

The following results show how LD properties for marginal measures carry over to the product measures. These are needed in the next section. The proof of Lemma 2.8 is obvious while that of Lemma 2.9 may be found in Lynch and Sethuraman (1984).

Let $\{P_n^i\}$ be a sequence of probability measures on a Polish space X^i, $i = 1,2$. Let $P_n = P_n^1 \times P_n^2$ be the product measure on $X = X^1 \times X^2$.

Lemma 2.8. If $\{P_n^i\}$ is LD tight for $i = 1,2$, then $\{P_n\}$ is LD tight.

Lemma 2.9. Let $\{P_n^i\}$ satisfy the WLDP with rate function $I^i(x_i)$, $i = 1,2$. Then $\{P_n\}$ satisfies the WLDP with rate function $I(x_1, x_2) = I^1(x_1) + I^2(x_2)$.

The following corollary follows from Lemmas 2.6, 2.8 and 2.9.

Corollary 2.10. Let $\{P_n^i\}$ be LD tight and satisfy the WLDP, $i = 1,2$. Then $P_n = P_n^1 \times P_n^2$ satisfies the LDP with rate function $I(x_1,x_2) = I^1(x_1) + I^2(x_2)$.

3. The LDP for the Sample Average Process. Let $X_1, X_2 \ldots$ be i.i.d. r.v.'s. We assume that the m.g.f., $\phi(\theta)$, is finite in some open interval about zero. Thus the mean is finite and we will without loss of generality assume that $E(X_1) = 0$. Recall that from (2.10) the large deviation rate for $\bar{X}_n$ is $J(a) = \sup_\theta\{\theta a - \log\phi(\theta)\}$.

To state precisely the LDP for the sample average process $\bar{X}_n(\cdot)$, where $\bar{X}_n(t) = P(nt)/n$ we make the following digression. We note that $J(\cdot)$ is convex with its minimum zero at zero since $E(X_1) = 0$. So, $J(a)/a$ is increasing (decreasing) in $a > 0$ ($a < 0$). Let $C_1 = \lim_{a\to\infty} J(a)/a$ and $C_2 = \lim_{a\to\infty} J(-a)/a$.

Let BV[0,1] denote the space of functions which are right continuous and of bounded variation on [0,1] endowed with the weak* topology - better known as the topology of weak convergence or convergence in distribution amongst statisticians and probabilists. Let C[0,1] denote the space of continuous function on [0,1] with the usual uniform topology.

For $f \in BV[0,1]$, let $f = h_1 - h_2$ denote the Hahn-Jordan decomposition of f, let $\dot{f}$ denote the Radon-Nikodym derivative of its absolutely continuous part and let h_1^* and h_2^* denote the singular parts of h_1 and h_2. Let $I(f) = \int_0^1 J(\dot{f})dt + C_1 h_1^*[0,1] + C_2 h_2^*[0,1]$ and $f \in BV[0,1]$ and $= \infty$ if not, where we adopt the convention that $0 \cdot \infty = 0$. In particular, if $C_1 = \infty = C_2$ then $I(f) = \int_0^1 J(\dot{f})dt$ if f absolutely continuous and ∞ if not.

If either C_1 or C_2 is finite (Case 1) we must view $\bar{X}_n(\cdot)$ as taking values in BV[0,1]. If C_1 and C_2 are both infinite (Case 2) we may view $\bar{X}_n(\cdot)$ as taking values in either BV[0,1] or C[0,1]. The reason we must distinguish these two cases is that $\Gamma_a = \{f: I(f) \leq a\}$ is not a compact subset of C[0,1] when Case 1 obtains.

Theorem 3.1. Let P_n denote the probability measure induced by $\bar{X}_n(\cdot)$. Then $\{P_n\}$ satisfies the LDP with rate function $I(\cdot)$.

The proof that $I(\cdot)$ is a regular rate function can be found in Lynch and Sethuraman (1984) for Case 1 and in Varadhan (1966) for Case 2 (c.f. Groeneboom et al., 1979). Here we only outline the proof of the upper and lower probability bounds in the principle.

To proof of these bounds is somewhat technical but in reality involves only three key ideas. First the process $\bar{X}_n(\cdot)$ is approximated by a finite dimensional random vector whose components are independent. Then Corollary 2.10 with Theorem 2.7 and the contraction principle establishes the LDP for

the finite dimensional process. Finally, Lemma 3.2 and Theorem 3.3, below, show that the approximations are suitable enough to establish the principle.

We need the following. For a partition $P = \{0 = t_0 < t_1 < \ldots < t_k = 1\}$ let $\Delta_i t = t_i - t_{i-1}$ and $\Delta_i f = f(t_i) - f(t_{i-1})$. Let $I_P(f) = \sum_1^k J(\Delta_i f/\Delta_i t)\Delta_i t$. Analogous to (2.5) let $I_P(A) = \inf_{f \in A} I_P(f)$. Let $A = \text{ess inf } X_1$ and $B = \text{ess sup } X_1$. Note that if A is finite then $J(A) = -\log P(X_1 = A)$. A similar statement about $J(B)$ can be made if B is finite. Also note that $J(\cdot)$ is continuous on $[A,B]$ (even if A or B is infinite) where if $J(A) = \infty$, the $\lim_{a \to A} J(a) = \infty$ etc. Thus $I_P(\cdot)$ is continuous on $[A,B]^k$.

Let $\mathcal{B}$ denote the Borel σ-field in $[0,1]$.

Lemma 3.2. $I_P(f) \uparrow I(f)$ as $\sigma(P) \uparrow \mathcal{B}$.

Theorem 3.3. (The Minimax Theorem) if F is closed, then $\sup_P I_P(F) = I(F)$.

Remark 3.4. The proof of Lemma 3.2 depends on the fact that $\{f_P, \sigma(P)\}$, where $f_P = E(f_P(U) \mid \sigma(P))$ and U is uniform on $[0,1]$, is a martingale and can be found in Lynch and Sethuraman (1984). A proof of the minimax theorem for Case 1 can also be found there. The proof for Case 2 is similar (c.f. Groeneboon et al., 1979, Lemma 2.4).

Outline - lower bound. We only do the proof for Case 1. The proof for Case 2 is similar but somewhat more complicated and may be found in Varadhan (1966).

Let G be an open set. If $I(G) = \infty$ there is nothing to prove. So assume that $I(G) < \infty$. For $\varepsilon > 0$ choose f such that $I(f) < I(G) + \varepsilon$. There is a partition $P = \{0 = t_0 < t_1 < \ldots < t_k = 1\}$ of continuity points f such that $N_{P,\varepsilon} = \{g : \max_{i \le k} |\Delta_1 g \quad \Delta_1 f| < \varepsilon\} \subset G$.

Since f is continuous at t_i, for all sufficiently large n, $\{\overline{X}_n(\cdot) \in N_{p,\varepsilon}\}$ $\supset \{|S_{[nt_1]}/n - (f(\frac{[nt_1]}{n}) - f(0))| < \varepsilon, |X_{[nt_1]+1}/n| < \varepsilon, |(S_{[nt_2]} - S_{[nt_1]+1})/n - (f(\frac{[nt_2]}{n}) - f(\frac{[nt_1]}{n}))| < \varepsilon, \frac{X_{[nt_2]+1}}{n} < \varepsilon$, etc.}. So for such an n, by independence,

$$(3.1) \qquad P(\overline{X}_n(\cdot) \in N_{p,\varepsilon}) \geq \prod_{i=0}^{k-1} P\left(\frac{S_{[nt_{i+1}]}-[nt_i]+1}{n} - \left(f\left(\frac{[nt_{i+i}]}{n}\right) - f\left(\frac{[nt_i]}{n}\right)\right| < \varepsilon\right) \times P^k\left(\frac{|X_1|}{n} < \varepsilon\right).$$

Since $P^k(\frac{|X_1|}{n} < \varepsilon) \to 1$ as $n \to \infty$, by (3.1) and Theorem 2.7,

$$\underline{\lim}\ n^{-1}\log P(\overline{X}_n(\cdot) \in N_{p,\varepsilon}) \geq -I_p(N_{p,\varepsilon}) \geq -I_p(f) \geq -I(f)$$

where the last inequality follows from Lemma 3.2 and the second to last from the definition of $I_p(\cdot)$.

Outline - the upper bound. Let F be a closed set. Fix a partition $P = \{0 = t_0 < t_1 < \ldots < t_k = 1\}$. For simplicity assume $k = 2$. Note that, from the definition of $I_p(\cdot)$,

$$P(\overline{X}_n(\cdot) \in F) \leq P(I_p(\overline{X}_n(\cdot)) \geq I_p(F)).$$

Thus, to establish the upper bound it suffices to show that

$$(3.2) \qquad \overline{\lim}\ n^{-1}\log P(I_p(\overline{X}_n(\cdot)) \geq a) \leq -a.$$

Now the event

$$(3.3). \qquad \{I_p(\overline{X}_n(\cdot) \geq a\} \subset \bigcup_{j=1}^{8} Ajn$$

where $A_{1n} = \{X_{[nt_1]+1} \geq 0,\ S_{[nt_1]} \geq 0,\ S_n - S_{[nt_1]+1} \geq 0,$ and $I_p(\overline{X}_n(\cdot)) \geq a\}$ and $A_{2n}, \ldots, A_{8n}$ denote the other possible sets resulting from the seven other choices of the first three inequalities. For A_{1n}, let K be a fixed finite positive number less than or equal to B. Then, since J is nonnegative and increasing in $a > 0$ and convex

$$(3.4) \qquad P(A_{1n}) \leq P\left(J\left(\frac{S_{[nt_1]}+K}{nt_1}\right)t_1 + J\left(\frac{S_n - S_{[nt_1]+1}+K}{n(1-t_1)}\right)(1-t_1) \geq a\right) +$$

$$P\left(\frac{X_{[nt_1]+1}}{nt_1} \geq K\right) = P_n + Q_n.$$

Since J is continuous on [A,B] and $0 < K \leq B$ it follows from Theorem 2.7, Corollary 2.10 and the contraction principle that

$$(3.5) \qquad \overline{\lim}\, n^{-1} \log P_n \leq -a.$$

By Markov's inequality,

$$(3.6) \qquad n^{-1}\log Q_n \leq -t_1\theta K + \frac{\psi(\theta)}{n} \to -t_1\theta K$$

as $n \to \infty$ for every θ for which $\psi(\theta) < \infty$. If $B < \infty$, then $\psi(\theta) < \infty$ for all $\theta > 0$, while if $B = \infty$, then K may be taken arbitrarily large. In any event (3.4) combined with (3.5) and (3.6) and the above observation show that $\overline{\lim}\, n^{-1}\log P(A_{1n}) \leq -a$. This with (3.3) proves (3.2) and completes the proof of the upper bound.

4. Functional E-R Laws. With the formulation the same as in Section 3, let $I(\cdot)$ denote the rate function for $\overline{X}_n(\cdot)$ and recall the definition of $\Delta_{m,n,a}(\cdot)$ in the Introduction. As before $\Gamma_a = \{f: I(f) \leq a\}$.

For a set A, let $A_\varepsilon = \{g: d(f,g) < \varepsilon \text{ for some } f \in A\}$ where these sets are viewed as subsets of BV[0,1] (Case 1) or C[0,1] (Case 2) and d is either the metric which induces the weak* topology (Case 1) or the uniform topology (Case 2).

The functional analogues of Shepp's (1964) and Erdös and Rényi's laws can be stated as follows.

Theorem 4.1. The set of cluster points of $\{\Delta_{n,n,a}(\cdot)\}$ is Γ_a.

Theorem 4.2. Let $\varepsilon > 0$. With probability 1,

(i) $\{\Delta_{m,n,a}(\cdot): m \le n\} \subset \Gamma_{a,\varepsilon}$ eventually

and

(ii) $\Gamma_a \subset \{\Delta_{m,n,a}(\cdot): m \le n\}_\varepsilon$.

(iii) In particular, the set of cluster points of the triangular array $\{\Delta_{m,n,a}(\cdot): m \le n\}$ is Γ_a.

To prove these theorems we need the following lemma.

Lemma 4.3. (i) For each $\varepsilon > 0$ there exists a $c > a$ for which $\Gamma_c \subset \Gamma_{a,\varepsilon}$ and (ii) Γ_a equals the closure of $\{f: I(f) < a\}$.

Proof. (i) Suppose not. Then for every $c > a$ there exists an $f_c \in \Gamma_c$ with $f_c \notin \Gamma_{a,\varepsilon}$. Fix $b > a$. Since Γ_b is compact the net $\{f_c: a < c < b\}$ has a subnet $\{f_d\}$ which converges, say, to f_0, as $d \downarrow a$. By the lsc of $I(\cdot)$ it follows that $I(f_0) \le \lim_{d \downarrow a} I(f_d) \le a$, and so $f_0 \in \Gamma_a$. But $f_c \notin \Gamma_{a,\varepsilon}$ implies that $d(f_0, f_c) > \varepsilon$ for every c which contradicts that $\{f_d\}$ converges to f_0.

(ii) Fix $f\epsilon\Gamma_a$ with $I(f) = a$. It suffices to show that $I(\alpha f) < \alpha$ for $0 < \alpha < 1$. This is immediate since I is convex with $I(\underline{0}) = 0$, where $\underline{0}$ denotes the function which is identically zero. ||

Proof of Theorems 4.1 and 4.2. Fix $a \geq 0$ and $\epsilon > 0$. We first show that

(4.1) for any $f\epsilon\Gamma_a$, $d((\Delta_{n,n,a}(\cdot), f) < \epsilon)$ infinitely often a.s..

Let $0 = \{g: d(g,f) < \epsilon\}$. Note that since 0 is open, $\alpha \cdot f \epsilon 0$ for some $\alpha\epsilon(0,1)$. By the LDP,

$$\underline{\lim}\, n^{-1} \log P(\bar{X}_n \epsilon 0) \geq -I(0) \geq -I(\alpha f).$$

Since $I(\cdot)$ is convex with $I(0) = 0$ and $\alpha\epsilon(0,1)$, $I(\alpha f) \leq \alpha I(f) \leq \alpha a$. Thus,

(4.2) $$P(\Delta_{n,n,a}(\cdot)\epsilon 0) \geq n^{-\alpha(1+0(1))}.$$

Since $\Sigma n^{-1} = \infty$, for $\alpha'\epsilon(\alpha,1)$ fixed it follows from Lemma (3.1) of Shepp (1964) that there exists a sequence $\{n(k)\}$ with $n(k+1) = n(k) + [a^{-1} \log n(k)]$ such that $\Sigma n(k)^{-\alpha'} = \infty$. This with (4.2) shows that $\Sigma P(\Delta_{n(k),n(k),a}(\cdot)\epsilon 0) = \infty$. Statement (4.1) follows from this and the divergent part of the Borel-Cantelli lemma since the events $\{\Delta_{n(k),n(k),a}(\cdot)\epsilon 0\}$, $k = 1, 2, \ldots$ are independent.

We now show that, with probability 1,

(4.3) $$\{\Delta_{m,n,a}(\cdot): m \leq n\} \subset \Gamma_{a,\epsilon} \text{ eventually.}$$

For all sufficiently large integer k, let $K = e^{ak+1}$ and let $l(k) = K-1$ if K is an integer and $= [K]$ if not. Then, since $k = [a^{-1} \log k]$ for $l(k-1) < n \leq l(k)$, $\Delta_{m,n,a}(\cdot) = \Delta_{m,l(k),a}(\cdot)$ for $m \leq n$. Thus, to prove (4.3), it suffices to show that

(4.4) $$P(\Delta_{m,l(k),a}(\cdot)\epsilon\Gamma_{\alpha,\epsilon} \text{ for some } m \leq l(k) \text{ i.o.}) = 0.$$

By Lemma 4.3, (i) there exists a $c > a$ such that $\Gamma_c \subset \Gamma_{a,\varepsilon}$. Thus $I(\Gamma^c_{a,\varepsilon}) \geq c$. So by the LDP,

$$\overline{\lim}\, n^{-1} \log P(\bar{X}_n(\cdot) \in \Gamma^c_{a,\varepsilon}) \leq -I(\Gamma^c_{a,\varepsilon}) \leq -c$$

since $\Gamma_{a,\varepsilon}$ is closed with $\Gamma^c_{a,\varepsilon} \subset \Gamma_c{}^c$. Thus, for k' sufficiently large,

$$\sum_{k \geq k'} P(\Delta_{m,l(k),a}(\cdot) \in \Gamma^c_{a,\varepsilon} \text{ for some } m \leq l(k))$$

$$\leq \sum_{k \geq k'} l(k) e^{-ck(1+0(1))} = \sum_{k \geq k'} e^{l} e^{(a-c)(1+0(1))} < \infty.$$

So (4.4) follows from the convergent part of Borel-Cantelli lemma. This completes the proof of (4.3).

Since Γ_a is closed, $\Gamma_{a,\varepsilon} \downarrow \Gamma_a$ as $\varepsilon \downarrow 0$. This with (4.1) and (4.3) completes the proof of Theorem 4.1 and (i) of Theorem 4.2.

To prove (ii) of Theorem 4.2, let $\varepsilon > 0$. Then, by Lemma 4.3 (ii), there is a finite collection $\{f_1, \ldots, f_k\} \subset \Gamma_a$ such that $\Gamma_a \subset \bigcup_{i=i}^{k} \{f_i\}_{\varepsilon/2}$.

Since $\{f_i\}_{\varepsilon/2}$ is an open set, by the LDP,

$$\underline{\lim}\, n^{-1} \log P(\bar{X}_n \in \{f_i\}_{\varepsilon/2}) \geq -I(\{f_i\}_{\varepsilon/2}) \geq -I(f_i).$$

So,

$$(4.5) \qquad P(\Gamma_a \not\subset \{\Delta_{m,n,a} : m \leq n\}_{\varepsilon/2})$$

$$\leq P(f_i \not\in \{\Delta_{m,n,a} : m \leq n\}_{\varepsilon/2} \text{ for some } i)$$

$$\leq \sum_{i=i}^{k} P(f_i \not\in \{\Delta_{m,n,a} : m \leq n\}_{\varepsilon/2})$$

$$\leq \sum_{i=i}^{k} P(f_i \not\in \{\Delta_{m,n,a} : m = \ell\lfloor a^{-1} \log n \rfloor + 1,\ \ell = 0, 1, \ldots + m \leq n\}_{\varepsilon/2})$$

$$\leq \sum_{i=i}^{k} P(\bar{X}_{[a^{-1}\log n]} \in \{f_i\}_{\varepsilon/2})^{(n/[a^{-1}\log n])-1}$$

$$= \sum_{i=i}^{k} (1-P(\bar{X}_{[a^{-1}\log n]} \in \{f_i\}_{\varepsilon/2}))^{(n/[a^{-1}\log n])-1}$$

$$\leq \sum_{i=i}^{k} (1-e^{-I(f_i)[a^{-1}\log n](1+0(1))})^{(n/[a^{-1}\log n])}$$

$$\leq \sum_{i=i}^{k} \ell^{-n(-I(f_i)a^{-1})(1+0(1))/a^{-1}\log n},$$

where in the last inequality we used that $(1-x) \leq \ell^{-x}$.

Since $I(f_i) < a$ for $i = i,\ldots,k$, it follows that $\sum P(\Gamma_a \notin \{A_{m,n,a} : m \leq n\}_{\varepsilon/2}) < \infty$. This with the convergent part of the Borel-Cantelli lemma completes the proof.

Acknowledgements. The authors would like to thank Alex de Acosta for pointing out the Acosta-Kuebbs (1983) version of the Erdös-Rényi Law.

References

Acosta, A. de and Kuebbs, J. (1983). Limit Theorems for moving averages of independent random variables. Z. Wahrscheinlichkeitstheorie Verw. Gebiete. 64, 67-123.

Billingsley, Patrick (1968). Convergence of Probability Measures. Wiley, New York.

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.

Cramer, H. (1937). Sar un nouveau theoreme limite de la theorie des probabilities. Colloquium on theory of probability. Paris-Herman.

Erdös, P. and Rényi, A. (1970). On a new law of large numbers. J. Analyse Math., 23, 103-111.

Groeneboom, P., Oosterhoff, J., and Ruymgaart, F.H. (1979). Large deviation theorems for empirical probability measures. Ann. Probab., 7, 553-586.

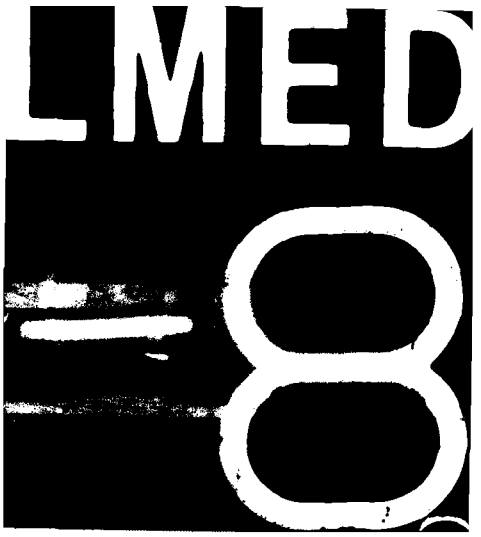
Lynch, J. and Sethuraman, J. (1984). Large deviations for processes with independent increments. FSU Tech Report No. M687.

Shepp, L. A. (1964). A limit theorem concerning moving averages. Ann. Math. Statists., 35, 424-428.

Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. Z . Wahrscheinlichkeitstheorie and Verw. Gebiete 3, 211-226.

Varadhan, S. R. S. (1966). Asymptotic probabilities and differential equations. Comm. Pure Appl. Math., 19, 261-286.

Varadhan, S. R. S. (1984). Large Deviations and Applications. SIAM, CBMS/NSF Regional Conference in Applied Math., 46 SIAM, Philadelphia


LMED
8