



LARGE DEVIATION PROBABILITIES AND SOME RELATED TOPICS

by

Josef Steinbach

Department of Mathematics and Statistics
Carleton University
Ottawa, Canada

Carleton Mathematical Lecture Notes No. 28 - Sept. 1980

QA
15
.S84

QA
15
S84

LARGE DEVIATION PROBABILITIES

AND SOME RELATED TOPICS

Josef Steinebach

Department of Mathematics and Statistics

Carleton University

Ottawa, Canada

Carleton Mathematical Lecture Notes No. 28, Sept. 1980.

UNIVERSITY
LIBRARY
SASKATOON
SASKATCHEWAN

CONTENTS

Preface and acknowledgements	1
Introduction	3
1. <u>Exponential convergence rates of large deviation probabilities</u>	5
1.1 Moment-generating functions (m.g.f.'s)	5
1.2 Chernoff's theorem	8
1.3 Necessity of moment-generating functions (i.i.d. case)	12
1.4 General sequences	14
1.5 Necessity of moment-generating functions (general case)	17
1.6 Some recent results	18
2. <u>Asymptotic relative efficiencies (A.R.E.'s) of test statistics and estimators.</u>	21
2.1 Kullback-Leibler information numbers	21
2.2 Stein's lemma. Large sample behaviour of the Neyman-Pearson test	27
2.3 Asymptotic effective variances and A.R.E.'s of estimators	32
2.4 Bahadur slopes and A.R.E.'s of test statistics	37
3. <u>Fluctuations of random processes. Erdős-Rényi laws.</u>	45
3.1 The Erdős-Rényi new law of large numbers	45
3.2 Fluctuations of renewal processes	56
3.3 Convergence rates in Erdős-Rényi laws	65
3.4 General versions of Erdős-Rényi laws	72
References	81

PREFACE AND ACKNOWLEDGEMENTS

Special interest has always been focussed by probabilists and statisticians as well to the study of limit theorems. The theory of large deviation probabilities as a part of those studies has been widely extended during the last three decades and has found various different applications in probability theory and statistics. Cramér's (1938) famous asymptotic expansion in the central limit theorem may be viewed as a starting point in this area, already providing a great deal of fundamental ideas. Heavy use is made of the concept of conjugate variables, introduced by Esscher (1932), which turns out to be the most adequate tool.

This short monograph is based on eight lectures given at Carleton University, Ottawa, during the 1980 winter term. I wish to express my sincere thanks to the Department of Mathematics for the kind invitation to spend the first nine months of 1980 there, in particular, to Professors M. Csörgö, D.A. Dawson and J.N.K. Rao who supported this visit by their NSERC grants. I am also grateful to the Deutsche Forschungsgemeinschaft for contributing to this support by a DFG-grant, too, which last not least enabled my family to join me here. So, we really could enjoy our stay in Ottawa all together, fortunately having got a very unusual "mild" winter with temperatures only down to -30°C and just a few inches of snow instead of a couple of feet. However, the warm and kind atmosphere in the stochastic section of the Department

(and not only there) made me forget about the cold temperatures outside and let me feel like a part of it. My special thanks are due to Professor Miklós Csörgő who acted as a host in every respect and also stimulated a great deal of the research work I was lucky to do during this period. Last not least it is my pleasure to thank Mrs. Gill Murray who did an outstanding job in typing my manuscripts here. The time of our visit really passed over too fast, but it has been an impressive experience for all of us.

Josef Steinebach

INTRODUCTION

For a short motivation of our present subject consider the following situation: Let $X_1, X_2, \dots$ be an infinite sequence of independent, identically distributed (i.i.d.) random variables (r.v.'s) on some probability space $(\Omega, \mathcal{A}, P_\theta)$, taking values in a measurable space $(X, \mathcal{B})$. The probability measure may depend about some parameter $\theta \in \Theta$. Suppose that $T_n = T_n(X_1, \dots, X_n)$ denotes a statistic estimating a parameter $\gamma = \gamma(\theta)$ in the following sense:

$$(1) \quad \alpha_n(\epsilon, \theta) = P_\theta(|T_n - \gamma(\theta)| \geq \epsilon) \rightarrow 0 \quad (n \rightarrow \infty)$$

for all $\epsilon > 0$ and $\theta \in \Theta$, i.e. T_n is a weakly consistent estimate of γ .

Example. Let $X_1, X_2, \dots$ be i.i.d. and $N(\theta, 1)$ -distributed, $\theta \in \mathbb{R}$.

Then, by the law of large numbers,

$$P_\theta(|\bar{X}_n - \theta| \geq \epsilon) \rightarrow 0 \quad (n \rightarrow \infty),$$

where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, and, moreover, the probabilities are not depending

on θ . But also

$$P_\theta(|M_n - \theta| \geq \epsilon) \rightarrow 0 \quad (n \rightarrow \infty),$$

where $M_n = \text{med}(X_1, \dots, X_n)$, and the question is, what estimate is preferable.

Sometimes there is even a whole class of estimators T_n satisfying (1) and the statistician would like to select the best (if any) in a

certain sense. Some concepts of efficiency have been introduced taking advantage of an exponential behaviour of probabilities of kind $\alpha_n(\varepsilon, \theta)$, which might be as follows

$$(2) \quad \lim_{n \rightarrow \infty} \alpha_n^{1/n}(\varepsilon, \theta) = \rho(\varepsilon, \theta) \quad ,$$

where $0 < \rho(\varepsilon, \theta) < 1$, sometimes depending on ε only. Then the estimator with "smaller" ρ in (2) should be preferred (if for two estimators the ρ 's are comparable at all), because it yields the same precision of the estimate much faster. Such considerations led to notions of A.R.E.'s of test statistics and estimators, the most famous having been developed by Bahadur, Hodges and Lehmann or Chernoff.

In Chapter 1 we collect some methods and results for dealing with exponential behaviour of large deviation probabilities under i.i.d. situations as well as more general ones. Necessary conditions are also mentioned. Chapter 2 reports on certain applications to A.R.E.'s of test statistics and estimators and is mainly taken from Bahadur's (1971) nice monograph. In a final Chapter 3 we treat a relatively new aspect of exponentially behaving large deviation probabilities dealing with fluctuations of random processes. Starting point is the Erdős-Rényi (1970) new law of large numbers which also provides a new solution to Rényi's (1962) so-called "Stochastic geyser problem". Some extensions and general versions are stated including convergence rate statements for such results. For the sake of completeness most proofs are given or at least outlined in a short way.

1. Exponential convergence rates of large deviation probabilities

The main tool in Chapter 1 is use of moment-generating functions and "conjugate" ("associated") distributions or variables defined via m.g.f.'s. So let us first collect some main facts.

1.1. Moment-generating functions

Let X be a real-valued r.v. on $(\Omega, \mathcal{A}, P)$ with m.g.f.

$$\phi(t) = E \exp(tX) = \int \exp(tX) dP < \infty, \quad t \in (t_0, t_1)$$

and cumulant-generating function (c.g.f.)

$$\psi(t) = \log \phi(t), \quad t \in (t_0, t_1).$$

We have the following

Lemma 1.1. (a) If $P(X = 0) < 1$, then ϕ is strictly convex on (t_0, t_1) .

(b) ϕ is infinitely often (i.o.) differentiable in (t_0, t_1) , especially, if $0 \in (t_0, t_1)$,

$$\phi^{(k)}(0) = EX^k, \quad k = 0, 1, \dots$$

(c) If X is non-degenerate, i.e. $P(X = a) < 1$ for all real a , then also ψ is strictly convex on (t_0, t_1) .

(d) If $0 \in (t_0, t_1)$, then

$$\psi'(0) = EX, \quad \psi''(0) = \text{Var}(X).$$

Using moment-generating functions Esscher (1932) introduced the concept

of "conjugate" distributions (or measures) which has later been successfully applied by Cramér (1938) to lay the foundations of large deviation theory.

Definition 1.1. (a) The family $\{P_t\}_{t \in (t_0, t_1)}$ of probability measures, defined by

$$P_t(A) = \int_A \frac{\exp(tX)}{\phi(t)} dP, \quad A \in \mathcal{A}$$

is said to be conjugate to (associated with) P (for a given r.v. X on $(\Omega, \mathcal{A}, P)$). We use the notation:

$$dP_t = \frac{\exp(tX)}{\phi(t)} dP, \quad t \in (t_0, t_1).$$

Sometimes it turns out being useful to have a family of r.v.'s whose distributions are determined by $\{P_t\}$. So, we add a second part to the above definition:

Definition 1.1. (b) The family $\{X(t)\}_{t \in (t_0, t_1)}$ of r.v.'s on $(\Omega, \mathcal{A}, P)$ is said to be conjugate to (associated with) X , if

$$P(X(t) \leq x) = P_t(X \leq x) = \int_{\{X \leq x\}} \frac{\exp(tX)}{\phi(t)} dP, \quad t \in (t_0, t_1).$$

Thus, two equivalent possibilities are provided for operating on $(\Omega, \mathcal{A})$: either use X and $\{P_t\}_{t \in (t_0, t_1)}$ or $\{X(t)\}_{t \in (t_0, t_1)}$ together with P . We shall take advantage of both of them depending on the situation. Some useful facts are collected next (see e.g. Bahadur (1971), Heathcote (1967)):

Lemma 1.2. (a) For $h \in (t_0, t_1)$ fixed, the m.g.f. of $X(h)$ is given by

$$\phi_h(t) = E_P \exp(tX(h)) = E_{P_h} \exp(tX) = \frac{\phi(t+h)}{\phi(h)}, \quad t \in (t_0-h, t_1-h).$$

(b) Using $\psi_h(t) = \log \phi(t+h) - \log \phi(h)$, we have

$$EX(h) = \psi'_h(0) = \psi'(h) = \frac{\phi'(h)}{\phi(h)},$$

$$\text{Var}(X(h)) = \psi''_h(0) = \psi''(h) = \frac{\phi''(h)\phi(h) - (\phi'(h))^2}{\phi^2(h)}.$$

(c) The following "inversion formula" holds:

$$P(X \in B) = \phi(t) \int_{\{X \in B\}} \exp(-tX) dP_t,$$

$t \in (t_0, t_1)$, B measurable.

(d) Let $X_1, \dots, X_n$ be n independent copies of X and $S_n = X_1 + \dots + X_n$. Then n independent copies $X_1(t), \dots, X_n(t)$ of $X(t)$ can be found such that, in distribution,

$$S_n(t) \stackrel{D}{=} X_1(t) + \dots + X_n(t), \quad t \in (t_0, t_1).$$

From Lemma 1.2 we derive the first result about exponential convergence rates of large deviation probabilities.

1.2. Chernoff's Theorem.

Theorem 1.1 (Chernoff, 1952) Consider an i.i.d. sequence $X_1, X_2, \dots$ of non-degenerate r.v.'s on $(\Omega, \mathcal{A}, P)$ satisfying

$$(i) \quad \phi(t) = E \exp(tX_1) < \infty, \quad t \in [0, t_1) .$$

Then, for $a \in A = \{\phi'(t)/\phi(t) : t \in (0, t_1)\}$, we have

$$(1.1) \quad P(S_n \geq na) \leq \rho^n(a) ,$$

where $S_n = X_1 + \dots + X_n$ and $\rho(a) = \inf_t \phi(t) \exp(-ta)$, and, moreover,

$$(1.2) \quad \lim_{n \rightarrow \infty} [P(S_n \geq na)]^{1/n} = \rho(a) .$$

Remark 1.1. (a) If t_1 in (i) has been chosen as large as possible, $A = \{\phi'(t)/\phi(t) : t \in (0, t_1)\} = (EX_1, \text{esssup } X_1)$.

(b) In (1.1) and (1.2) as well, the symbol ' $\geq$ ' can be replaced by ' $>$ '.

(c) By strict convexity properties of $\psi(t) = \log \phi(t)$,

$$\rho(a) = \phi(h) \exp(-ha) ,$$

where h is the unique solution of $a = \phi'(h)/\phi(h)$. Hence

$$0 < \rho(a) < 1 \quad \text{for } a \in A .$$

(d) In (1.2), $P(S_n \geq na)$ can be replaced by $P(S_n \geq na_n)$, whenever $a_n \rightarrow a$ ($n \rightarrow \infty$).

(e) If $\phi(t) < \infty$ for $|t| < t_1$, a two-sided large deviation result can immediately be derived from (1.2), i.e.

$$(1.3) \quad \lim_{n \rightarrow \infty} [P(|\frac{S_n}{n} - EX_1| \geq \epsilon)]^{1/n} = \max(\rho_+(\epsilon), \rho_-(\epsilon)) ,$$

for $\epsilon > 0$ such that $EX_1 \pm \epsilon \in A$, where

$$\rho_+(\epsilon) = \inf_t \phi(t) \exp \{-t(EX_1 + \epsilon)\} ,$$

$$\rho_-(\epsilon) = \inf_t \phi(t) \exp \{-t(EX_1 - \epsilon)\} .$$

Let us now turn to the proof of Chernoff's theorem which differs from the original one and is much simpler.

Proof of Theorem 1.1. (a) Upper bound: Since the solution h of $a = \phi'(h)/\phi(h)$ is positive, we have

$$P(S_n \geq na) \leq \int_{\{h(S_n - na) \geq 0\}} \exp\{h(S_n - na)\} dP \leq \phi^n(h) \exp(-hna) = \rho^n(a) ,$$

which proves (1.1) and, moreover, yields

$$(1.4) \quad \limsup_{n \rightarrow \infty} [P(S_n \geq na)]^{1/n} \leq \rho(a) .$$

(b) Lower bound: Using the inversion formula in Lemma 1.2c for $S_n = X_1 + \dots + X_n$, we obtain, for each $b > a$ and $h < t < t_1$,

$$\begin{aligned} (1.5) \quad P(S_n \geq na) &= \phi^n(t) \int_{\{S_n \geq na\}} \exp(-tS_n) dP_{t,n} \\ &\geq \phi^n(t) \int_{\{na \leq S_n \leq nb\}} \exp(-tS_n) dP_{t,n} \\ &\geq \phi^n(t) \exp(-tnb) P_{t,n}(na \leq S_n \leq nb) . \end{aligned}$$

If t is close to h such that

$$a = \frac{\phi'(h)}{\phi(h)} < EX_1(t) = \frac{\phi'(t)}{\phi(t)} < b ,$$

it follows, from Lemma 1.2d and the law of large numbers,

$$(1.6) \quad P_{t,n} \left(a \leq \frac{S_n}{n} \leq b \right) = P \left(a \leq \frac{S_n(t)}{n} \leq b \right) \rightarrow 1 \quad (n \rightarrow \infty) .$$

Using (1.5) and (1.6) together, we have

$$\liminf_{n \rightarrow \infty} [P(S_n \geq na)]^{1/n} \geq \phi(t) \exp(-tb) ,$$

$t > h$, $b > a$. Letting first $t \rightarrow h$ and then $b \rightarrow a$, this yields

$$(1.7) \quad \liminf_{n \rightarrow \infty} [P(S_n \geq na)]^{1/n} \geq \phi(h) \exp(-ha) = \rho(a) ,$$

which completes the proof. From our derivation of (1.4) and (1.7) it is also obvious that Remark 1.1 (d) holds true.

Remark 1.2. From Chernoff's theorem we get for $\varepsilon > 0$ and n sufficiently large

$$\alpha_n = (\rho(a) - \varepsilon)^n \leq P(S_n \geq na) \leq \rho^n(a) = \beta_n ,$$

but we observe that the bounds are not asymptotically equivalent, since

$$\beta_n / \alpha_n \rightarrow \infty \quad (n \rightarrow \infty) .$$

Under certain situations, however, better bounds are available. Consider for instance, the following

Example 1.1. Let $X_1, X_2, \dots$ be i.i.d. and $N(0,1)$ -distributed. From Feller's (1968, p.166) inequalities

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \frac{1}{(2\pi)^{1/2}} e^{-x^2/2} < 1 - \Phi(x) < \frac{1}{x} \frac{1}{(2\pi)^{1/2}} e^{-x^2/2} \quad (x > 0)$$

for the tail of a standard normal distribution function we obtain, inserting $x = n^{1/2}a$ ($a > 0$),

$$\alpha'_n = \left(\frac{1}{n^{1/2}a} - \frac{1}{(n^{1/2}a)^3}\right) \frac{1}{(2\pi)^{1/2}} e^{-na^2/2} < P(S_n \geq na)$$

$$\frac{1}{n^{1/2}a} \frac{1}{(2\pi)^{1/2}} e^{-na^2/2} = \beta'_n,$$

which implies

$$(1.8) \quad P(S_n \geq na) \sim \frac{1}{(2\pi n)^{1/2}a} (e^{-a^2/2})^n.$$

Checking out that, in the $N(0,1)$ -case, $\rho(a) = \exp(-a^2/2)$ ($a > 0$),

(1.8) means a considerable improvement compared to the general assertion

(1.2). This may raise the question, whether or not an improvement of

Theorem 1.1 might be available generally. A positive answer has been

given by Bahadur and Ranga Rao (1960), which result we state here

without proof:

Theorem 1.2 (Bahadur and Ranga Rao, 1960): Under the conditions of

Theorem 1.1 we have

$$(1.9) \quad P(S_n \geq na) = \frac{1}{(2\pi n)^{1/2}} \rho^n(a) b_n (1 + o(1)),$$

where the b_n 's are independent of n unless X is lattice with $P(X = a) = 0$.

Remark 1.3. The lattice case has been treated earlier by Blackwell and Hodges (1959). Roughly spoken, in both cases (lattice and non-lattice as well) the improved version is made possible by using some expansions in the central limit theorem instead of the law of large numbers, used in (1.6), for the conjugate variables.

However, since use of moment-generating functions via conjugate variables is the main tool in the proof of Theorem 1.1, the question is, what happens if m.g.f.'s do not exist. We shall see that in such situations exponential large deviation rates are no longer available.

1.3 Necessity of moment-generating functions (i.i.d. case)

First results about necessary and sufficient conditions for exponential convergence rates in the law of large numbers have been stated by Baum, Katz and Read (1962), who treated the two-sided case. We are going to prove a one-sided result which is due to Petrov and Širokova (1973).

Theorem 1.3 (Petrov and Širokova, 1973). Let $X_1, X_2, \dots$ be an i.i.d. sequence on $(\Omega, \mathcal{A}, P)$ satisfying

$$(i) \quad P(S_n \geq na) \leq A\rho^n \quad (n = 1, 2, \dots)$$

for some constants a, A and ρ ($0 \leq \rho < 1$). Then there exists some

$t_1 > 0$ such that

$$(1.10) \quad \phi(t) = E \exp(tX_1) < \infty \quad \text{for all } t \in [0, t_1) .$$

Proof. For $a' > 0$ we estimate

$$\begin{aligned} P(S_n \geq na) &\geq P(X_1 \geq n(a+a'), X_2 \geq -a', \dots, X_n \geq -a') \\ &= P(X_1 \geq n(a+a')) [P(X_2 \geq -a')]^{n-1} . \end{aligned}$$

Setting $a_1 = a + a'$ we get

$$P(X_1 \geq na_1) \leq A \rho^n / \rho'^{n-1} = A \rho_1^n ,$$

where $\rho' = P(X_2 \geq -a')$, $A_1 = A\rho'$, $\rho_1 = \rho/\rho'$. Now, a' can be fixed in such a way that

$$\rho_1 = \rho/\rho' < 1 .$$

If, without loss of generality, $a_1 > 0$ we further estimate for $t > 0$

$$\begin{aligned} \phi(t) = E \exp(tX_1) &= \int_{\{X_1 < 0\}} \exp(tX_1) dP + \sum_{n=0}^{\infty} \int_{\{na_1 \leq X_1 < (n+1)a_1\}} \exp(tX_1) dP \\ &\leq P(X_1 < 0) + \sum_{n=0}^{\infty} e^{(n+1)ta_1} P(X_1 \geq na_1) \\ &\leq P(X_1 < 0) + A_1 e^{ta_1} \sum_{n=0}^{\infty} (e^{ta_1} \rho_1)^n . \end{aligned}$$

Hence, choosing $t_1 > 0$ such that $e^{t_1 a_1} \rho_1 = 1$, the latter series is finite for $t < t_1$, which completes the proof.

The following simple corollary is used later on:

Corollary 1.1. Let $X_1, X_2, \dots$ be an i.i.d. sequence satisfying

$$(i) \quad \phi(t) = \infty \text{ for all } t > 0 .$$

Then, for all constants a and ρ ($0 < \rho < 1$),

$$(1.11) \quad \limsup_{n \rightarrow \infty} \frac{P(S_n \geq na)}{\rho^n} = \infty .$$

Remark 1.4. Baum, Katz and Read (1962) also treated the case of independent but not necessarily identically distributed random variables. Then, under certain conditions, finiteness of the corresponding m.g.f.'s is available on some common interval. But examples show that in the non-independence case the close connection between exponential convergence rates and existence of m.g.f.'s gets lost.

However, we shall see in our next section that for general sequences at least sufficient conditions can be imposed to retain an exponential large deviation behaviour.

1.4. General Sequences

Sievers (1969) seems to have been the first who generally derived exponential properties for large deviation probabilities from a certain exponential behaviour of the corresponding m.g.f.'s via the inversion formula in Lemma 1.2. But most of Sievers' assumptions turned out to be only technical ones and thus could be removed (see Plachky (1971), Plachky and Steinebach (1975)). We state a recent version given in Steinebach (1976):

Theorem 1.4 (Sievers-Plachky-Steinebach). Let $\{W_n\}_{n=1,2,\dots}$ be a sequence of real-valued r.v.'s on (Ω, A, P) satisfying

- (i) $\phi_n(t) = E \exp(tW_n) < \infty$ for $t \in [0, t_1]$;
- (ii) $\lim_{n \rightarrow \infty} \frac{1}{n} \psi_n(t) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \phi_n(t) = c(t)$ exists on some subinterval (t_0, t_1) ;
- (iii) c is strictly convex on (t_0, t_1) .

Then, for $a_n \rightarrow a \in \{c'_+(t) : t \in (t_0, t_1)\}$, we have

$$(1.12) \quad \lim_{n \rightarrow \infty} [P(W_n \geq na_n)]^{1/n} = \inf_t \exp \{c(t) - ta\} \\ = \exp \{c(h) - ha\} = \rho(a),$$

where h is the unique solution of $a = c'_+(h)$ and $c'_+(h)$ denotes the right derivative of the convex function c at h .

The proof follows the lines of our proof of Theorem 1.1, deriving upper and lower bounds for the large deviation probabilities separately.

Proof of Theorem 1.4. a) Upper bound: Take the unique positive solution h of $a = c'_+(h)$ and estimate

$$P(W_n \geq na_n) \leq \int_{\{h(W_n - na_n) \geq 0\}} \exp\{h(W_n - na_n)\} dP \leq \phi_n(h) \exp(-hna_n),$$

which yields, using (ii) and $a_n \rightarrow a$,

$$\limsup_{n \rightarrow \infty} [P(W_n \geq na_n)]^{1/n} \leq \exp \{c(h) - ha\} = \rho(a).$$

b) Lower bound: For each $b > a$ and $t > h$, same estimations, used to derive (1.5) in the proof of Chernoff's theorem, yield

$$(1.13) \quad P(W_n \geq na_n) \geq \phi_n(t) \exp(-tnb) P_{t,n}(na_n \leq W_n \leq nb),$$

where now

$$P_{t,n}(A) = \int_A \frac{\exp(tW_n)}{\phi_n(t)} dP, \quad A \in \mathcal{A}.$$

Using some convex analysis, it can be shown again that

$$(1.14) \quad \lim_{n \rightarrow \infty} P_{t,n}(na_n \leq W_n \leq nb) = 1,$$

whenever t is chosen such that $a < c'_+(t) < b$, which is possible since c'_+ is strictly increasing (for details see [37], pp.3-4). To give a motivation for (1.14) note that

$$\begin{aligned} \limsup_{n \rightarrow \infty} E_{P_{t,n}} \left(\frac{W_n}{n} \right) &= \limsup_{n \rightarrow \infty} \frac{1}{n} \psi'_n(t) \leq c'_+(t), \\ \liminf_{n \rightarrow \infty} E_{P_{t,n}} \left(\frac{W_n}{n} \right) &= \liminf_{n \rightarrow \infty} \frac{1}{n} \psi'_n(t) \geq c'_-(t), \end{aligned}$$

where $c'_-(t)$ denotes the left-hand derivative of c at t .

Combining (1.13) and (1.14) we get

$$\liminf_{n \rightarrow \infty} [P(W_n \geq na_n)]^{1/n} \geq \exp \{c(t) - tb\},$$

$t > h$, $b > a$. Letting again $t \rightarrow h$ and then $b \rightarrow a$, the proof is complete.

To prove Theorem 1.4 the so-called moment-generating function technique has been used. Bahadur (1971, Theorem 2.2) obtained a similar result, also using m.g.f.'s and imposing some "standard conditions" on them. Different techniques have been used by several authors to establish exponential large deviation results in other situations. For an extensive list of references see e.g. S. Csörgö (1979).

In the next section we add a few remarks on the necessity of m.g.f.'s for obtaining exponential behaviours under general sequences.

1.5. Necessity of moment-generating functions (general case)

Baum, Katz and Read (1962) construct a Markov-sequence $X_1, X_2, \dots$ satisfying

$$E \exp(tX_i) = \infty \quad \text{for } t \neq 0, i = 1, 2, \dots, \text{ but}$$

$$\lim_{n \rightarrow \infty} [P(S_n \neq 0)]^{1/n} = \frac{1}{2}.$$

Hence exponential rates are possible for partial sum sequences without finiteness of any moment-generating function.

On the other hand, if m.g.f.'s exist, they need not behave exponentially in order to obtain exponential large deviation rates.

Lynch (1978) gives an example of a sequence $\{W_n\}_{n=1,2,\dots}$ satisfying

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log E \exp(tW_n) = \infty \quad \text{for all } t \neq 0, \text{ but}$$

$$\lim_{n \rightarrow \infty} [P(W_n \geq na)]^{1/n} = \exp(-a^2/2), \quad a > 0.$$

Thus, condition (ii) of Theorem 1.4 is sufficient but not necessary.

Lynch, however, stated "a curious converse of Sievers' theorem"

(as he called it) giving some kind of a necessary condition. We mention his interesting result without proof:

Theorem 1.5 (Lynch, 1978). Let $\{W_n\}_{n=1,2,\dots}$ be a sequence on (Ω, A, P) such that $W_n/n \xrightarrow{P} 0$ ($n \rightarrow \infty$). Suppose that

$$(i) \quad \lim_{n \rightarrow \infty} [P(W_n \geq na)]^{1/n} = \rho(a)$$

exists for $a \in [0, a_0]$, where $\rho(a_0) > 0$ ($a_0 > 0$). Put

$$c(t) = \sup \{ta + \log \rho(a) : a \in [0, a_0]\}$$

and define

$$V_n = \begin{cases} W_n & , \text{ if } W_n \leq na_0 \\ na_0 & , \text{ if } W_n > na_0 \end{cases}$$

Setting $\psi_n(t) = \log E \exp(tV_n)$, we then have

$$(1.15) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \psi_n(t) = c(t) \quad \text{for all } t > 0.$$

1.6. Some recent results.

Without claiming completeness let us mention a few recent results about exponential convergence rates for large deviation probabilities.

Bárfai (1977) extended the Chernoff theorem to sequences $X_1, X_2, \dots$ of i.i.d. r.v.'s taking values in $\mathbb{R}^k$ and possessing finite m.g.f.'s $\phi(t) = E \exp \langle t, X_1 \rangle$ for t 's in some neighborhood of the

origin. The results are

$$(1.16) \quad P\left(\frac{S_n}{n} \in A\right) \leq \rho^n(A)$$

for convex sets A , where $\rho(A) = \sup_{a \in A} \rho(a)$ and $\rho(a) = \inf_t \phi(t) \exp(-\langle t, a \rangle)$

Moreover,

$$(1.17) \quad \lim_{n \rightarrow \infty} [P\left(\frac{S_n}{n} \in A\right)]^{1/n} = \rho(A),$$

if A is convex and open.

In [39] we extended Theorem 1.4 to sequences $\{W_n\}_{n=1,2,\dots}$ of random vectors in $\mathbb{R}^k$ whose m.g.f.'s satisfy certain conditions. For special sets A , i.e. convex or complements of bounded sets in $\mathbb{R}^k$, we proved that

$$(1.18) \quad \lim_{n \rightarrow \infty} [P\left(\frac{W_n}{n} \in A\right)]^{1/n} = \rho(A),$$

the value $\rho(A)$ again being well-determined.

Several authors (see e.g. [4], [11], [18]) dealt with large deviations of partial sums of an i.i.d. sequence $X_1, X_2, \dots$ taking values in some topological vector space V and using the moment-generating functional $\phi(t) = E \exp \{t(X_1)\}$, $t \in D$, where D denotes the dual space of all linear functionals $t = t(v)$ on V . The results are either of the type

$$(1.19) \quad \lim_{n \rightarrow \infty} [P\left(\frac{S_n}{n} \in A\right)]^{1/n} = \rho(A),$$

where A is convex and open (see [11]), or yield upper and lower bounds, i.e.

$$(1.20) \quad \limsup_{n \rightarrow \infty} [P(\frac{S_n}{n} \in A)]^{1/n} \leq \rho(A) ,$$

for closed sets A , or

$$(1.21) \quad \liminf_{n \rightarrow \infty} [P(\frac{S_n}{n} \in B)]^{1/n} \geq \rho(B)$$

for open sets B (see [4], [18]).

Finally, we mention a recent result of Groeneboom, Oosterhoff and Ruymgaart (1979) dealing with large deviation probabilities for functionals T of empirical probability measures $\hat{P}_n$ based upon an i.i.d. sequence $X_1, X_2, \dots$, again taking values in some general topological space. The essential result is

$$(1.22) \quad \lim_{n \rightarrow \infty} [P(T(\hat{P}_n) \geq a_n)]^{1/n} = \rho(a) ,$$

where $a_n \rightarrow a$ and $\rho(a)$ depends on certain Kullback-Leibler information numbers.

For a nice collection of the recent literature about large deviations we call again attention to S. Csörgö's (1979) paper. For a list of references up to 1976 see also [37].

2. Asymptotic relative efficiencies (A.R.E.'s) of test statistics and estimators

Let us now turn to some statistical applications of exponential convergence rate results for certain large deviation probabilities. As already mentioned, notions of efficiency have been based on such properties by several authors (e.g. Bahadur, Chernoff, Hodges and Lehmann). A main tool is use of Kullback-Leibler information numbers which have a definite relationship to certain exponential large deviation rates. So, let us first collect some basic facts about those numbers. These, as most parts of our second chapter, have been taken from Bahadur's (1971) monograph.

2.1. Kullback-Leibler information numbers

Let $(X, \mathcal{B})$ be a measurable space (sample space). Let p and q be probability measures on $\mathcal{B}$. If $q \ll p$ on $\mathcal{B}$, let $r(x)$ be a density, $0 \leq r(x) < \infty$, i.e. $dq = r(x)dp$ on $\mathcal{B}$.

Definition 2.1. The value

$$(2.1) \quad K(q, p) = \begin{cases} \int \log r(x) dq, & \text{if } q \ll p, \\ \infty, & \text{else,} \end{cases}$$

is called Kullback-Leibler information number of q with respect to p . Some basic properties of K are collected in

Lemma 2.1. a) K is well-defined.

b) $0 \leq K \leq \infty$. Moreover, $K = 0$ if and only if $p(B) = q(B)$ for all $B \in \mathcal{B}$.

c) If $q \ll p$, we also have

$$(2.2) \quad K(q,p) = \int r(x) \log r(x) dp .$$

d) Let $q \ll p$ and $p \ll q$. Then $0 < r(x) < \infty$. Setting $f(t) = \int \exp \{t \log r(x)\} dp = \int \{r(x)\}^t dp$, we have

$$(2.3) \quad K(q,p) = f'(1-) , \quad -K(p,q) = f'(0+) .$$

Note that $f(0) = f(1) = 1$, and $f(t) < \infty$ for all $t \in [0,1]$.

Outline of proof. a) This follows from the fact that $r(x) > 0$ [q], and, if $r_1(x), r_2(x)$ are two p-densities of q, $r_1(x) = r_2(x)$ [p], which also means $r_1(x) = r_2(x)$ [q].

b) Using inequality $x \log x \geq x-1$, with equality only if $x = 1$, we obtain in case $q \ll p$

$$\begin{aligned} K(q,p) &= \int \log r(x) dq = \int r(x) \log r(x) dp \\ &\geq \int (r(x)-1) dp = q(X) - p(X) = 0 . \end{aligned}$$

c) follows from the definition of densities.

d) follows from properties of m.g.f.'s, since $f(t)$ is nothing but the m.g.f. of $\log r(x)$ with respect to p.

Remark 2.1. In case of p and q having densities with respect to some σ -finite measure μ , i.e. $dp = f_1 d\mu$, $dq = f_2 d\mu$, and $0 < f_1(x), f_2(x) < \infty$ [μ], we get the special form

$$(2.4) \quad K(q,p) = \int \log \left(\frac{f_2(x)}{f_1(x)} \right) f_2(x) d\mu ,$$

which is also denoted $K(f_2, f_1)$ occasionally.

Example 2.1. Let $(X, B) = (\mathbb{R}^k, B^k)$ and $p = f_1(x)\lambda^k$, $q = f_2(x)\lambda^k$, where

$$f_i(x) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)' \Sigma^{-1} (x - \mu_i) \right\}, \quad i=1,2,$$

$\mu_i \in \mathbb{R}^k$, Σ being a positive-definite matrix. Since

$$\log \left(\frac{f_2(x)}{f_1(x)} \right) = \frac{1}{2} (\mu_1' \mu_1 - 2(\mu_1' - \mu_2') \Sigma^{-1} x - \mu_2' \mu_2),$$

we obtain

$$K(q, p) = K(f_2, f_1) = \frac{1}{2} (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2).$$

Suppose now, that $Y = Y(x)$ is a real-valued measurable function on (X, B) , and that p is a given probability measure on B . Let

$$\phi(t) = \int \exp(tY(x)) dp, \quad \text{and}$$

$$\rho(a) = \inf_{t \geq 0} \phi(t) \exp(-ta).$$

Define a set A of probability measures q on B by

$$A = \{q : \int Y(x) dq \text{ exists and } \geq a\},$$

a being fixed. Setting

$$(2.5) \quad K(A, p) = \begin{cases} \inf \{K(q, p) : q \in A\}, & \text{if } A \neq \emptyset, \\ +\infty, & \text{if } A = \emptyset, \end{cases}$$

we have the following reformulation of the exponential rate $\rho(a)$ appearing in Chernoff's Theorem:

Theorem 2.1 (Hoeffding, 1965). $\rho(a) = \exp [-K(A,p)]$

Proof. a) We first prove

$$(2.6) \quad K(A,p) \geq -\log \rho(a) .$$

If $K(A,p) = \infty$ or $\rho(a) = 1$, (2.6) holds trivially. Hence suppose $K(A,p) < \infty$ and $\rho(a) < 1$. Then $q \in A$ and $t > 0$ exist such that $dq = r(x)dp$ and $\phi(t) < \infty$. We estimate

$$\begin{aligned} \phi(t) &= \int \exp(tY(x))dp \geq \int_{\{r(x) > 0\}} \exp(tY(x))dp \\ &= \int_{\{r(x) > 0\}} \exp(tY(x) - \log r(x))r(x)dp \\ &= \int_X \exp(tY(x) - \log r(x))dq \\ &\geq \exp(ta - K(q,p)) , \end{aligned}$$

where Jensen's inequality and the definition of A and K have been used in the last line. Thus

$$K(q,p) \geq \sup_{t \geq 0} \{ta - \log \phi(t)\} = -\log \rho(a) .$$

Taking $\inf_{q \in A}$ on the left-hand side, (2.6) is proved.

b) To prove

$$(2.7) \quad K(A,p) \leq -\log \rho(a) ,$$

let us first consider the case $\phi(t) < \infty$ for some $t > 0$. If

$\int Y(x)dp \geq a$, then $p \in A$ and $\rho(a) = 1$. Since $K(p,p) = 0$, (2.7) holds in this case.

If $\int Y dp < a < a_1 = \text{esssup } Y \text{ (under } p)$, let $h > 0$ be such that

$$a = \phi'(h)/\phi(h). \text{ Put } dq = \frac{\exp(hY(x))}{\phi(h)} dp. \text{ Then } \int Y(x) dq = \frac{\phi'(h)}{\phi(h)} = a,$$

hence $q \in A$. Moreover

$$K(q, p) = \int (hY(x) - \log \phi(h)) dq = ha - \log \phi(h) = -\log \rho(a),$$

which implies (2.7).

If $a_1 = \text{esssup } Y \text{ (under } p) < \infty$, note that $\rho(a) = 0$ for $a > a_1$ and $\rho(a_1) = p(Y = a_1)$.

If $\rho(a) = 0$, (2.7) holds trivially. Hence suppose $p(Y = a_1) > 0$ and consider the case $a = a_1$. Setting $B = \{Y = a_1\}$ and $r(x) = 1/p(B)$ ($x \in B$), $r(x) = 0$ ($x \in X \setminus B$). Let $dq = r(x)dp$. Then $\int Y(x) dq = a_1 \frac{1}{p(B)} p(B) = a$, hence $q \in A$. On the other hand

$$K(q, p) = \log\left(\frac{1}{p(B)}\right) \frac{1}{p(B)} p(B) = -\log p(Y = a_1) = -\log \rho(a_1).$$

Taking $\inf$ over A again, the proof is complete.

To treat the general case, let

$$Y_k(x) = \begin{cases} Y(x), & \text{if } Y(x) \leq k, \\ k, & \text{else,} \end{cases}$$

$k = 1, 2, \dots$. Then $\phi_k(t) = \int \exp(tY_k(x)) dp < \infty$ for some $t > 0$.

Hence our above considerations yield

$$K(A_k, p) \leq -\log \rho_k(a),$$

where $A_k = \{q : \int Y_k(x) dq \geq a\}$, $\rho_k(a) = \inf_{t \geq 0} \phi_k(t) \exp(-ta)$.

Since $Y_k(x) \geq Y(x)$, we have $A_k \subset A$, which implies

$$K(A_k, p) \geq K(A, p) .$$

Using $\rho_k(a) \rightarrow \rho(a)$ ($k \rightarrow \infty$) (see Bahadur (1971, Lemma 3.2)), the proof of (2.7) is complete.

Hoeffding's theorem affords the following reformulation of the Chernoff theorem:

Suppose that $\{x\} \in \mathcal{B}$ for all $x \in X$. Let $Y(x)$ be a real-valued measurable function on $(X, \mathcal{B})$ and p be a given probability measure on $\mathcal{B}$. Let $K(A, p)$ defined in (2.5) be finite. Consider an i.i.d. sequence $X_1, X_2, \dots$ taking values in X and suppose $P_{X_1} = p$. Define the *empirical probability measure* $\hat{p}_n$ on $\mathcal{B}$ by

$$(2.8) \quad \hat{p}_n(B) = \frac{1}{n} \#\{X_i : i \in \{1, \dots, n\}, X_i \in B\} , \quad B \in \mathcal{B} .$$

Then

$$\int Y(x) d\hat{p}_n = \frac{1}{n} \sum_{i=1}^n Y(X_i)$$

and

$$\hat{p}_n \in A \quad \text{if and only if} \quad \sum_{i=1}^n Y(X_i) \geq na .$$

In view of Theorem 2.1, assertion (1.2) in Chernoff's theorem is therefore equivalent to

$$(2.9) \quad P(\hat{p}_n \in A | p) = \exp[-nK(A, p) + o(n)] .$$

Results like (2.9) are available for essentially larger classes of sets A than those of the above special form (see Sanov (1957), Hoadley (1967)). A rather general version has recently been obtained by Groeneboom, Oosterhoff and Ruymgaart (1979).

2.2. Stein's Lemma. Large sample behaviour of the Neyman-Pearson test.

Certain optimality properties of test-statistics and estimators can be described in terms of Kullback-Leibler information numbers. We shall start with an unpublished result of Stein, a first published proof seems to be given by Chernoff (1956). We follow the lines of Bahadur (1971) here:

Consider an i.i.d. sequence $X_1, X_2, \dots$ taking values in some measurable sample space $(X, \mathcal{B})$. Let P_θ , $\theta \in \Theta$, denote the (common) distribution of the X_i 's. We use the following notations:

$$X^{(n)} = (X_1, \dots, X_n), \quad X^{(\infty)} = (X_1, X_2, \dots),$$

$$\chi^{(n)} = \prod_{i=1}^n \chi, \quad \chi^{(\infty)} = \prod_{i=1}^{\infty} \chi,$$

$$\mathcal{B}^{(n)} = \bigotimes_{i=1}^n \mathcal{B}, \quad \mathcal{B}^{(\infty)} = \bigotimes_{i=1}^{\infty} \mathcal{B},$$

$$P_\theta^{(n)} = \bigotimes_{i=1}^n P_\theta, \quad P_\theta^{(\infty)} = \bigotimes_{i=1}^{\infty} P_\theta.$$

For $\theta_0, \theta_1 \in \Theta$ denote

$$K(\theta_1, \theta_0) = K(P_{\theta_1}, P_{\theta_0}) .$$

Note that, by independence properties,

$$K(P_{\theta_1}^{(n)}, P_{\theta_0}^{(n)}) = nK(P_{\theta_1}, P_{\theta_0}) = nK(\theta_1, \theta_0) .$$

Now, choose and fix $\theta_0, \theta_1 \in \Theta$ and consider the Neyman-Pearson testing problem

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta = \theta_1$$

(simple hypotheses). Let β be given, $0 < \beta < 1$, and require the test to have power $1-\beta$ against θ_1 , i.e.

$$E_{\theta_1} \phi_n = 1 - \beta ,$$

$\phi_n \in \Phi_n = \{\phi_n : X^{(n)} \rightarrow [0,1] \mid \phi_n \text{ measurable}\}$. For each n , let $\alpha_n(\beta)$ denote the infimum of all available sizes under the stated power requirement, i.e.

$$\alpha_n(\beta) = \inf \{E_{\theta_0} \phi_n : \phi_n \in \Phi_n, E_{\theta_1} \phi_n = 1 - \beta\}.$$

From the Neyman-Pearson-theory it is well-known that a (possibly randomized) test ϕ_n^* exists such that

$$\alpha_n(\beta) = E_{\theta_0} \phi_n^* , \quad E_{\theta_1} \phi_n^* = 1 - \beta .$$

The following theorem, known as Stein's lemma, holds true (see Bahadur (1971, Lemma 6.1)):

Theorem 2.2 (Stein's Lemma). For each β , $0 < \beta < 1$,

$$(2.10) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha_n(\beta) = -K(\theta_1, \theta_0) .$$

Outline of proof. We shall consider four different cases.

Case 1: $K(\theta_1, \theta_0) = 0$. Then $P_{\theta_1} = P_{\theta_0}$ on B , which also implies $P_{\theta_1}^{(n)} = P_{\theta_0}^{(n)}$ for each n . Hence $\alpha_n(\beta) = E_{\theta_0} \phi_n = E_{\theta_1} \phi_n = 1 - \beta$, yielding

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha_n(\beta) = 0 = -K(\theta_1, \theta_0) .$$

Case 2: $0 < K(\theta_1, \theta_0) < \infty$. In this case a density $r(x)$ exists, $0 \leq r(x) < \infty$, such that $dP_{\theta_1} = r(x)dP_{\theta_0}$ on X . We have

$$K(\theta_1, \theta_0) = E_{\theta_1} \log r(x) ,$$

$$0 < r(x) < \infty [P_{\theta_1}] ,$$

$$dP_{\theta_1}^{(n)} = \prod_{i=1}^n r(x_i) dP_{\theta_0}^{(n)} ,$$

$$0 < r_n = \prod_{i=1}^n r(x_i) < \infty [P_{\theta_1}^{(n)}] .$$

Hence constants c_n and δ_n exist, $0 < c_n < \infty$, $0 \leq \delta_n \leq 1$, such that for

$$\phi_n^* = \begin{cases} 1 , & \text{if } r_n > c_n , \\ \delta_n , & \text{if } r_n = c_n , \\ 0 , & \text{if } r_n < c_n , \end{cases}$$

we have

$$E_{\theta_1} \phi_n^* = 1 - \beta .$$

From the Neyman-Pearson theory we then know

$$\alpha_n(\beta) = E_{\theta_0} \phi_n^* .$$

Denote $Y(x) = \log r(x)$ and

$$Y_i = Y(x_i) , \quad u_n = \frac{1}{n} \log c_n .$$

The special structure of ϕ_n^* yields

$$P_{\theta_1} \left(\sum_{i=1}^n Y_i > nu_n \right) \leq 1 - \beta \leq P_{\theta_1} \left(\sum_{i=1}^n Y_i \geq nu_n \right) .$$

Since $0 < 1 - \beta < 1$, and, by the law of large numbers,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_i = E_{\theta_1} Y = K(\theta_1, \theta_0) [P_{\theta_1}] ,$$

it follows

$$\lim_{n \rightarrow \infty} u_n = K(\theta_1, \theta_0) .$$

Now, again from the structure of ϕ_n^* ,

$$P_{\theta_0} \left(\sum_{i=1}^n Y_i > nu_n \right) \leq \alpha_n(\beta) \leq P_{\theta_0} \left(\sum_{i=1}^n Y_i \geq nu_n \right) .$$

Applying Chernoff's theorem yields the desired conclusion, but separate consideration of the case $P_{\theta_0}(Y \leq K(\theta_1, \theta_0)) = 1$ is required. Direct evaluations are also available (see Bahadur (1971, Lemma 6.1)).

Case 3: $P_{\theta_1} \ll P_{\theta_0}$, but $K(\theta_1, \theta_0) = \infty$. In this case use a truncation argument and case 2 again.

Case 4: $P_{\theta_1} \not\ll P_{\theta_0}$. Then a set $B \subset X$ exists such that $P_{\theta_0}(B) = 0$, but $P_{\theta_1}(B) = \gamma > 0$. Define

$$\phi_n(x^{(n)}) = \begin{cases} 1, & \text{if } x_i \in B \text{ for some } i=1, \dots, n; \\ 0, & \text{otherwise.} \end{cases}$$

Size and power of this test are given by

$$\alpha_n = E_{\theta_0} \phi_n \leq \sum_{i=1}^n P_{\theta_0}(B) = 0,$$

$$\pi_n = E_{\theta_1} \phi_n = 1 - [P_{\theta_1}(\bar{B})]^n = 1 - (1-\gamma)^n \rightarrow 1 \quad (n \rightarrow \infty).$$

Hence, for n such that $\pi_n \geq 1-\beta$, put

$$\phi_n^* = (1-\beta)^{\pi_n^{-1}} \phi_n,$$

which yields

$$\alpha_n(\beta) = E_{\theta_0} \phi_n^* = 0, \quad E_{\theta_1} \phi_n^* = 1-\beta,$$

completing the proof.

For an extension of Theorem 2.2 to optimal tests for finite composite hypotheses see e.g. [32].

2.3. Asymptotic effective variances and A.R.E.'s of estimators

In the same framework as in the preceeding section we shall now consider a theory of estimation:

Let γ be a real-valued functional on the parameter space Θ , and, for each n , let $T_n = T_n(X_1, \dots, X_n)$ be a point estimator of γ . For $\epsilon > 0$ and $\theta \in \Theta$, put

$$\alpha_n(\epsilon, \theta) = P_\theta(|T_n - \gamma(\theta)| \geq \epsilon) .$$

Now, if T_n is a consistent estimator, i.e. $\lim_{n \rightarrow \infty} \alpha_n(\epsilon, \theta) = 0$ for all $\epsilon > 0$, define

$$M(\epsilon, \delta, \theta) = \text{smallest sample size } m \text{ such that } \alpha_n(\epsilon, \theta) \leq \delta \\ \text{for all } n \geq m ,$$

$0 < \delta < 1$. In other words, $M(\epsilon, \delta, \theta)$ is the smallest available sample size under which $\gamma(\theta)$ is estimated within a precision smaller than ϵ and with a probability higher than $1-\delta$. In this sense, $M = M(\epsilon, \delta, \theta)$ measures the efficiency of T_n as a point estimate, but having the drawback of depending upon three parameters. So, let us consider some asymptotic properties of $M(\epsilon, \delta, \theta)$, to get rid of one or two of those parameters.

If T_n is asymptotically normal with variance $v(\theta)$, i.e.

$$\lim_{n \rightarrow \infty} P_\theta \left(n^{1/2} \frac{T_n - \gamma(\theta)}{v^{1/2}(\theta)} \leq x \right) = \Phi(x) \text{ for all } x \in \mathbb{R}, \theta \in \Theta ,$$

where ϕ denotes the standard normal d.f., T_n is certainly consistent.

Letting $\epsilon \rightarrow 0$, we obtain

$$(2.11) \quad \lim_{\epsilon \rightarrow 0} M(\epsilon, \delta, \theta) / (v(\theta) \frac{u_{\delta/2}^2}{\epsilon}) = 1,$$

where $1 - \phi(u_{\delta/2}) = \delta/2$. For a motivation of (2.11), notice that, by definition,

$$\alpha_M(\epsilon, \theta) \leq \delta < \alpha_{M-1}(\epsilon, \theta),$$

which implies

$$P_{\theta} \left(M^{\frac{1}{2}} \frac{|T_M - \gamma(\theta)|}{v^{\frac{1}{2}}(\theta)} \geq M^{\frac{1}{2}} \frac{\epsilon}{v^{\frac{1}{2}}(\theta)} \right) \sim \delta \quad (\epsilon \rightarrow 0).$$

By asymptotic normality, the latter relation can only be true if (2.11) holds.

Thus, if $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are two different, asymptotically normal sequences of point estimators for $\gamma(\theta)$,

$$(2.12) \quad e_{1,2}(\theta) = \lim_{\epsilon \rightarrow 0} \frac{M^{(2)}(\epsilon, \delta, \theta)}{M^{(1)}(\epsilon, \delta, \theta)} = \frac{v^{(2)}(\theta)}{v^{(1)}(\theta)}$$

serves as a measure of A.R.E. of $\{T_n^{(1)}\}$ against $\{T_n^{(2)}\}$. $e_{1,2}(\theta) > 1$ means that, under P_{θ} , $\{T_n^{(1)}\}$ is (asymptotically) more efficient than $\{T_n^{(2)}\}$.

Suppose now that $\{T_n\}$ is a consistent, but not necessarily asymptotically normal sequence of point estimators for γ , i.e.

$\alpha_n(\varepsilon, \theta) \rightarrow 0$ ($n \rightarrow \infty$). In typical cases, we have an exponential rate, e.g.

$$(2.13) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha_n(\varepsilon, \theta) = -\frac{1}{2} \kappa(\varepsilon, \theta) ,$$

$0 < \kappa < \infty$. In such a case it follows that

$$(2.14) \quad \lim_{\delta \rightarrow 0} M(\varepsilon, \delta, \theta) / (2 \log (1/\delta) / \kappa(\varepsilon, \theta)) = 1 ,$$

which can immediately be derived from the inequalities

$$\frac{1}{M} \log \alpha_M \leq \frac{1}{M} \log \delta \leq \frac{M-1}{M} \frac{1}{M-1} \log \alpha_{M-1} .$$

Hence, if $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are two different sequences of estimators, both holding (2.13),

$$(2.15) \quad e_{1,2}(\varepsilon, \theta) = \lim_{\delta \rightarrow 0} \frac{M^{(2)}(\varepsilon, \delta, \theta)}{M^{(1)}(\varepsilon, \delta, \theta)} = \frac{\kappa^{(1)}(\varepsilon, \theta)}{\kappa^{(2)}(\varepsilon, \theta)}$$

can serve as a measure of A.R.E. of $\{T_n^{(1)}\}$ against $\{T_n^{(2)}\}$. If there is no interest in a special ε , Bahadur (see e.g. [1]) proposes

$$(2.16) \quad e_{1,2}(\theta) = \lim_{\varepsilon \rightarrow 0} e_{1,2}(\varepsilon, \theta) = \lim_{\varepsilon \rightarrow 0} \frac{\kappa^{(1)}(\varepsilon, \theta)}{\kappa^{(2)}(\varepsilon, \theta)}$$

as an A.R.E. measure.

In the considered framework, Bahadur (1967) also introduced a notion of "effective variances": $\tau_n^2 = \tau_n^2(\varepsilon, \theta)$ is called effective variance of T_n , if

$$(2.17) \quad \alpha_n(\varepsilon, \theta) = P_\theta(|T_n - \gamma(\theta)| \geq \varepsilon) = P(|N(0,1)| \geq \varepsilon/\tau_n) ,$$

i.e. τ_n^2 is the variance of a standard normal estimate which yields the

same precision as T_n with same probability (under P_θ). Using (2.13) and e.g. Feller's (1968) expansion for the normal d.f. (see Example 1.1), we get the following asymptotic result for τ_n^2 :

$$(2.18) \quad \lim_{n \rightarrow \infty} n \tau_n^2(\epsilon, \theta) = \epsilon^2 / \kappa(\epsilon, \theta) .$$

We remark that, under certain assumptions, lower bounds like

$$(2.19) \quad \liminf_{\epsilon \rightarrow 0} \liminf_{n \rightarrow \infty} n \tau_n^2(\epsilon, \theta) \geq v^*(\theta)$$

are available, which may be attained e.g. for certain maximum-likelihood estimators. $\{T_n\}$ is sometimes called an asymptotically efficient sequence of estimators, if "=" holds true in (2.19). The value

$$(2.20) \quad w(\epsilon, \theta) = \epsilon^2 / \kappa(\epsilon, \theta)$$

in (2.18) is called asymptotic effective variance. In fact, under certain conditions, $w(\epsilon, \theta)$ turns out to be a limit of variances. Remember, for instance, Theorem 1.4 in Section 1.4. If $\{T_n\}$ holds the regularity conditions assumed there, in a two-sided sense, the limit in (2.13) could be of the form

$$-\frac{1}{2} \kappa(\epsilon, \theta) = \max \{c_\theta(h_+) - h_+ \epsilon, c_\theta(h_-) + h_- \epsilon\},$$

where $c_\theta(t)$ is strictly convex (and probably differentiable) in some open interval about the origin. Here $c_\theta(0) = 0$, and $h_\pm = h_\pm(\epsilon)$ are solutions of $\epsilon = \pm c'_\theta(h_\pm)$. If c_θ should even be twice differentiable about the origin, we have, by Taylor series expansion,

$$(2.21) \quad \kappa(\varepsilon, \theta) = \frac{\varepsilon^2}{c_{\theta}''(0)} + o(\varepsilon^2) \quad (\varepsilon \rightarrow 0) .$$

Moreover, if

$$c_{\theta}(t) = \lim_{n \rightarrow \infty} \frac{1}{n} \log E \exp (nT_n) , \quad |t| < t_1 ,$$

under some regularity conditions, also

$$c_{\theta}''(t) = \lim_{n \rightarrow \infty} \frac{1}{n} (\log E \exp (nT_n))'' ,$$

in particular,

$$(2.22) \quad c_{\theta}''(0) = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Var}(nT_n) = \lim_{n \rightarrow \infty} \text{Var}(n^{1/2}T_n) .$$

Combining (2.18) and (2.20)-(2.22) , we actually see that asymptotic effective variances turn out to be (asymptotically) the variances of the (standardized) estimators under consideration. Note that, under normal distributions τ_n^2 equals exactly the variance of T_n for each ε . Under (2.21), we also obtain

$$(2.23) \quad e_{1,2}(\theta) = \lim_{\varepsilon \rightarrow 0} \frac{\kappa^{(1)}(\varepsilon, \theta)}{\kappa^{(2)}(\varepsilon, \theta)} = \frac{c_{\theta}^{(2)''}(0)}{c_{\theta}^{(1)''}(0)} ,$$

which, by (2.22) , might be a ratio of (asymptotic) variances again.

Hence (2.23) is very similar to (2.12) stated under asymptotically normal distributions.

2.4. Bahadur slopes and A.R.E.'s of test statistics

Let us next consider a theory of testing hypotheses $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0$ in terms of the framework stated before. Let $T_n = T_n(X_1, \dots, X_n)$ be a suitable (possibly standardized) test statistic and suppose large values of T_n being significant. For the sake of simplicity, assume that T_n has a null distribution F_n , i.e.

$$P_\theta(T_n \leq x) = F_n(x) \quad (x \in \mathbb{R}),$$

for all $\theta \in \Theta_0$. Define the level attained by T_n to be the random variable

$$L_n = 1 - F_n(T_n).$$

L_n is an equivalent test statistic, since H_0 is to be rejected at level α if and only if $L_n \leq \alpha$ ($0 < \alpha < 1$ fixed). In typical cases we observe

$$L_n \xrightarrow{D} L_0 \quad (n \rightarrow \infty) \quad \text{for all } \theta \in \Theta_0,$$

where L_0 possesses a rectangular distribution on $(0,1)$, and

$$L_n \rightarrow 0 \quad (n \rightarrow \infty) [P_\theta] \quad \text{for all } \theta \in \Theta \setminus \Theta_0.$$

If the latter convergence is even exponentially fast, the following definition makes sense (cf. Bahadur (1967, 1971)):

Definition 2.2. The sequence $\{T_n\}_{n=1,2,\dots}$ of statistics for testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ has exact slope $c(\theta)$ (when θ obtains) if

$$(2.24) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log L_n = -\frac{1}{2} c(\theta) [P_\theta] .$$

Remark 2.2. For a motivation, consider Fisher's transformation $V_n = -2 \log L_n$. Then, $V_n \xrightarrow{D} V_0$ (under H_0) in typical cases, where V_0 is χ^2_2 -distributed. Plotting $\{(n, V_n), n=1,2,\dots\}$ in $\mathbb{R}^2$, the points move out to infinity in the direction of a ray from the origin, the angle between ray and x-axis given by $\arctan c(\theta)$.

Exact slopes of $\{T_n\}$ have to be distinguished from (so-called) approximate slopes which have also been introduced by Bahadur (1967, 1971) under the following situation: Suppose T_n has an asymptotic null distribution F , i.e.

$$\lim_{n \rightarrow \infty} P_\theta(T_n \leq x) = F(x) \quad (x \in \mathbb{R})$$

for all $\theta \in \Theta_0$. Define

$$L_n^{(a)} = 1 - F(T_n)$$

to be the approximate level attained by T_n .

Definition 2.3. $\{T_n\}$ has approximate slope $c^{(a)}(\theta)$ (when θ obtains) if

$$(2.25) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log L_n^{(a)} = -\frac{1}{2} c^{(a)}(\theta) [P_\theta] .$$

Remark 2.3. Under certain regularity conditions, it can be proved that

$$c(\theta) \sim c^{(a)}(\theta) \quad (\theta \rightarrow \theta_0 \in \Theta_0) .$$

But our personal impression is that approximate slopes have to be used very carefully, if $\theta \in \Theta_1$ is supposed to be fixed.

In the sequel we restrict our considerations to exact slopes. Define, for $0 < \delta < 1$,

$$N(\delta) = \text{smallest integer } m \text{ such that } L_n \leq \delta \text{ for all } n \geq m ,$$

if such a (random) integer exists, otherwise put $N(\delta) = \infty$. In other words. $N(\delta)$ is the smallest sample size required in order that the sequence $\{T_n\}$ (or equivalently $\{L_n\}$) becomes (and remains) significant at the level δ . That means, $N(\delta)$ measures the efficiency of the sequence $\{T_n\}$ of test statistics in this certain sense. For small δ the random variable $N(\delta)$ can be approximated in terms of exact slopes:

If (2.24) holds and $0 < c(\theta) < \infty$, then

$$(2.26) \quad \lim_{\delta \rightarrow 0} N(\delta)/(2 \log (1/\delta)/c(\theta)) = 1 \quad [P_\theta] .$$

Hence, if $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are two different sequences of test statistics, both holding (2.24) with $c^{(1)}(\theta)$ resp. $c^{(2)}(\theta)$ under P_θ ,

$$(2.27) \quad e_{1,2}(\theta) = \lim_{\delta \rightarrow 0} \frac{N^{(2)}(\delta)}{N^{(1)}(\delta)} = \frac{c^{(1)}(\theta)}{c^{(2)}(\theta)} \quad [P_\theta]$$

serves as a measure of A.R.E. of $\{T_n^{(1)}\}$ with respect to $\{T_n^{(2)}\}$, when θ obtains.

A useful method for finding exact slopes of certain test statistics has been given by Bahadur in the following

Theorem 2.3 (Bahadur, 1971). Suppose that

$$(i) \quad \lim_{n \rightarrow \infty} n^{-1/2} T_n = b(\theta) \quad [P_\theta] ,$$

and

$$(ii) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log [1 - F_n(n^{1/2}x)] = -f(x) ,$$

for x in an open interval I , where f is continuous on I .

If $b(\theta) \in I$, assertion (2.24) holds with

$$(2.28) \quad c(\theta) = 2f(b(\theta)) .$$

Proof. Let $\epsilon > 0$ be such that $b-\epsilon, b+\epsilon \in I$. Since F_n is nondecreasing, $n^{1/2}(b-\epsilon) \leq T_n \leq n^{1/2}(b+\epsilon)$ implies $1 - F_n(n^{1/2}(b-\epsilon)) \geq L_n \geq 1 - F_n(n^{1/2}(b+\epsilon))$. Applying (ii) we thus obtain

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log L_n \leq -f(b-\epsilon) \quad [P_\theta] ,$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log L_n \leq -f(b+\epsilon) \quad [P_\theta] .$$

Continuity of f now yields the desired result.

The following theorem shows that the exact slope of any sequence $\{T_n\}$ cannot exceed a certain upper bound. Let $K(\theta, \theta_0) = K(P_\theta, P_{\theta_0})$ be as defined above, $\theta, \theta_0 \in \Theta$, and put

$$(2.29) \quad J(\theta) = \inf \{K(\theta, \theta_0) : \theta_0 \in \Theta_0\} .$$

Then $0 \leq J(\theta) \leq \infty$ for all θ , $J(\theta) = 0$ for $\theta \in \theta_0$, and (in typical cases) $0 < J(\theta) \leq \infty$ for $\theta \in \theta_1 = \theta \setminus \theta_0$. We prove (see e.g. [2])

Theorem 2.4. (Bahadur-Raghavachari-Berk). For each $\theta \in \theta$,

$$(2.30) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log L_n \geq -J(\theta) \quad [P_\theta],$$

that means $c(\theta) \leq 2J(\theta)$, if the exact slope exists.

Proof. It suffices to consider θ 's for which $J(\theta) < \infty$. Choose and fix such a θ , and let $\varepsilon > 0$ be a constant. From the definition of $J = J(\theta)$ it follows that a $\theta_0 \in \theta_0$ exists such that

$$0 \leq K = K(\theta, \theta_0) < J + \varepsilon < \infty.$$

Since $K < \infty$, we have a density $r(x)$ such that $dP_\theta = r(x)dP_{\theta_0}$ on $(X, \mathcal{B})$. Let

$$r_n = \prod_{i=1}^n r(x_i)$$

be a density of $dP_\theta^{(n)}$ with respect to $dP_{\theta_0}^{(n)}$ on $(X^{(n)}, \mathcal{B}^{(n)})$. By definition of K and the law of large numbers, we have

$$(2.31) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log r_n = K(\theta, \theta_0) \quad [P_\theta].$$

Put

$$A_n = \{L_n < \exp[-n(K + 2\varepsilon)]\},$$

$$B_n = \{r_n < \exp[n(K + \varepsilon)]\}.$$

Then

$$P_{\theta}^{(\infty)}(A_n \cap B_n) = \int_{A_n \cap B_n} dP_{\theta}^{(n)} = \int_{A_n \cap B_n} r_n dP_{\theta_0}^{(n)} \\ \leq \exp[n(K+\epsilon)] \int_{A_n \cap B_n} dP_{\theta_0}^{(n)} \leq \exp[n(K+\epsilon)] P_{\theta_0}^{(n)}(L_n < \exp[-n(K+2\epsilon)]) .$$

Using $P_{\theta_0}^{(n)}(L_n \leq x) \leq x$ for $0 \leq x \leq 1$ and $\theta_0 \in \Theta_0$ (see e.g. Bahadur (1971, Theorem 7.4)), we obtain

$$P_{\theta}^{(\infty)}(A_n \cap B_n) \leq \exp(-n\epsilon) .$$

Hence $\sum_n P_{\theta}^{(\infty)}(A_n \cap B_n) < \infty$, which, by Borel-Cantelli lemma, implies that, if θ obtains, then, with probability 1, A_n and B_n together occurs atmost finitely often. However, in view of (2.31), this means that with probability 1,

$$L_n \geq \exp[-n(K+2\epsilon)] \text{ for } n \text{ sufficiently large,}$$

if θ obtains. Thus,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log L_n \geq J(\theta) - 3\epsilon \quad [P_{\theta}] .$$

Since ϵ can be chosen arbitrarily small, the proof is complete.

Let us close this section by considering a standard example:

Example 2.2. Let $(X, \mathcal{B}) = (\mathbb{R}, \mathcal{B})$, and, for

$$dP_{\theta} = \frac{1}{(2\pi)^{1/2}} \exp[-(x-\theta)^2/2] ,$$

consider the testing problem $H_0 : \theta = 0$ against $H_1 : \theta > 0$ (i.e. $\Theta_0 = \{0\}$, $\Theta_1 = (0, \infty)$). From the Neyman-Pearson theory, it is well-known that

$$T_n^{(1)} = n^{-1/2} \sum_{i=1}^n x_i$$

is an optimal test statistic for this problem, and that

$$T_n^{(2)} = T_n^{(1)} / s_n, \quad \text{where} \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

is asymptotically equivalent in a certain sense. We shall see that these notions of optimality coincide with Bahadur's notions of efficiency and exact slopes.

From (1.8) in section 1.2 we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_0(T_n^{(1)} > n^{1/2}x) = -x^2/2$$

$0 < x < \infty$, hence $f_1(x) = x^2/2$, using the notations of Theorem 2.3.

Furthermore,

$$\begin{aligned} P_0(T_n^{(2)} > n^{1/2}x) &= \frac{1}{2} P_0((n^{1/2}\bar{X})^2 - x^2 \frac{n}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 > 0) \\ &= \frac{1}{2} P_0(W_n \geq 0). \end{aligned}$$

Since

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log E \exp(tW_n) = -\frac{1}{2} \log(1+tx^2) = c(t),$$

$t \in (0,1)$, we obtain by the same methods, used to derive the upper bound in Theorem 1.4,

$$\limsup_{n \rightarrow \infty} [P_o(T_n^{(2)} \geq n^{1/2}x)]^{1/n} \leq \inf_{0 \leq t \leq 1} \exp c(t) = (1+x^2)^{-1/2}.$$

On the other hand,

$$P_o(T_n^{(2)} \geq n^{1/2}x) \geq P_o(n^{1/2}x \leq T_n^{(2)} \leq n^{1/2}(x+\epsilon))$$

for all $\epsilon > 0$. Evaluating the right-hand side probability by using the explicit density, we also obtain

$$\liminf_{n \rightarrow \infty} [P_o(T_n^{(2)} \geq n^{1/2}x)]^{1/n} \geq (1 + (x+\epsilon)^2)^{-1/2}$$

Hence, after letting ϵ tend to 0 ,

$$f_2(x) = \frac{1}{2} \log (1 + x^2) ,$$

$0 < x < \infty$. Obviously,

$$\lim_{n \rightarrow \infty} n^{-1/2} T_n^{(1)} = \lim_{n \rightarrow \infty} n^{-1/2} T_n^{(2)} = \theta \quad [P_\theta] ,$$

hence $b^{(1)}(\theta) = b^{(2)}(\theta) = \theta$. Thus, by Theorem 2.3. the exact slopes of $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ (if θ obtains) are given by

$$c^{(1)}(\theta) = \theta^2$$

$$c^{(2)}(\theta) = \log (1 + \theta^2) .$$

Note that $e_{1,2}(\theta) = c^{(1)}(\theta)/c^{(2)}(\theta) > 1$ for each non-null θ , and that $e_{1,2}(\theta) \rightarrow 1$ as $\theta \rightarrow 0$. Moreover,

$$K(\theta, 0) = E_\theta \log \left(\frac{dP_\theta}{dP_o} \right) = \frac{1}{2} \theta^2 = J(\theta) .$$

Hence

$$c^{(1)}(\theta) = 2J(\theta) ,$$

which, by Theorem 2.4, means that $\{T_n^{(1)}\}$ is optimal in the sense of Bahadur-slopes. This coincides with well-known results from the Neyman-Pearson theory for UMP-tests.

3. Fluctuations of random processes. Erdős-Rényi laws.

In this final chapter we shall deal with another application of exponentially converging large deviation probabilities, i.e. the asymptotic description of certain fluctuations of random processes. The results are interesting from a probabilistic as well as a statistical point of view, since, on one hand, asymptotic properties of increments of certain stochastic processes can be characterized with probability 1 (w.p.1), and, on the other hand, such properties even determine the whole underlying distribution in several cases, which is of definite statistical interest. Erdős and Rényi (1970) were the first who, in the case of partial sum sequences, pointed out such a connection between increments of processes and underlying distributions.

3.1 The Erdős-Rényi new law of large numbers.

Consider gains (say) $X_1, X_2, \dots$ of a player in independent repetitions of a fair game, e.g. $P(X_i = \pm 1) = 1/2$ ($i = 1, 2, \dots$). From the strong law of large numbers we know

$$(3.1) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i = 0 \quad \text{w.p.1.}$$

On the other hand, by Borel-Cantelli lemma,

$$X_i = 1 \quad \text{i.o.} \quad \text{w.p.1.},$$

or, in a weaker form,

$$(3.2) \quad \lim_{N \rightarrow \infty} \max(X_1, \dots, X_N) = 1 \quad \text{w.p.1.}$$

Assertion (3.1) means, that we know what (on average) happens in a long series $X_1, \dots, X_N$ of trials, whereas (3.2) says that, again and again, at least one trial is successful. But the question is, what may happen (on the average) in subseries (say) $X_{n+1}, \dots, X_{n+K}$ of length K , which are comparably short in the long series $X_1, \dots, X_N$. Erdős and Rényi (1970) proved that asymptotically each average gain between 0 and 1 can be attained in a subseries of suitable length K depending upon N . More precisely, putting $S_0 = 0$, $S_n = X_1 + \dots + X_n$, and

$$D(N, K) = \max_{0 \leq n \leq N-K} (S_{n+K} - S_n) \quad (1 \leq K \leq N),$$

then, for each $a \in (0, 1)$, a constant $C = C(a)$, uniquely determined by a via the underlying distribution, exists such that

$$(3.3) \quad \lim_{N \rightarrow \infty} \frac{D(N, [C \log N])}{[C \log N]} = a \quad \text{w.p.1.}$$

Here $[\cdot]$ denotes the integer part of a real number. Assertion (3.3) is a special form of the (so-called) Erdős-Rényi new law of large numbers,

which, in a way, bridges the gap between the strong law in (3.1) and an extreme value law like (3.2).

Erdős and Rényi proved a general version of their new law, which we present in the following form:

Theorem 3.1 (Erdős and Rényi, 1970). Let $X_1, X_2, \dots$ be a sequence of non-degenerate i.i.d. random variables on $(\Omega, \mathcal{A}, P)$ satisfying

$$(i) \quad \phi(t) = E \exp(tX_1) < \infty \quad \text{for all } t \in [0, t_1] .$$

Then, for $a \in \{\phi'(t)/\phi(t) : t \in (0, t_1)\}$ and $C = C(a)$ such that $\exp(-1/C) = \inf_t \phi(t) \exp(-ta) = \rho(a)$, we have

$$(3.4) \quad \lim_{N \rightarrow \infty} \max_{0 \leq n \leq N - [C \log N]} \frac{S_{n+[C \log N]} - S_n}{[C \log N]} = a \quad \text{w.p.1} .$$

Note that $\{\phi'(t)/\phi(t) : t \in (0, t_1)\} = (EX_1, \text{esssup } X_1)$, if t_1 has been chosen as large as possible. The proof is heavily based on Chernoff's theorem (Theorem 1.1). Erdős and Rényi used the Bahadur and Ranga Rao refinement (Theorem 1.2), but this turns out to be unnecessary. It is just the exponential behaviour

$$\lim_{n \rightarrow \infty} [P(S_n \geq na)]^{1/n} = \rho(a) ,$$

which is essential for the proof. In the sequel we need certain properties of the rate $\rho(a)$. Thus we are going to collect them, in a more general form, in the following

Lemma 3.1. Let $c : (0, t_1) \rightarrow \mathbb{R}$ be strictly convex and twice continuously differentiable. For $a \in \{c'(t) : t \in (0, t_1)\}$ put $\rho(a) = \inf_t \exp \{c(t) - ta\}$. Then

a) $\rho(a)$ has a representation

$$(3.5) \quad \rho(a) = \exp \{c(t(a)) - at(a)\},$$

where $t(a)$ is the unique solution of $a = c'(t)$.

b) The function $t = t(a)$ is strictly increasing on (a_0, a_1) , where $a_0 = \inf \{c'(t) : t \in (0, t_1)\}$, $a_1 = \sup \{c'(t) : t \in (0, t_1)\}$.

c) $-\log \rho(a)$ is differentiable on (a_0, a_1) with

$$(3.6) \quad (-\log \rho(a))' = t(a).$$

d) $\rho(a)$ is strictly decreasing on (a_0, a_1) .

e) Let $a = a(t)$ be inverse to $t = t(a)$, i.e. $a = c'(t)$. Then

$$(3.7) \quad c(t) = \log \rho(a(t)) + ta(t), \quad t \in (0, t_1).$$

Proof. (a) Since $c(t) - ta$ is still convex, its infimum is attained at $t = t(a)$ such that $c'(t) - a = 0$.

b) $c'(t)$ is strictly increasing, hence is $t = t(a)$.

c) By properties of implicit functions and c being twice differentiable, $t = t(a)$ is also differentiable with $c''(t(a))t'(a) = 1$. Hence, $t'(a) = 1/c''(t(a)) > 0$. Using this and (3.5), we have

$$(-\log \rho(a))' = t(a) + t'(a)\{a - c'(t(a))\} = t(a).$$

d) Since $t(a) > 0$ on (a_0, a_1) , $-\log \rho(a)$ is strictly increasing by (3.6). Hence $\rho(a)$ is strictly decreasing.

e) Using the identity $t(a(t)) = t$, (3.5) yields (3.7).

Remark 3.1. If X_1 is non-degenerate, $c(t) = \log \phi(t)$ is strictly convex and i.o. differentiable on $(0, t_1)$. Thus, all assertions of Lemma 3.1 hold for $\rho(a) = \inf_t \phi(t) \exp(-ta)$, in particular, $\rho(a)$ is strictly decreasing on (a_0, a_1) . By this and Chernoff's theorem, for $a'' < a < a'$, $a', a'' \in (a_0, a_1)$,

$$(3.8) \quad P(S_n \geq na') \leq \rho'^n,$$

where $\rho(a') < \rho' < \rho(a)$, and,

$$(3.9) \quad P(S_n \geq na'') \geq \rho''^n,$$

where $\rho(a'') > \rho'' > \rho(a)$, if only n is sufficiently large.

Inequalities (3.8) and (3.9) are essential for the

Proof of Theorem 3.1. a) We first prove

$$(3.10) \quad \limsup_{N \rightarrow \infty} \frac{D(N, K_N)}{K_N} \leq a \quad \text{w.p.1},$$

where $K = K_N = [C \log N]$. Let $\varepsilon > 0$ be such that $a' = a + \varepsilon \in (a_0, a_1)$.

From (3.8) we obtain, for all sufficiently large N ,

$$P(D(N, K_N) \geq K_N a') \leq N P(S_{K_N} \geq K_N a') \leq N \rho'^{K_N}.$$

Defining N_j to be the largest integer such that $[C \log N_j] = j$,
we have for j sufficiently large,

$$P(D(N_j, K_{N_j}) \geq K_{N_j} a') \leq \exp[(j+1)/C] \rho'^j = e^{1/C} (e^{1/C} \rho')^j.$$

By definition of C , $\exp(-1/C) = \rho(a)$, hence $e^{1/C} \rho' = \rho'/\rho(a) < 1$,
by choice of ρ' . Therefore,

$$\sum_j P(D(N_j, K_{N_j}) \geq K_{N_j} a') < \infty,$$

which, by Borel-Cantelli lemma, implies

$$\limsup_{j \rightarrow \infty} \frac{D(N_j, [C \log N_j])}{[C \log N_j]} \leq a' \text{ w.p.1.}$$

Realizing that, for $N_{j-1} < N \leq N_j$,

$$[C \log N] = j,$$

$$D(N, [C \log N]) \leq D(N_j, [C \log N_j]),$$

the same holds for the whole sequence, i.e.

$$\limsup_{N \rightarrow \infty} \frac{D(N, [C \log N])}{[C \log N]} \leq a' \text{ w.p.1.}$$

Since $a' = a + \varepsilon$, where $\varepsilon > 0$ is arbitrarily small, this proves (3.10).

b) To prove

$$(3.11) \quad \liminf_{N \rightarrow \infty} \frac{D(N, [C \log N])}{[C \log N]} \geq a \text{ w.p.1.},$$

let $\varepsilon > 0$ be such that $a'' = a - \varepsilon \in (a_0, a_1)$. By inequality (3.9) we have,
for all sufficiently large N ,

$$\begin{aligned} P(D(N, K_N) < K_N a'') &\leq P\left(\max_{i=1, \dots, [N/K_N]} (S_{iK_N} - S_{(i-1)K_N}) < K_N a''\right) \\ &= P(S_{K_N} < K_N a'')^{[N/K_N]} \leq \{1 - \rho''^{K_N}\}^{[N/K_N]} \leq \exp(-\rho''^{K_N} [N/K_N]) . \end{aligned}$$

Now, since $\rho'' > \rho(a) = \exp(-1/C)$, for a suitable $\delta > 0$ it holds

$$\rho''^{K_N} \geq \rho''^{C \log N} = \rho(a)^{(1-2\delta)C \log N} = N^{-(1-2\delta)} ,$$

and, for N sufficiently large,

$$[N/K_N] = [N/[C \log N]] \geq N^{1-\delta} .$$

Hence, for $N \geq N_0$,

$$P(D(N, K_N) < K_N a'') \leq \exp(-N^\delta) .$$

Since $\int_{x_0}^{\infty} \exp(-x^\delta) dx < \infty$, we have by integral test

$$\sum_N P(D(N, K_N) < K_N a'') < \infty ,$$

which, by Borel-Cantelli lemma again, implies

$$\liminf_{N \rightarrow \infty} \frac{D(N, [C \log N])}{[C \log N]} \geq a'' \quad \text{w.p.1 .}$$

Letting ϵ tend to 0, (3.11) is proved, which renders the whole proof complete.

Erdős and Rényi (1970) give an interesting application of their "new law of large numbers" to the so-called "stochastic geyser problem".

We state this problem, introduced by Rényi (1962), in a more general version, given by Bartfai (1966), who also gave a first proof:

Stochastic geyser problem. Consider a sequence $S'_1 = S_1 + R_1$, $S'_2 = S_2 + R_2, \dots$ of observations, where $S_1, S_2, \dots$ are partial sums of an i.i.d. sequence $X_1, X_2, \dots$ and $R_1, R_2, \dots$ are random error terms. For the sake of simplicity, suppose that the errors are bounded, i.e. $|R_n| \leq r_n$ ($n = 1, 2, \dots$). The question is, of what magnitude are the error terms allowed to be such that the underlying distribution function $F(x) = P(X_1 \leq x)$ can still be reconstructed from the sequence $S'_1, S'_2, \dots$ with probability 1?

Rényi (1962) posed this problem for the case $S'_n = [S_n]$, the case of rounding-off errors (say). There are several answers to the above question, depending on the magnitude r_n of the error terms:

1) In the case $r_n = o(1)$, define $X'_1 = S'_1$, $X'_i = S'_i - S'_{i-1}$ ($i=2, 3, \dots$). Then $|X'_i - X_i| = o(1)$, and, by Glivenko-Cantelli theorem,

$$\lim_{n \rightarrow \infty} \sup_x |F_n(x) - F(x)| = 0 \text{ w.p.1,}$$

where F_n denotes the empirical distribution function based on $X_1, \dots, X_n$.

Hence it is obvious that $F_n(x)$ can be replaced by $n^{-1} \sum_{i=1}^n I_{\{X'_i \leq x\}}$,

thus determining $F(x)$ from the erroneous sequence $S'_1, S'_2, \dots$, w.p.1.

However, if $r_n \rightarrow \infty$, this method does not work any longer.

2) In the case $r_n = o(\log n)$, Bártfai (1966) was still able to show that the problem is solvable by recursive estimation of all moments, if the corresponding moment-generating function $\phi(t) = E \exp(tX_1)$ is finite in some interval about the origin. Bártfai conjectured that this might be the maximal magnitude admitting a solution of the problem.

3) Bártfai's conjecture was confirmed in two papers of Komlós, Major and Tusnády (1975/76) who proved, that, in the case $r_n = \varepsilon \log n$, even a normal distribution cannot be detected by only observing the erroneous sequence $S'_1, S'_2, \dots$.

Hence, $r_n = o(\log n)$ is indeed the least favourable case still providing a positive answer. But then, however, this answer is also an immediate consequence of Theorem 3.1. For, in this case, assertion (3.4) remains true, even if the sequence $\{S_n\}$ is replaced by the erroneous sequence $\{S'_n\}$. That means, also $\{S'_n\}$ determines the functional dependence $C = C(a)$ on some interval (a_0, a_1) . But, since $1/C(a) = -\log \rho(a)$, we thus also get knowledge of

$$\log \phi(t) = \log \rho(a(t)) + ta(t)$$

on some interval, using Lemma 3.1. Since the moment-generating function uniquely determines the underlying distribution function (by inversion), the problem is solved.

Now, remember that the proof of Theorem 3.1 essentially uses exponentially decreasing large deviation probabilities, which, by

Chernoff's theorem, are available under finite moment-generating functions. On the other hand, by Theorem 1.3 of Section 1.3, this is the only case where exponential rates are possible. This might raise the question, what happens to the Erdős-Rényi law, if no moment-generating function exists? The answer is negative, the law cannot be retained any longer:

Theorem 3.2. (Steinebach, 1978). If $\phi(t) = \infty$ for all $t > 0$, then

$$(3.12) \quad \limsup_{N \rightarrow \infty} \max_{0 \leq n \leq N - [C \log N]} \frac{S_{n+[C \log N]} - S_n}{[C \log N]} = \infty \quad \text{w.p.1}$$

for any constant $C > 0$.

Proof. For arbitrary a and $\rho < 1$, Corollary 1.1 implies the existence of a subsequence $\{N_k\}_{k=1,2,\dots}$ of integers such that

$$P(S_{N_k} \geq N_k a) \geq \rho^{N_k}, \quad k=1,2,\dots$$

Noting that $\log\{(N+1)/N\} \rightarrow 0$, this subsequence can be chosen in such a way that

$$N_k = [C \log N'_k], \quad k = 1,2,\dots$$

for another subsequence $\{N'_k\}_{k=1,2,\dots}$. Now,

$$\begin{aligned} & P(D(N'_k, [C \log N'_k]) < [C \log N'_k] a) = P(D(N'_k, N_k) < N_k a) \\ & \leq P\left(\max_{i=1, \dots, [N'_k/N_k]} (S_{iN_k} - S_{(i-1)N_k}) < N_k a\right) \\ & = \{1 - P(S_{N_k} \geq N_k a)\}^{[N'_k/N_k]} \leq \{1 - \rho^{N_k}\}^{[N'_k/N_k]} \\ & \leq \exp(-\rho^{N_k} [N'_k/N_k]) . \end{aligned}$$

If $\rho = \exp(-1/C')$, where $C' > C$, we have

$$\rho^{N'_k} \geq \rho^{C \log N'_k} = N'_k^{-C/C'} = N'_k^{-(1-2\delta)}$$

for some $\delta > 0$. Furthermore, for k sufficiently large, say $k \geq k_0$,

$$[N'_k/N_k] = [N'_k/[C \log N'_k]] \geq N'_k^{1-\delta}.$$

Hence,

$$P(D(N'_k, [C \log N'_k]) < [C \log N'_k]a) \leq \exp(-N'_k^\delta),$$

$k \geq k_0$, which implies

$$\sum_k P(D(N'_k, [C \log N'_k]) < [C \log N'_k]a) < \infty,$$

by the integral test. Thus, the Borel-Cantelli lemma yields

$$\liminf_{k \rightarrow \infty} \frac{D(N'_k, [C \log N'_k])}{[C \log N'_k]} \geq a \text{ w.p.1,}$$

which immediately implies (3.12), since a can be chosen arbitrarily large.

Using analogous results about the increments of Wiener processes and the Komlós-Major-Tusnády (1975/76) strong approximation, M.Csörgö and Révész (1979) extended the Erdős-Rényi law for partial sums to cases considering increments over blocks of length K_N , where $K_N/\log N \rightarrow \infty$. Taking advantage of some appropriate large deviation theorems Book (see e.g. [10]) was able to also treat non-i.i.d. situations. Moreover, Book obtained some results for increments over blocks of length K_N , where $K_N/\log N \rightarrow 0$, making use of certain properties

of probabilities for very large deviations. These results, however, are (in a way) strong invariance principles and thus no longer afford a reconstruction of the underlying distribution. Nevertheless they provide interesting characterizations of the asymptotic properties of partial sum sequences. Extensions of the mentioned results are also available for partial sum sequences, indexed by parameters taking values on lattices (for details see e.g. [42] and [43]).

All results have in common the use of exponential properties of certain large deviation probabilities and independence as well as some kind of stationarity assumptions about the processes under consideration. In our next section, however, using some different methods, we shall also deal with processes having neither independent nor stationary increments.

3.2. Fluctuations of renewal processes.

Consider an industrial replacement problem. On a positive time-axis random events occur (replacements of units), the interoccurrence-times being denoted X_i (failure time between i^{th} and $(i-1)^{\text{th}}$ item). Let $N(T)$ be the number of replacements up to time T , i.e.

$$N(T) = \max \{n \geq 0 : S_n \leq T\},$$

where $S_0 = 0$, $S_n = X_1 + \dots + X_n$ ($n = 1, 2, \dots$). Moreover, let Y_j be the cost of replacement for the j^{th} failure. Then

$$Z(T) = \sum_{j=1}^{N(T)} Y_j$$

is the total cost of replacement incurred by time T .

Under appropriate conditions, we have

$$\lim_{T \rightarrow \infty} \frac{Z(T)}{T} = \frac{EY_1}{EX_1} \quad \text{w.p.1,}$$

describing the average cost over a long time. But we might ask again, what can happen in least favourable cases over comparably short time intervals of length K , say e.g. $K = K(T) \sim C \log T$. In view of the Erdős-Rényi law, one would expect

$$\lim_{T \rightarrow \infty} \sup_{0 \leq t \leq T - C \log T} \frac{Z(t + C \log T) - Z(t)}{C \log T} = a \quad \text{w.p.1,}$$

where there is a certain one-to-one correspondence between a and C . Indeed, under mild conditions, such a result is available:

Theorem 3.3. Let $\{X_i\}_{i=1,2,\dots}$ and $\{Y_j\}_{j=1,2,\dots}$ be sequences of nonnegative r.v.'s, each i.i.d. and independent of each other, satisfying:

- (i) $P(X_1 > 0) > 0$;
- (ii) $EX_1^3 < \infty$;
- (iii) $f(s) = E \exp(sY_1) < \infty$, $s \in (0, s_1)$;
- (iv) either the X_i 's or the Y_j 's are nondegenerate.

Then, for $a \in \{g'(f(s))f'(s) : s \in (0, s_1)\}$, where $g(x) = \lim_{t \rightarrow \infty} \frac{1}{t} \log EX^N(t)$, we have

$$(3.13) \quad \lim_{T \rightarrow \infty} \sup_{0 \leq t \leq T - C \log T} \frac{Z(t + C \log T) - Z(t)}{C \log T} = a \quad \text{w.p.1},$$

where $C = C(a)$ is uniquely determined by the equation $\exp(-1/C) = \inf_s \exp \{g(f(s)) - sa\} \equiv \rho(a)$.

The proof is heavily based on an exponential behaviour of the large deviation probabilities for the process $\{Z(t)\}_{t \geq 0}$, i.e.

$$(3.14) \quad \lim_{t \rightarrow \infty} [P(Z(t) \geq ta)]^{1/t} = \inf_s \exp \{g(f(s)) - sa\},$$

for $a \in \{g'(f(s))f'(s) : s \in (0, s_1)\}$, which follows from the following

Lemma 3.2. Denote $\phi(s) = E \exp(sX_1)$ (being finite at least for $s \leq 0$). Put $S_{N(t)} = \sum_{i=1}^{N(t)} X_i$. Then

$$(3.15) \quad E(\exp(sS_{N(t)+1}) / \{\phi(s)\}^{N(t)+1}) = 1 \quad (s < 0),$$

$$(3.16) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \log E\left(\frac{1}{\phi(s)}\right)^{N(t)} = -s \quad (s > 0).$$

Outline of proof. a) The following equality holds true:

$$(3.17) \quad \begin{aligned} \phi^n(s) &= E \exp(sS_n) \\ &= \int_{\{N(t) < n\}} \exp\{sS_{N(t)+1} + s(S_n - S_{N(t)+1})\} dP + \int_{\{N(t) \geq n\}} \exp(sS_n) dP. \end{aligned}$$

The first integral can be written as follows:

$$\begin{aligned} &\sum_{v=0}^{n-1} \int_{\{N(t)=v\}} \exp\{sS_{v+1} + s(S_n - S_{v+1})\} dP \\ &= \sum_{v=0}^{n-1} \int_{\{N(t)=v\}} \exp(sS_{v+1}) \phi^{n-(v+1)}(s) dp \end{aligned}$$

$$= \phi^n(s) \int_{\{N(t) < n\}} \exp(s S_{N(t)+1}) / \{\phi(s)\}^{N(t)+1} dP ,$$

remembering $\{N(t) = v\} \in A(X_1, \dots, X_{v+1})$. Using $\{N(t) \geq n\} = \{S_n \leq t\}$, we further have

$$\int_{\{N(t) \geq n\}} \exp(s S_n) / \phi^n(s) dP = P(S_n(s) \leq t) ,$$

where $S_n(s) = X_1(s) + \dots + X_n(s)$ is conjugate to S_n . Since the $X_i(s)$ can be chosen to be i.i.d., and $EX_1(s) = \phi'(s)/\phi(s) > 0$,

$$\lim_{n \rightarrow \infty} P(S_n(s) \leq t) = 0 ,$$

by the law of large numbers. Hence (3.15) follows from (3.17) upon dividing by $\phi^n(s)$ and letting n tend to infinity.

b) Noting that $S_{N(t)+1} > t$, we obtain from (3.15)

$$E \exp(st) \left(\frac{1}{\phi(s)}\right)^{N(t)+1} \geq 1 \quad (s < 0) ,$$

which implies

$$\liminf_{t \rightarrow \infty} \frac{1}{t} \log E \left(\frac{1}{\phi(s)}\right)^{N(t)} \geq -s \quad (s < 0) .$$

To show that also

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \log E \left(\frac{1}{\phi(s)}\right)^{N(t)} \leq -s \quad (s < 0) ,$$

use a truncation argument on the X_i 's (for details see e.g. [41]).

Now, by (3.16), the function $g(x) = \lim_{t \rightarrow \infty} \frac{1}{t} \log Ex^{N(t)}$, appearing in Theorem 3.3, is well-defined on some interval $(1, x_1)$, and, moreover,

$$g\left(\frac{1}{\phi(s)}\right) = -s \quad (s < 0) .$$

By independence of $\{X_i\}$ and $\{Y_j\}$ and identical distributions within each sequence, we have

$$E \exp(sZ(t)) = \sum_{v=0}^{\infty} f^v(s) P(N(t) = v) = E\{f(s)\}^{N(t)} .$$

Hence

$$(3.18) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \log E \exp(sZ(t)) = g(f(s)) \equiv c(s) ,$$

$s \in (0, s_1)$, where c is a strictly convex function under assumption (iv) of Theorem 3.3. By a continuous-index version of Theorem 1.4 (see e.g. [37]), (3.14) is an immediate consequence of (3.18).

Let us now turn to the proof of our renewal theoretic analogue of the Erdős-Rényi law:

Proof of Theorem 3.3. Denote

$$D(T, K) = \sup_{0 \leq t \leq T-K} \{Z(t+K) - Z(t)\}$$

the maximal cost over time-intervals $(t, t+K]$, and

$$V(t) = S_{N(t)+1} - t ,$$

the forward recurrence times of the renewal process.

a) We first prove

$$(3.19) \quad \limsup_{T \rightarrow \infty} \frac{D(T, C \log T)}{C \log T} \leq a \quad \text{w.p.1} .$$

For $a \in A = \{f'(f(s))f'(s) : s \in (0, s_1)\}$, $C = C(a)$ as defined in Theorem 3.3, and $a' = a + \epsilon \in A$ ($\epsilon > 0$), we have

$$P(D(T, K) \geq Ka') \leq P\left(\sup_{0 \leq t \leq [T]-K+1} \{Z([t]+K+1) - Z([t])\} \geq Ka'\right) \\ \leq \sum_{t=0}^{[T]} P(Z(t+K+1) - Z(t) \geq Ka') ,$$

t , of course, taking integer values only in the last line. Now, for fixed t ,

$$Z(t+K+1) - Z(t) = \sum_{j=N(t)+1}^{N(t+K+1)} Y_j$$

can be viewed as a cumulative sum associated with a sequence V_t , $X_{N(t)+2}$, $X_{N(t)+3}$, ... of 'failure times'. This sum is less or equal to

$$Y_{N(t)+1} + \sum_{j=1}^{N_t(K+1)} Y_{N(t)+1+j} ,$$

where $N_t(K+1)$ denotes the number of renewals in the random interval $[t + V_t, t + V_t + K + 1]$ associated with the sequence $X_{N(t)+2}$, $X_{N(t)+3}$, ... Given $N(t)$, the distribution is the same as that of

$$Y_0 + \sum_{j=1}^{N(K+1)} Y_j \quad \text{or} \quad \sum_{j=1}^{N(K+1)+1} Y_j ,$$

where $Y_0; Y_1, Y_2, \dots$ are i.i.d. . Using (3.14) we thus obtain, for K sufficiently large,

$$P(Z(t+K+1) - Z(t) \geq Ka') \leq P(Y_0 + Z(K+1) \geq Ka') \leq \rho'^K ,$$

where $\rho(a) > \rho' > \rho(a')$. Same arguments as used in the proof of Theorem 3.1, then yield (3.19). Note, that difficulties arose from $\{Z(t)\}_{t \geq 0}$ not necessarily having stationary increments.

b) To prove

$$(3.20) \quad \liminf_{T \rightarrow \infty} \frac{D(T, C \log T)}{C \log T} \geq a \quad \text{w.p.1},$$

define a sequence $\{\tau_{i,T}\}_{i=0,1,\dots}$ of random points by

$$\tau_{0,T} \equiv 0,$$

$$\tau_{i,T} = \text{next renewal point after } \tau_{i-1,T}^{+K(T)}, i=1,2,\dots$$

where $K = K(T) = C \log T$. Then, by construction, the $\{Z(\tau_{i,T}^{+K}) - Z(\tau_{i,T})\}$'s are i.i.d., all having the same distribution as $Z(K) - Z(0)$. Moreover

$$D(T, K) \geq \max_{\tau_{i,T}^{+K} \leq T} \{Z(\tau_{i,T}^{+K}) - Z(\tau_{i,T})\}.$$

Defining $M(T) = \#\{\tau_{i,T} : \tau_{i,T}^{+C \log T} \leq T\}$, we certainly have

$$M(T) \leq [T/C \log T].$$

By some estimations, we can also show that

$$\sum_{T=1}^{\infty} P(M(T) < [\delta T/C \log T]) < \infty$$

for $0 < \delta < 1$. That means $M(T) \approx [T/C \log T]$ in an appropriate asymptotic sense. Using (3.14) again and similar arguments as in the proof of Theorem 3.1, part b), (3.20) can be obtained, which renders the proof complete.

Remark 3.2. Assertion (3.13) in Theorem 3.3 remains valid in a 'discrete-time' version. More precisely, setting

$$D^*(N, [C \log N]) = \max_{n=0, \dots, N-[C \log N]} \{Z(n+[C \log N]) - Z(n)\},$$

we also have

$$(3.21) \quad \lim_{N \rightarrow \infty} \frac{D^*(N, [C \log N])}{[C \log N]} = a \quad \text{w.p.1.}$$

For a proof, consider real sequences $\{T_N^{(1)}\}_{N=1,2,\dots}$ and $\{T_N^{(2)}\}_{N=1,2,\dots}$, defined by

$$T_N^{(1)} = \inf \{T \geq N : T - C \log T \geq N - [C \log N]\},$$

$$T_N^{(2)} : T_N^{(2)} \leq N-2 \text{ and } [C \log T_N^{(2)}] + 2 = [C \log N],$$

which is possible for all sufficiently large N . Then estimate

$$\begin{aligned} D^*(N, [C \log N]) &\leq \sup_{0 \leq t \leq T_N^{(1)} - C \log T_N^{(1)}} \{Z(t + C \log T_N^{(1)}) - Z(t)\} \\ &= D(T_N^{(1)}, C \log T_N^{(1)}), \end{aligned}$$

and also

$$\begin{aligned} D(T_N^{(2)}, C \log T_N^{(2)}) &\leq \sup_{0 \leq t \leq T_N^{(2)} - C \log T_N^{(2)}} \{Z([t] + [C \log T_N^{(2)}] + 2) - Z([t])\} \\ &\leq D^*(N, [C \log N]), \end{aligned}$$

using $T_N^{(2)} - C \log T_N^{(2)} \leq N-2 - [C \log T_N^{(2)}] = N - [C \log N]$ in the

last line. Noting that $T_N^{(i)} \rightarrow \infty$ and $\log T_N^{(i)} / \log N \rightarrow 1$ ($N \rightarrow \infty$, $i = 1, 2$), (3.21) immediately follows from (3.13) by these estimations.

The discrete-time version (3.21) of Theorem 3.3 now affords the solution of a "stochastic geyser problem for the renewal process $\{N(t)\}_{t \geq 0}$ ": For (3.21) implies, that even an erroneously and at discrete points observed sequence (say) $\{M(n)\}_{n=1,2,\dots}$, where $M(n) = N(n) + R(n)$, determines the functional dependence $C = C(a)$ in Theorem 3.3 with probability 1, whenever $|R(n)| \leq r_n = o(\log n)$. Since, in the considered case, the one-to-one correspondence between a and C is given by

$$-\frac{1}{C} = \sup_s \{sa - g(e^s)\},$$

knowledge of $C = C(a)$, by Lemma 3.1, implies knowledge of $g = g(e^s)$. By Lemma 3.2, however,

$$g\left(\frac{1}{\phi(s)}\right) = -s,$$

$s < 0$. Thus, if g is known, the m.g.f. ϕ is also available on some interval, which solves the problem, since ϕ determines the underlying distribution. It would be interesting to know whether $r_n = o(\log n)$ is again the worst case admitting a positive solution, which is to be expected from its analogue in the partial sum situation.

For $Y_j \equiv 1$ ($j = 1, 2, \dots$) Theorem 3.3 has been extended in [43] to cover also cases of increments over subintervals of length $K(T)$,

where $K(T)/\log T \rightarrow \infty$ or $K(T)/\log T \rightarrow 0$ ($T \rightarrow \infty$). Moreover, 'generalized' renewal processes are considered there, where the X_i 's are allowed to take negative values, too. A stochastic geyser problem can still be solved under this slightly more general situation (see e.g. [45]).

3.3. Convergence rates in Erdős-Rényi laws.

Recent results, obtained by Révész (1980) for Wiener processes, indicated that convergence rate statements should be available in the original Erdős-Rényi law, at least for normal distributions. Stimulated by Révész's (1980) results we prove in [16] that such statements hold true for other distributions, too.

Theorem 3.4. (M. Csörgö and Steinebach, 1980). Under the assumptions of Theorem 3.1 we have

$$(3.22) \quad \lim_{N \rightarrow \infty} \left(\max_{0 \leq n \leq N - [C \log N]} \frac{S_n + [C \log N]^{-S_n}}{[C \log N]^{\frac{1}{2}}} - [C \log N]^{\frac{1}{2}} a \right) = 0 \text{ w.p.1.}$$

A similar result can be stated in the case of considering increments over blocks of length K_N , where $K_N/\log N \rightarrow \infty$.

Theorem 3.5. (M. Csörgö and Steinebach, 1980). Let $\{X_i\}_{i=1,2,\dots}$ be an i.i.d. sequence with $EX_1 = 0$, $EX_1^2 = 1$, $\phi(t) = E \exp(tX_1) < \infty$ ($|t| < t_1$). Let $\{K_N\}_{N=1,2,\dots}$ be a nondecreasing integer sequence satisfying

- (i) $1 \leq K_N \leq N$,
- (ii) K_N/N is nonincreasing,
- (iii) $\lim_{N \rightarrow \infty} \frac{(\log(N/K_N))^{\frac{1}{2}}}{\log \log N} = \infty$
- (iv) $\tilde{K}_N / \log N \rightarrow \infty$, where $K_N = [\tilde{K}_N]$.

Put $\rho(a) = \inf_t \phi(t) \exp(-ta)$, and, for N sufficiently large, let $a_N > 0$ be the unique positive solution of

$$(v) \quad \rho^{K_N}(a_N / K_N^{\frac{1}{2}}) = K_N / N.$$

Then we have

$$(3.23) \quad \lim_{N \rightarrow \infty} \left(\max_{0 \leq n \leq N - K_N} \frac{S_{n+K_N} - S_n}{K_N^{\frac{1}{2}}} - a_N \right) = 0 \text{ w.p.1.}$$

Before we turn to the proofs, remember that in the standard normal case $\rho(a) = \exp(-a^2/2)$ for all $a > 0$. Hence, by definition of a_N , in this case

$$K_N \frac{1}{2} \left(\frac{a_N}{K_N^{\frac{1}{2}}} \right)^2 = \log \left(\frac{N}{K_N} \right), \text{ i.e. } a_N = (2 \log(N/K_N))^{\frac{1}{2}}.$$

Generally, we have by Taylor series expansion of $-\log \rho(a)$ in a neighborhood about the origin

$$-\log \rho(a) \sim a^2/2 \quad (a \rightarrow 0).$$

Therefore,

$$a_N^2/2 \sim \log(N/K_N), \text{ i.e. } a_N \sim (2 \log(N/K_N))^{\frac{1}{2}} \quad (N \rightarrow \infty).$$

But, using the refined version $-\log \rho(a) = a^2/2 + O(a^3)$, we also obtain

$$\begin{aligned} a_N - (2 \log (N/K_N))^{1/2} &= a_N - (2K_N[-\log \rho(a_N/K_N^{1/2}))]^{1/2} \\ &= a_N(1 - (2K_N[-\log \rho(a_N/K_N^{1/2})]/a_N^2)^{1/2}) \\ &= a_N(1 - [1 + O(a_N/K_N^{1/2})]^{1/2}) = O(a_N^2/K_N^{1/2}) = O(\log N/K_N^{1/2}), \end{aligned}$$

if $K_N/\log N \rightarrow \infty$. That means, however, if $K_N/\log N \rightarrow \infty$, the absolute difference

$$|a_N - (2 \log (N/K_N))^{1/2}|$$

need not tend to 0, or, in other words, the limit behaviour in (3.23) is different from that under normal distributions.

That it is invariant under all distributions in cases $K_N^{1/2}/\log N \rightarrow \infty$, follows from Révész's (1980) result

$$(3.24) \quad \lim_{T \rightarrow \infty} \left(\sup_{0 \leq t \leq T-K_T} \frac{W(t+K_T) - W(t)}{K_T^{1/2}} - (2 \log(T/K_T))^{1/2} \right) = 0 \quad \text{w.p.1}$$

for standard Wiener processes $\{W(t)\}_{t \geq 0}$ and nondecreasing functions $\{K_T\}_{T \geq 0}$ satisfying conditions (i)-(iii) of Theorem 3.5 (in a continuous-parameter version), and from the Komlós-Major-Tusnády (1975/76) strong approximation for the partial sums of the i.i.d. sequence $\{X_i\}$ in Theorem 3.5, which is e.g.

$$(3.25) \quad \sup_{0 \leq t \leq N} |W(t) - S_{[t]}| = O(\log N) \quad \text{w.p.1}.$$

Assertions (3.24) and (3.25) immediately yield a proof of Theorem 3.5 in the case $K_N^{1/2}/\log N \rightarrow \infty$. But (3.25) also indicates that different methods are necessary in cases $K_N^{1/2}/\log N \not\rightarrow \infty$. We shall use the following lemma to prove Theorem 3.4:

Lemma 3.3. Under the assumptions of Theorem 3.4 we have, for $\varepsilon > 0$ and K sufficiently large,

$$a) \quad P(S_K \geq Ka + K^{1/2}\varepsilon) \leq \rho^K(a) \exp(-h\varepsilon K^{1/2}),$$

where $h : a = \phi'(h)/\phi(h)$;

$$b) \quad P(S_K \leq Ka - K^{1/2}\varepsilon) \geq A \rho^K(a) \exp(h\varepsilon K^{1/2}/2) \quad (\exists A > 0).$$

To prove Lemma 3.3, use conjugate distributions and the central limit theorem (for details see [16]).

Outline of Proof of Theorem 3.4. a) To prove, for $\varepsilon > 0$,

$$\limsup_{N \rightarrow \infty} \left(\max_{0 \leq n \leq N - [C \log N]} \frac{S_{n+[C \log N]} - S_n}{[C \log N]^{1/2}} - [C \log N]^{1/2} a \right) \leq \varepsilon \quad \text{w.p.1,}$$

follow the lines of the proof of Theorem 3.1, but instead of (3.8) use part a) of the above lemma. For $N_j =$ largest integer N such that $[C \log N] = j$, we then have

$$\begin{aligned} P\left(\max_{0 \leq n \leq N_j - [C \log N_j]} (S_{n+[C \log N_j]} - S_n) \geq [C \log N_j] a + [C \log N_j]^{1/2} \varepsilon \right) \\ = O(\exp(-h\varepsilon j^{1/2})). \end{aligned}$$

Hence, by Borel-Cantelli lemma,

$$\limsup_{j \rightarrow \infty} \left(\max_{0 \leq n \leq N_j - [C \log N_j]} \frac{S_{n+[C \log N_j]} - S_n}{[C \log N_j]^{\frac{1}{2}}} - [C \log N_j]^{\frac{1}{2}} a \right) \leq \varepsilon \quad \text{w.p.1,}$$

which again turns out to hold for the whole sequence.

b) Using part b) of Lemma 3.3 instead of (3.9) in the proof of Theorem 3.1, we obtain

$$\begin{aligned} P\left(\max_{0 \leq n \leq N - [C \log N]} (S_{n+[C \log N]} - S_n) \leq [C \log N] a - [C \log N]^{\frac{1}{2}} \varepsilon \right) \\ = O(\exp \{-A \exp(h\varepsilon [C \log N]^{\frac{1}{2}}/2) / [C \log N]\}) . \end{aligned}$$

By integral test, the series of these probabilities converges. Hence, using Borel-Cantelli lemma again, we also have

$$\liminf_{N \rightarrow \infty} \left(\max_{0 \leq n \leq N - [C \log N]} \frac{S_{n+[C \log N]} - S_n}{[C \log N]^{\frac{1}{2}}} - [C \log N]^{\frac{1}{2}} a \right) \geq -\varepsilon \quad \text{w.p.1,}$$

which renders the proof complete.

For the proof of Theorem 3.5 we need the following

Lemma 3.4. Under the assumptions of Theorem 3.5 we have, for $\varepsilon > 0$ and N sufficiently large,

$$\begin{aligned} \text{a) } P(S_{K_N}/K_N^{\frac{1}{2}} \geq a_N + \varepsilon) &\leq \rho^{K_N(a_N/K_N^{\frac{1}{2}})} \exp(-\delta(\log(N/K_N))^{\frac{1}{2}}) ; \\ \text{b) } P(S_{K_N}/K_N^{\frac{1}{2}} > a_N - \varepsilon) &\geq A \rho^{K_N(a_N/K_N^{\frac{1}{2}})} \exp(\delta(\log(N/K_N))^{\frac{1}{2}}) , \end{aligned}$$

where A and δ are positive constants depending on ε .

The proof is similar to that of Lemma 3.3 (see [16]).

Outline of Proof of Theorem 3.5. As mentioned above, by Révész's results and the Komlós-Major-Tusnády strong approximation, the proof is finished in the case $K_N^{1/2}/\log N \rightarrow \infty$. Hence let us assume that $K_N/(\log N)^p \rightarrow 0$ for some $p > 2$.

a) By Lemma 3.4, part a), and definition of a_N in (v), we have

$$P\left(\max_{0 \leq n \leq N-K_N} \frac{S_{n+K_N} - S_n}{K_N^{1/2}} \geq a_N + \varepsilon\right) \leq N \rho^{K_N} (a_N/K_N^{1/2}) \exp(-\delta(\log(N/K_N))^{1/2}) \\ = K_N \exp(-\delta(\log(N/K_N))^{1/2}).$$

Since $K_N = o((\log N)^p)$, assumption (iii) of Theorem 3.5 certainly holds, that is

$$\delta(\log(N/K_N))^{1/2} \geq q \log \log N \geq q \log(K_N)^{1/p} = \log K_N^{q/p}$$

for all q and N sufficiently large. Defining N_j = largest integer such that $K_{N_j} = j$, we obtain

$$P\left(\max_{0 \leq n \leq N_j - K_{N_j}} \frac{S_{n+K_{N_j}} - S_n}{K_{N_j}^{1/2}} \geq a_{N_j} + \varepsilon\right) \leq j^{1-(q/p)},$$

j sufficiently large. Hence, choosing $q > 2p$, Borel-Cantelli lemma yields

$$\limsup_{j \rightarrow \infty} \left(\max_{0 \leq n \leq N_j - K_{N_j}} \frac{S_{n+K_{N_j}} - S_n}{K_{N_j}^{1/2}} - a_{N_j} \right) \leq \varepsilon \quad \text{w.p.1.}$$

By some further estimations it can again be shown that this also holds for the whole sequence (for details see [16]).

b) By part b) of Lemma 3.5, we obtain

$$P \left(\max_{0 \leq n \leq N-K_N} \frac{S_{n+K_N} - S_n}{K_N^{1/2}} \leq a_N^{-\varepsilon} \right) = O(\exp \{-A' \exp(\delta' (\log N)^{1/2})\}) ,$$

for some $A', \delta' > 0$ and all N sufficiently large. Hence, by integral test, the series of these probabilities is finite, which, using Borel-Cantelli lemma, implies

$$\liminf_{N \rightarrow \infty} \left(\max_{0 \leq n \leq N-K_N} \frac{S_{n+K_N} - S_n}{K_N^{1/2}} - a_N \right) \geq -\varepsilon \quad \text{w.p.1} .$$

This completes the proof.

In a rewritten form, assertions (3.22) and (3.23) in Theorems 3.4 and 3.5 are obviously convergence rate statements. Assertion (3.22), for instance, can be written as

$$\lim_{N \rightarrow \infty} [C \log N]^{1/2} \left(\max_{0 \leq n \leq N-[C \log N]} \frac{S_{n+[C \log N]} - S_n}{[C \log N]} - a \right) = 0$$

and thus yields a convergence rate $o((\log N)^{-1/2})$ in the original Erdős-Rényi law (Theorem 3.1). Similarly, (3.23) means nothing but

$$\lim_{N \rightarrow \infty} (2 \log(N/K_N))^{1/2} \left(\max_{0 \leq n \leq N-K_N} \frac{S_{n+K_N} - S_n}{(2K_N \log(N/K_N))^{1/2}} - \frac{a_N}{(2 \log(N/K_N))^{1/2}} \right) = 0 .$$

Remembering

$$\frac{a_N}{(2 \log(N/K_N))^{1/2}} - 1 = O\left(\frac{\log N}{(2K_N \log(N/K_N))^{1/2}}\right) ,$$

if $K_N/\log N \rightarrow \infty$, we also have

$$(3.26) \quad \lim_{N \rightarrow \infty} (2 \log(N/K_N))^{1/2} \left(\max_{0 \leq n \leq N-K_N} \frac{S_{n+K_N} - S_n}{(2K_N \log(N/K_N))^{1/2}} - 1 \right) = 0,$$

in case that $(\log N)^2/K_N \rightarrow 0$ ($N \rightarrow \infty$). This yields a convergence rate of order $o((\log(N/K_N))^{-1/2})$ in such an extended version of the Erdős-Rényi law, dealing with increments over blocks of length K_N . If $K_N/\log N \rightarrow \infty$, but $K_N/(\log N)^2 \not\rightarrow \infty$, the same convergence rate can only be retained if the constant limit point 1 in (3.26) is replaced by the sequence $a_N/(2 \log(N/K_N))^{1/2}$.

Similar convergence rate results, as stated above, are also available in other Erdős-Rényi type laws (e.g. Theorem 3.3), whenever improved large deviation estimates, as given in Lemmas 3.3 and 3.4, can be obtained. In the case of the renewal counting process $\{N(t)\}_{t \geq 0}$, such estimates are an immediate consequence of those for partial sums via the identity $P(N(t) \geq n) = P(S_n \leq t)$.

3.4. General versions of Erdős-Rényi laws.

The methods, used in Section 3.2 to derive an Erdős-Rényi type analogue in renewal theory, provide some useful tools for dealing with this kind of limit theorems under rather general situations. For the sake of completeness, in this final section we add some recent results of [44], where such methods are successfully applied.

Various generalizations of the Erdős-Rényi (1970) law have already been found, e.g. for weighted sums of i.i.d. rv's or non-i.i.d. rv's with absolutely continuous distributions (see Book (1973,1976)), for continuous functionals of multivariate partial sum differences (see Bártfai (1977)) or, generally, for such functions of moving blocks of i.i.d. rv's and those of empirical measures of these blocks, for which a first order large deviation theorem holds (see S. Csörgö (1979), who also gives a very complete list of references). Moreover, it was the last author's aim to provide a unified approach to Erdős-Rényi laws and thus "to avoid the trap of proving so many E-R laws separately". For reasons of statistical applications, S. Csörgö (1979) stated two different theorems, which, however, could be combined into one of the following type:

Theorem 3.6 (S. Csörgö, 1979). Let $\{X_i\}_{i=1,2,\dots}$ be an i.i.d. sequence taking values in some measurable space $(X, \mathcal{B})$ and let T_n be real-valued test-statistics depending upon $\{X_i\}_{i=1,2,\dots}$ only through the first n observations, i.e. $T_n = T_n(X_1, \dots, X_n)$. Suppose that

$$(i) \quad \lim_{K \rightarrow \infty} P(T_K \geq Ka)^{1/K} = \rho(a)$$

exists for a -values in some interval (a_0, a_1) , where ρ is a strictly decreasing function, $0 < \rho(a) < 1$.

Then, for $a \in (a_0, a_1)$ and $C = C(a)$ such that $\exp(-1/C) = \rho(a)$ we have

$$(3.27) \quad \lim_{N \rightarrow \infty} \max_{0 \leq n \leq N - [C \log N]} \frac{T_{[C \log N]}^{(X_{n+1}, \dots, X_{n+[C \log N]})}}{[C \log N]} = a \quad \text{w.p.1.}$$

Remark 3.3. It is indeed an exponential large deviation behaviour like (i) in Theorem 3.6, which is essential for the proof of an Erdős-Rényi type law, besides stationarity and independence of disjoint blocks in the i.i.d. sequence $\{X_i\}_{i=1,2,\dots}$. This was S. Csörgö's (1979) main observation, enabling him to get away from the sum-function of real-valued i.i.d. variables and state such a general version. Of course, Theorem 3.6 includes Theorem 3.1, since, by Chernoff's (1952) result, existence of m.g.f. implies an exponential behaviour of the tails of the sum distribution.

But, although Theorem 3.6 already covers rather general cases, it still depends upon an i.i.d. situation, which turns out to be superfluous. In Section 3.2, for instance, we have seen that there are other typical situations also providing certain Erdős-Rényi laws. Our main goal is to extend Theorem 3.6 such that those situations are covered, too. Some further examples will be mentioned indicating the need for such an extension.

Before stating the main theorem let us first give a reformulated version of S. Csörgö's (1979) result, forgetting about the special dependence upon an i.i.d. sequence and substituting it by the basic assumptions:

Theorem 3.7. Let $\{T_{n,K}\}_{n=0,1,2,\dots}^{K=1,2,\dots}$ be a double sequence of real-valued random variables on $(\Omega, \mathcal{A}, P)$ satisfying

- (i) $\lim_{K \rightarrow \infty} P(T_{0,K} \geq Ka)^{1/K} = \rho(a)$
exists for a -values in some interval (a_0, a_1) , where ρ is a strictly decreasing function, $0 < \rho(a) < 1$;
- (ii) for each K , $\{T_{n,K}\}_{n=0,1,2,\dots}$ is a stationary sequence;
- (iii) for each K , $\{T_{iK,K}\}_{i=0,1,2,\dots}$ is an i.i.d. sequence.

Then, for $a \in (a_0, a_1)$ and $C = C(a)$ such that $\exp(-1/C) = \rho(a)$, we have

$$(3.28) \quad \lim_{N \rightarrow \infty} \max_{0 \leq n \leq N - [C \log N]} \frac{T_{n, [C \log N]}}{[C \log N]} = a \quad \text{w.p.1.}$$

Obviously, Theorem 3.7 includes Theorem 3.6. But, for purposes described before, we state the following generalized version:

Theorem 3.8 (Steinebach, 1980). Let $\{T_{n,K}\}_{n=0,1,2,\dots, K=1,2,\dots}$ be a double sequence of real-valued random variables on $(\Omega, \mathcal{A}, P)$ satisfying:

- (i) for a -values in some interval (a_0, a_1) there exists a strictly decreasing function $\rho_1(a)$, $0 < \rho_1(a) < 1$, such that for each $C > 0$:

$$\limsup_{N \rightarrow \infty} P(T_{n, [C \log N]} \geq [C \log N]a)^{1/[C \log N]} \leq \rho_1(a)$$

holds uniformly with respect to (w.r.t.) n ;

- (ii) there exists a sequence $v_{0,N} < v_{1,N} < \dots$ of random indices such that, for each $N = 1, 2, \dots$ and $C > 0$:

- (ii.1) $\{T_{v_{i,N}, [C \log N]}\}_{i=0,1,2,\dots}$ is a sequence of independent rv's

$$(ii.2) \quad \liminf_{N \rightarrow \infty} P(T_{v_{i,N}}; [C \log N] \geq [C \log N] a)^{1/[C \log N]} \geq \rho_2(a)$$

holds uniformly w.r.t. i , where $\rho_2(a)$, $0 < \rho_2(a) < 1$,
is another strictly decreasing function on (a_0, a_1) ;

$$(ii.3) \quad \sum_{N=1}^{\infty} P(L_N \leq (1-\delta)[N/C \log N]) < \infty$$

for some $0 < \delta < 1$, where $L_N = \#\{v_{i,N} : v_{i,N} \leq N - [C \log N]\}$.

Then, for $a \in (a_0, a_1)$ and $C_1 = C_1(a)$ such that $\exp(-1/C_1) = \rho_1(a)$,
we have

$$(3.29) \quad \limsup_{N \rightarrow \infty} \max_{0 \leq n \leq N - [C_1 \log N]} \frac{T_{n, [C_1 \log N]}}{[C_1 \log N]} \leq a \quad \text{w.p.1},$$

and, for $a \in (a_0, a_1)$ and $C_2 = C_2(a)$ such that $\exp(-1/C_2) = \rho_2(a)$,
we have

$$(3.30) \quad \liminf_{N \rightarrow \infty} \max_{0 \leq v_{i,N} \leq N - [C_2 \log N]} \frac{T_{v_{i,N}, [C_2 \log N]}}{[C_2 \log N]} \geq a \quad \text{w.p.1}.$$

Remark 3.4. Using the deterministic sequence $v_{i,N} = i[C \log N]$
($i=0,1,2,\dots$) of indices and $\rho_1(a) = \rho_2(a)$ we see that Theorem 3.7 is
an immediate consequence of Theorem 3.8. However, the latter theorem
is mainly stated for other than i.i.d. situations.

Trying not to overload the proof let us first formulate an
auxiliary lemma:

Lemma 3.5. Under the assumptions of Theorem 3.8, we have for

$a'' < a < a'$, $a', a'' \in (a_0, a_1)$, and N sufficiently large

$$(3.31) \quad P\left(\max_{0 \leq n \leq N - [C_1 \log N]} T_{n, [C_1 \log N]} \geq [C_1 \log N] a'\right) \leq N^{-\beta} \quad (\beta > 0),$$

and, setting $\ell_N = [(1-\delta)[N/C_2 \log N]]$,

$$(3.32) \quad P\left(\max_{i=0,1,\dots,\ell_N-1} T_{v_{i,N}; [C_2 \log N]} < [C_2 \log N] a''\right) \leq \exp(-N^\gamma) \quad (\gamma > 0).$$

Proof. a) By assumption (i) of Theorem 3.8 we have, for N sufficiently large,

$$P\left(\max_{0 \leq n \leq N - [C_1 \log N]} T_{n, [C_1 \log N]} \geq [C_1 \log N] a'\right) \leq N \rho_1'^{C_1 \log N}$$

where $\rho_1(a) > \rho_1' > \rho_1(a')$. Remembering $\rho_1(a) = \exp(-1/C_1)$ and thus setting $\rho_1' = \exp(-1/C_1')$, where $C_1(a) > C_1' > C_1(a')$, we obtain

$$N \rho_1'^{C_1 \log N} = N^{1 - (C_1/C_1')} = N^{-\beta}$$

$\beta = (C_1/C_1') - 1 (> 0)$, which completes the proof of (3.31).

b) By assumptions (ii.1) and (ii.2) we have, for N sufficiently large,

$$\begin{aligned} & P\left(\max_{i=0,1,\dots,\ell_N-1} T_{v_{i,N}; [C_2 \log N]} < [C_2 \log N] a''\right) \\ & \leq (1 - \rho_2'')^{C_2 \log N \ell_N} \leq \exp(-\ell_N \rho_2''^{C_2 \log N}), \end{aligned}$$

where $\rho_2(a'') > \rho_2'' > \rho_2(a)$. Now, choose $\rho_2'' = \exp(-1/C_2'')$ with $C_2(a'') > C_2'' > C_2(a)$, and make use of $\ell_N \geq N^{1-\kappa}$ ($\kappa > 0$) if N is sufficiently large, then

$$\frac{C_2 \log N}{N^{1/2}} \geq N^{1-\kappa-(C_2/C_2'')} .$$

Since $C_2 = C_2(a) < C_2''$, $\gamma = 1 - \kappa - (C_2/C_2'')$ is positive for small κ , which renders the proof complete.

Now, let us turn to the

Proof of Theorem 3.8. a) Let $a' > a$ be close to a , $a' \in (a_0, a_1)$, but fixed, and $C_1 = C_1(a)$ as defined in Theorem 3.8.

We prove

$$(3.33) \quad \limsup_{N \rightarrow \infty} \max_{0 \leq n \leq N - [C_1 \log N]} \frac{T_{n, [C_1 \log N]}}{[C_1 \log N]} \leq a' \quad \text{w.p.1} .$$

Defining $N_j =$ largest integer N s.t. $[C_1 \log N] = j$ (which is possible for all sufficiently large j), we have

$$e^{j/C_1} \leq N_j < e^{(j+1)/C_1}$$

and thus, using Lemma 3.4, (3.31),

$$\sum_{j=1}^{\infty} P\left(\max_{0 \leq n \leq N_j - [C_1 \log N_j]} T_{n, [C_1 \log N_j]} \geq [C_1 \log N_j] a' \right) < \infty$$

Hence, the Borel-Cantelli lemma implies

$$(3.34) \quad \limsup_{j \rightarrow \infty} \max_{0 \leq n \leq N_j - [C_1 \log N_j]} \frac{T_{n, [C_1 \log N_j]}}{[C_1 \log N_j]} \leq a' \quad \text{w.p.1} .$$

Since, for $N_{j-1} < N \leq N_j$, by definition

$$(3.35) \quad [C_1 \log N] = [C_1 \log N_j] = j ,$$

and therefore

$$(3.36) \quad \max_{0 \leq n \leq N_j - [C_1 \log N_j]} \frac{T_{n, [C_1 \log N_j]}}{[C_1 \log N_j]} \geq \max_{0 \leq n \leq N - [C_1 \log N]} \frac{T_{n, [C_1 \log N]}}{[C_1 \log N]},$$

assertion (3.33) follows by combining (3.34)-(3.36) together. Letting a' tend to a in (3.33), assertion (3.29) is proved.

b) Now let $a'' < a$ be fixed, $a'' \in (a_0, a_1)$, and $C_2 = C_2(a)$ as defined in Theorem 3.8. We prove

$$(3.37) \quad \liminf_{N \rightarrow \infty} \max_{0 \leq v_{i,N} \leq N - [C_2 \log N]} \frac{T_{v_{i,N}, [C_2 \log N]}}{[C_2 \log N]} \geq a'' \text{ w.p.1.}$$

First we estimate

$$\begin{aligned} & P\left(\max_{0 \leq v_{i,N} \leq N - [C_2 \log N]} T_{v_{i,N}, [C_2 \log N]} < [C_2 \log N] a''\right) \\ & \leq P(L_N \leq \ell_N) + P\left(\max_{i=0,1,\dots,\ell_N-1} T_{v_{i,N}, [C_2 \log N]} < [C_2 \log N] a''\right) \end{aligned}$$

$\ell_N = [(1-\delta)[N/C_2 \log N]]$. Hence, from assumption (ii.3) and Lemma 3.5, (3.32), we get

$$\sum_{N=1}^{\infty} P\left(\max_{0 \leq v_{i,N} \leq N - [C_2 \log N]} T_{v_{i,N}, [C_2 \log N]} < [C_2 \log N] a''\right) < \infty$$

which, by use of Borel-Cantelli lemma again, implies (3.37). Letting a'' tend to a , assertion (3.37) immediately yields (3.30) and the proof of Theorem 3.8 is complete.

Remark 3.5. It is obvious from the proofs of Lemma 3.5 and Theorem 3.8 that $\max_{0 \leq n \leq N - [C_1 \log N]}$ in assertion (3.29) can be replaced by $\max_{0 \leq n \leq N}$. In the same vein, $\max_{0 \leq v_{i,N} \leq N - [C_2 \log N]}$ in assertion (3.30) can be replaced by any $\max_{i=0,1,\dots,L'_N}$, whenever assumption (ii.3) holds for the random number L'_N , too.

There is a series of examples, given by S. Csörgö (1979), which is already covered by Theorem 3.6 or its rewritten version Theorem 3.7. In Section 3.2, however, we treated a situation, which cannot be dealt with by using Theorem 3.7, but which can be derived from Theorem 3.8. There are some other examples. Book (1973, 1976), for instance, extended the original Erdős-Rényi (1970) law to the weighted sum case or to non-i.i.d. situations. This could also be done by using Theorem 3.8. Moreover, certain Erdős-Rényi type analogues are available for waiting-times in queuing models. They can also be obtained as corollaries of Theorem 3.8. Roughly spoken, this theorem handles situations, where processes, again and again, start independently from certain random index points.

REFERENCES

- [1] Bahadur, R.R. (1967). Rates of convergence of estimates and test statistics. Ann. Math. Statist. 38, 303-324.
- [2] Bahadur, R.R. (1971). Some Limit Theorems in Statistics. Regional Conference Series in Appl. Math. No. 4, SIAM, Philadelphia.
- [3] Bahadur, R.R., Ranga Rao, R. (1960). On deviations of the sample mean. Ann. Math. Statist. 31, 1015-1027.
- [4] Bahadur, R.R., Zabell, S. (1979). Large deviations of the sample mean in general vector spaces. Ann. Probability 7, 587-621.
- [5] Bártfai, P. (1966). Die Bestimmung der zu einem wiederkehrenden Prozess gehörenden Verteilungsfunktion aus den mit Fehlern behafteten Daten einer einzigen Realisation. Studia Sci. Math. Hungar. 1, 161-168.
- [6] Bártfai, P. (1977). On the multivariate Chernoff theorem. Math. Inst. Hungar. Acad. Sci. (Preprint).
- [7] Baum, L.E., Katz, M., Read, R.R. (1962). Exponential convergence rates for the law of large numbers. Trans. Amer. Math. Soc. 102, 187-199.
- [8] Blackwell, D., Hodges, J.L. (1959). The probability in the extreme tail of a convolution. Ann. Math. Statist. 30, 1113-1120.
- [9] Book, S.A. (1973). The Erdős-Rényi new law of large numbers for weighted sums. Proc. Amer. Math. Soc. 38, 165-171.
- [10] Book, S.A. (1976). Large deviation probabilities and the Erdős-Rényi law of large numbers. Canad. J. Statist. 4, 185-210.
- [11] Borovkov, A.A., Mogul'skii, A.A. (1978). Probabilities of large deviations in topological spaces I. Siberian Math. J. 19, 697-709.
- [12] 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.
- [13] Chernoff, H. (1956). Large sample theory - parametric case. Ann. Math. Statist. 27, 1-22.
- [14] Cramér, H. (1938). Sur un nouveaux théorème limite de la théorie des probabilités. Actualités Sci. Indust. No. 736, 5-23. Hermann, Paris.

- [15] Csörgő, M., Révész, P. (1979). How big are the increments of a Wiener process? Ann. Probability 7, 731-737.
- [16] Csörgő, M., Steinebach, J. (1980). Improved Erdős-Rényi and strong approximation laws for increments of partial sums. Carleton Mathematical Series 166, 19 pp. (Preprint)
- [17] Csörgő, S. (1979). Erdős-Rényi laws. Ann. Statist. 7, 772-787.
- [18] Donsker, M.D., Varadhan, S.R.S. (1977). Some problems of large deviations. Publicazione dell' Istituto Nazionale di Alta Matematica. Symposia Mathematica, Vol. XXI, 313-318, Academic Press, London.
- [19] Erdős, P., Rényi, A. (1970). On a new law of large numbers. J. Analyse Math. 23, 103-111.
- [20] Esscher, F. (1932). On the probability function in the collective theory of risk. Skand. Aktuarietidskr. 15, 175-195.
- [21] Feller, W. (1968). An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed., Wiley, New York.
- [22] Groeneboom, P., Oosterhoff, J., Ruymgaart, F. H. (1979). Large deviation theorems for empirical probability measures. Ann. Probability 7, 553-586.
- [23] Heathcote, C.R. (1967). Complete Exponential Convergence and Some Related Topics. Methuen's Review Series in Appl. Probab. Vol. 7, Methuen, London.
- [24] Hoadley, A.B. (1967). On the probability of large deviations of functions of several empirical cdf's. Ann. Math. Statist. 38, 360-382.
- [25] Hoeffding, W. (1965). On probabilities of large deviations. Proc. Fifth Berkeley Symp. Math. Statist. Prob. 1, 203-219.
- [26] Komlós, J., Major, P., Tusnády, G. (1975). An approximation of partial sums of independent rv's and the sample df. I. Z. Wahrscheinlichkeitstheorie verw. Gebiete 32, 111-131.
- [27] Komlós, J., Major, P., Tusnády, G. (1976). An approximation of partial sums of independent rv's and the sample df. II. Z. Wahrscheinlichkeitstheorie verw. Gebiete 34, 33-58.
- [28] Lynch, J. (1978). A curious converse of Sievers' theorem. Ann. Probability 6, 169-173.

- [29] Petrov, V.V., Širokova, I.V. (1973). The exponential rate of convergence in the law of large numbers. Vestnik Leningrad. Univ. No. 7 Mat. Meh. Astronom. 2, 155-157, 165.
- [30] Plachky, D. (1971). On a theorem of G. L. Sievers. Ann. Math. Statist. 42, 1442-1443.
- [31] Plachky, D., Steinebach, J. (1975). A theorem about probabilities of large deviations with an application to queuing theory. Period. Math. Hungar. 6, 343-345.
- [32] Plachky, D., Steinebach, J. (1977). A generalization of a result of Chernoff in large sample theory. Math. Operationsforschung Statist., Ser. Statistics, 8, 375-379.
- [33] Rényi, A. (1962). A sztochasztikus geizir. Publ. Math. Inst. Hungar. Acad. Sci. 7, 643.
- [34] Révész, P. (1980). How to characterize the asymptotic properties of a stochastic process by four classes of deterministic curves? Carleton Mathematical Series 164, 34 pp. (Preprint).
- [35] Sanov, I.N. (1957). On the probability of large deviations of random variables. Sel. Trans. Math. Statist. Prob. 1, 213-244.
- [36] Sievers, G.L. (1969). On the probability of large deviations and exact slopes. Ann. Math. Statist. 40, 1908-1921.
- [37] Steinebach, J. (1976a). Exponentielles Konvergenzverhalten von Wahrscheinlichkeiten grosser Abweichungen. Dissertation, University of Düsseldorf.
- [38] Steinebach, J. (1976b). Das Gesetz der grossen Zahlen von Erdős-Rényi im Wartemodell G/G/1. Studia Sci. Math. Hungar. 11, 459-466 (1980).
- [39] Steinebach, J. (1978a). Convergence rates of large deviation probabilities in the multidimensional case. Ann. Probability 6, 751-759.
- [40] Steinebach, J. (1978b). On a necessary condition for the Erdős-Rényi law of large numbers. Proc. Amer. Math. Soc. 68, 97-100.
- [41] Steinebach, J. (1978c). A strong law of Erdős-Rényi type for cumulative processes in renewal theory. J. Appl. Prob. 15, 96-111.



- [42] Steinebach, J. (1979a). On the increments of partial sum processes with multidimensional indices. University of Düsseldorf (Preprint).
- [43] Steinebach, J. (1979b). Erdős-Rényi-Zuwächse bei Erneuerungsprozessen und Partialsummen auf Gittern. Habilitationsschrift, University of Düsseldorf.
- [44] Steinebach, J. (1980). On general versions of Erdős-Rényi laws. Carleton Mathematical Series 168, 18 pp. (Preprint).
- [45] Steinebach, J. (1981). The stochastic geyser problem for first-passage times. J. Appl. Prob. (to appear).



