MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A
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).
The Large Deviation Principle
for the Sample Average Process
and
Functional Erdős-Rényi Laws
by
James Lynch¹ and Jayaram Sethuraman²
June, 1986

FSU Technical Report No. M-730

¹Department 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.

²Department 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

Approved for public release; distribution unlimited.

The view, opinions, and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.
The Large Deviation Principle for the Sample Average Process and Functional Erdős-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 Erdős-Rényi type laws of Erdős 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 $D_n(t) = P(t)/\sigma \sqrt{n}$ and $S_n(t) = P(t)/\sqrt{n \log \log n}$. Let $J(a) = a^2/2$ which is the large deviation rate (Section 2) for the normal distribution. Let $I(f) = \int_0^1 J(f(t)) \, dt$ if $f$ is absolutely continuous ($\dot{f}$ denotes the Radon-Nikodym of $f$) and $= \infty$ if not. Let $r_a = \{ f : I(f) \leq a \}$. Then, the invariance principles of Donsker (see Billingsley (1968) and Strassen (1964)) state that $\{D_n(t) : 0 \leq t \leq 1\} \Rightarrow \{W(t) : 0 \leq t \leq 1\}$ 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 $r_a$ 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 Erdős-Rényi (E-R) type laws of Erdős 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 \( \Delta_m(n,a)(s) = \Delta_m(s[a^{-1}\log n])/[a^{-1}\log n] \) where \([x]\) indicates the integer part of \( x \). We further 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 
\[
\Gamma_a^J = \{ b: J(b) \leq a \}.
\]
Then \( \Gamma_a^J \) 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 \( \lim \Delta_{m,n,a} = b_1 \) and \( \lim \Delta_{m,n,a} = b_2 \) a.s.

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

**The Erdös-Rényi Law.** Let \( \epsilon > 0 \). Then with probability 1,

\[
\{\Delta_{m,n,a}: m \leq n \} \in (\Gamma_a^J)_{\epsilon} \text{ eventually}
\]

and

\[
\Gamma_a^J \cap \{\Delta_{m,n,a}: m \leq n \} \in (\Gamma_a^J)_{\epsilon} \text{ eventually}.
\]

In particular, the set of cluster points of the triangular array \( \{\Delta_{m,n,a}: m \leq n \} \) 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 \( \Gamma_a^I = \Gamma_a = \{ f: I(f) \leq a \} \) 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 Erdős and Rényi's laws:

(1.1) The set of cluster points of \( \{A_{n,n,a}(\cdot)\} \) is \( \Gamma_a \).

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

\[
\{\Lambda_{m,n,a}(\cdot) : m \leq n\} \subset \Gamma_a \quad \text{eventually}
\]

and

\[
\Gamma_a = \{\Lambda_{m,n,a}(\cdot) : m \leq n\}
\]

In particular, the set of cluster points of the triangular array \( \{\Lambda_{m,n,a}(\cdot) : m \leq n\} \) is \( \Gamma_a \).

Remark 1.1. Since \( \Gamma_a \) is "compact", we have from (1.1) that the set of cluster points of \( H(\Lambda_n) \) is \( H(\Gamma_a) \) and \( \sup_{f \in \Gamma_a} H(\Lambda_{n,n,a}(\cdot)) = \sup_{f \in \Gamma_a} H(f) \).

for any "continuous" function \( H \). In particular, if \( H(f) = f(1) \), then the set of cluster points of \( \{\Lambda_{n,n,a}\} \) is \( \{f(1) : f \in \Gamma_a\} \) and \( \sup_{f \in \Gamma_a} \Lambda_{n,n,a} = \sup_{f \in \Gamma_a} f(1) \) which is a restatement of Shepp's law. A similar remark can be made concerning the Erdős-Rényi law.

The original Erdős-Rényi Law, which states that \( \lim \max_{m \leq n} \Lambda_{m,n,a} = b_2 \) and \( \lim \min_{m \geq n} \Lambda_{m,n,a} = b_1 \) 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 \( \Lambda(\cdot) \) 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 \( \overline{X}_n(t) = P(\text{nt}/n) \). Roughly speaking, the principle says that \( P(\overline{X}_n(\cdot) \in df) \approx e^{-n\mathbb{I}(f)} df \). 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 $F$ be the Borel $\sigma$-field in $X$. Let $(P_n)$ be a family of probability measures on $(X,F)$. 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

1. $0 \leq I(x) \leq \infty$,
2. $I(\cdot)$ is lower semi-continuous (Lsc) and
3. for each $c \leq \infty$, $\Gamma_c = \{x : I(x) \leq c\}$ is compact.

For any subset $A$ of $X$, define

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

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

$$I(\cdot) \text{ is a regular rate function},$$
for each closed set $F$,
\[
\limsup \frac{1}{n} \log P_n(F) \leq -I(F),
\]
and
\[
(2.7) \quad \text{for each open set } G,
\lim \frac{1}{n} \log P_n(G) \geq -I(G).
\]

**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:
\[
(2.8) \quad \text{for each compact set } K, \lim_{n \to \infty} \frac{1}{n} \log P_n(K) \leq -I(K).
\]

**Definition 2.4.** The measures $\{P_n\}$ are large deviation tight (LD tight) if, for each $M > 0$, there exists a compact set $K_M$ such that
\[
(2.9) \quad \lim_{n \to \infty} \frac{1}{n} \log P_n(K_M^C) \leq -M.
\]

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 integrals. 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

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

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, 1983). 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 MLDP 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. \( \phi(\theta) = \mathbb{E}(e^{\theta X_1}) \). Let \( \psi(\theta) = \log \phi(\theta) \) denote the cumulant generating function of \( X_1 \) and define

\[
J_{X}(a) = J(a) = \sup_{\theta} \{ \theta a - \psi(\theta) \}.
\]

Let \( \bar{X}_n \) 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 MLDP with rate function \( I_i(x^i) \), \( i = 1,2 \). Then \( \{P_n\} \) satisfies the MLDP 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 MLDP, \( i = 1,2 \). 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 \( J(\alpha) = \sup_{\Theta} \{\alpha \cdot \log \phi(\theta)\} \).

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, \( J(\alpha)/\alpha \) is increasing (decreasing) in \( \alpha > 0 \) (\( \alpha < 0 \)). Let \( C_1 = \lim_{\alpha \to \infty} J(\alpha)/\alpha \) and \( C_2 = \lim_{\alpha \to -\infty} J(\alpha)/\alpha \).
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 \(\tilde{f}\) 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(\tilde{f}) \, dt + C_1 h_1^*[0,1] + C_2 h_2^*[0,1] \quad \text{and} \quad f \in BV\([0,1]\) \text{ and } = \infty \text{ if not},
\]

where we adopt the convention that \(0 \cdot \infty = 0\). In particular, if \(C_1 = \infty = C_2\) then

\[
I(f) = \int_0^1 J(\tilde{f}) \, dt \text{ if } f \text{ absolutely continuous and } = \infty \text{ if not}.
\]

If either \(C_1\) or \(C_2\) is finite (Case 1) we must view \(X_n(\cdot)\) as taking values in BV\([0,1]\). If \(C_1\) and \(C_2\) are both infinite (Case 2) we may view \(X_n(\cdot)\) 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 \(X_n(\cdot)\). 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 \(X_n(\cdot)\) 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
\[
\Delta_{i} t = t_{i} - t_{i-1} \quad \text{and} \quad \Delta_{i} f = f(t_{i}) - f(t_{i-1}).
\]
Let \( I_{p}(f) = \sum_{i=1}^{k} J(\Delta_{i} f/\Delta_{i} t) \Delta_{i} t. \) 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
\[
\lim_{a \uparrow A} J(a) = \infty \quad \text{etc. Thus } I_{p}(\cdot) \text{ is continuous on } [A, B]^{k}.
\]

Let \( B \) denote the Borel \( \sigma \)-field in \([0, 1]\).

**Lemma 3.2.** \( I_{p}(f) + I(f) \) as \( \sigma(P) + B. \)

**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 \( f_{p} = E(f_{p}(U) | \sigma(P)) \) 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 \( \epsilon > 0 \) choose \( f \) such that \( I(f) < I(G) + \epsilon. \) There is a partition \( P = \{0 = t_0 < t_1 < \ldots < t_k = 1\} \) of continuity points \( f \) such that \( N_{p, \epsilon} = \{g: \max_{i \leq k} \left\{ \frac{f}{\epsilon} \right\} \in G. \)
Since \( f \) is continuous at \( t_i \), for all sufficiently large \( n \), \( \{X_n(\cdot) \in N_{p_x} \} \)

\[
\geq \left| \frac{S_{\lfloor nt_1 \rfloor}}{n} - (f(\frac{\lfloor nt_1 \rfloor}{n}) - f(0)) \right| < \varepsilon, \left| \frac{S_{\lfloor nt_2 \rfloor} + 1}{n} - f(\frac{\lfloor nt_1 \rfloor}{n}) \right| < \varepsilon, \frac{X_{\lfloor nt_2 \rfloor} + 1}{n} < \varepsilon, \text{etc.}. \]

So for such an \( n \), by independence,

\[
P(X_n(\cdot) \in N_{p_x}) \geq \prod_{i=0}^{k-1} P\left( \frac{S_{\lfloor nt_1 \rfloor} - \lfloor nt_1 \rfloor + 1}{n} - (f(\frac{\lfloor nt_1 \rfloor}{n}) - f((\frac{\lfloor nt_1 \rfloor}{n})) \right| < \varepsilon) \times P\left( \frac{X_1}{n} < \varepsilon \right).
\]

Since \( p^k \left( \frac{X_1}{n} < \varepsilon \right) \to 1 \) as \( n \to \infty \), by (3.1) and Theorem 2.7,

\[
\lim n^{-1} \log P(X_n(\cdot) \in N_{p_x}) \geq -I_p(N_{p_x}) \geq -I_p(f) \geq -I(f)
\]

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 \( I_p(\cdot) \),

\[
P(X_n(\cdot) \in F) \leq P(I_p(X_n(\cdot)) \geq I_p(F)).
\]

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

\[
\lim n^{-1} \log P(I_p(X_n(\cdot)) \geq a) \leq -a.
\]

Now the event

\[
\{I_p(X_n(\cdot) \geq a) \in \bigcup_{j=1}^8 \text{Aj}_n \}.
\]
where \( A_{ln} = \{ X_{nt1} + 1 \geq 0, S_{nt1} \geq 0, S_n - S_{nt1} + 1 \geq 0, \text{ and } I_p(\chi_n(\cdot)) \geq a \} \) and \( A_{2n}, \ldots, A_{8n} \) denote the other possible sets resulting from the seven other choices of the first three inequalities. For \( A_{ln} \), 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

\[
P(A_{ln}) \leq P(J \left( \frac{S_{nt1} + K}{nt1} \right) + J(\frac{S_n - S_{nt1} + 1 + K}{n(1-t_1)}) (1-t_1) \geq a) +
\]

\[
\frac{X_{nt1} + 1}{nt1} \geq K = P_n + Q_n.
\]

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

\[
(3.5) \quad \overline{\lim} \, n^{-1} \log P_n \leq -a.
\]

By Markov's inequality,

\[
(3.6) \quad n^{-1} \log Q_n \leq -t_1 \theta K + \frac{\psi(\theta)}{n} + -t_1 \theta K
\]

as \( n \to \infty \) for every \( \theta \) 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

\[
\overline{\lim} \, n^{-1} \log P(A_{ln}) \leq -a.
\]

This with (3.3) proves (3.2) and completes the proof of the upper bound.

\section*{4. Functional E-R Laws.} With the formulation the same as in Section 3, let \( I(\cdot) \) denote the rate function for \( \chi_n(\cdot) \) and recall the definition of \( A_{ln,a}(\cdot) \) in the Introduction. As before \( r_a = \{ \hat{f} : I(\hat{f}) \leq a \} \).
For a set $A$, let $\Lambda_\varepsilon = \{ g : d(f, g) < \varepsilon \text{ for some } f \in A \}$ 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 Erdős and Rényi's laws can be stated as follows.

**Theorem 4.1.** The set of cluster points of $\{ \Lambda_{n,n,a}(\cdot) \}$ is $\Gamma_a$.

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

(i) $\{ \Lambda_{m,n,a}(\cdot) : m \leq n \} \in \Gamma_{a,\varepsilon}$ eventually and

(ii) $\Gamma_a \subseteq \{ \Lambda_{m,n,a}(\cdot) : m \leq n \} \in \Gamma_{a,\varepsilon}$.

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

To prove these theorems we need the following lemma.

**Lemma 4.3.** (i) For each $\varepsilon > 0$ there exists a $c > a$ for which $\Gamma_c \subseteq \Gamma_{a,\varepsilon}$ and (ii) $\Gamma_a$ 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 $\Gamma_b$ 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 $I(f_0) \leq \lim I(f_d) \leq a$, 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 $f \in \mathcal{A}$ with $I(f) = a$. It suffices to show that $I(af) < a$ for $0 < a < 1$. This is immediate since $I$ is convex with $I(0) = 0$, where 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

\begin{equation}
(4.1) \quad \text{for any } f \in \mathcal{A}, \; d((\Delta_n, n, a(\cdot), f) < \epsilon) \text{ infinitely often a.s.}
\end{equation}

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

\[\lim_{n \to \infty} n^{-1} \log P(\cap_{n \geq 0} a_n) \geq I(0) = -I(a).\]

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

\begin{equation}
(4.2) \quad P(\cap_{n \geq 0} a_n \varepsilon 0) \geq n^{-a(1+O(1))}.
\end{equation}

Since $\sum_{n}^{-1} = -$, for $a' \in (a,1)$ 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 $\sum_{n(k)} a' = 0$. This with (4.2) shows that $\sum P(\cap_{n \geq 0} a_n) = 0$.

Statement (4.1) follows from this and the divergent part of the Borel-Cantelli lemma since the events $\{\Delta_n(k), n(k), a(\cdot) \varepsilon 0\}, \; k = 1, 2, \ldots$ are independent.

We now show that, with probability 1,

\begin{equation}
(4.3) \quad \{\Delta_m, n, a(\cdot): m \leq n \in \cap_{a, \epsilon} \text{ eventually.}
\end{equation}

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

\begin{equation}
(4.4) \quad P(\Delta_m, 1(k), a(\cdot) \varepsilon 0 \text{ for some } m \leq 1(k) \text{i.o.}) = 0.
\end{equation}
By Lemma 4.3, (i) there exists a $c > a$ such that $r_c \in r_a$. Thus $I(r_c^a, e) \geq c$. So by the LDP,

$$\lim n^{-1} \log P(X_n, e_r^a, e) \leq -I(r_c^a, e) \leq -c$$

since $r_a^e$ is closed with $r_c^a \in r_c^a$. Thus, for $k'$ sufficiently large,

$$\sum_{k \geq k'} \mathbb{P}(A_{m, l(k)}, e_r^a, e) \geq c$$

for some $m \leq l(k)$

$$\leq \sum_{k \geq k'} 1(k)e^{-ck(1+O(1))} \leq e^{l(a-c)(1+O(1))} \cdot e.$$ 

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

Since $r_a$ is closed, $r_a^e \uparrow r_a$ as $e \downarrow 0$. 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 $e > 0$. Then, by Lemma 4.3 (ii), there is a finite collection $\{f_1, \ldots, f_k\} \subset r_a$ such that $r_a \in \bigcup_{i} f_i$.

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

$$\lim n^{-1} \log P(X_n, e(f_i)_{e/2}) \geq -I(\{f_i\}_{e/2}) \geq -I(f_i).$$

So,

$$P(r_a \notin \{A_m, n, a : m \leq n\}_{e/2}) \leq P(f_i \notin \{A_m, n, a : m \leq n\}_{e/2}) \leq \sum_{i=1}^k P(f_i \notin \{A_m, n, a : m \leq n\}_{e/2}) \leq \sum_{i=1}^k P(f_i \notin \{A_m, n, a : m \leq n\}_{e/2}) \leq \sum_{i=1}^k P(f_i \notin \{A_m, n, a : m \leq n\}_{e/2})$$
\[
\sum_{i=1}^{k} P(\mathcal{X}_{a^{-1} \log n} \notin \mathcal{F}_i \epsilon_{\sqrt{2}}) (n/[a^{-1} \log n])^{-1}
\]
\[
= \sum_{i=1}^{k} (1-P(\mathcal{X}_{a^{-1} \log n} \notin \mathcal{F}_i \epsilon_{\sqrt{2}})) (n/[a^{-1} \log n])^{-1}
\]
\[
\leq \sum_{i=1}^{k} (1-e^{-I(\mathcal{F}_i)[a^{-1} \log n]}(1+0(1))(n/[a^{-1} \log n])
\]
\[
\leq \sum_{i=1}^{k} e^{-\varepsilon n(-I(\mathcal{F}_i)a^{-1})(1+0(1))/a^{-1} \log n}
\]
where in the last inequality we used that \((1-x) \leq e^{-x}\).

Since \(I(\mathcal{F}_i) < a\) for \(i = i, \ldots, k\), it follows that \(E P(\mathcal{F}_i \notin \mathcal{F}_{\mathcal{F}_i} \epsilon_{\sqrt{2}}) < \infty\).

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 Erdős-Rényi Law.
References


