The Annals of Statistics
1979, Vol. 7, No. 4, 772–787

ERDOS-RENYI LAWS

BY SANDOR CSORGO

Szeged University

Almost sure limit theorems are proved for maxima of functions of moving blocks of size $c \log n$ of independent rv's and for maxima of functions of the empirical probability measures of these blocks. It is assumed that for the functions considered a first-order large deviation statement holds. It is well known that the indices of these large deviations are, in most cases, expressible in terms of Kullback-Leibler information numbers, and the a.s. limits of the above maxima are the inverses of these indices evaluated at $1/c$. Several examples are presented as corollaries for frequently used test statistics and point estimators.

1. Introduction and summary. Let $Y_1, Y_2, \cdots$ be a sequence of i.i.d. nondegenerate real rv's with a finite moment-generating function $R(t)$ in a nondegenerate interval around the origin. Set $Z_0 = 0$, $Z_n = Y_1 + \cdots + Y_n$, $I(x) = \sup_t(tx - \log R(t))$, and $I(x) = \infty$ for all those x's, where the former expression is meaningless. For any $c > 0$ we have

$$P\left\{\lim_{n\to\infty} \max_{0\leqslant i \leqslant n-[c \log n]} \frac{Z_{i+[c \log n]} - Z_i}{[c \log n]} = \alpha(c)\right\} = 1,$$

where $[\cdot]$ denotes integer part, the limit $\alpha(c)$ is defined by $\alpha(c) = \sup\{x : I(x) \leqslant 1/c\}$ and, as a function of c, it uniquely determines the common distribution of the summands $Y_1, Y_2, \cdots$. This is the "new law of large numbers" of Erdős and Rényi (1970). (For an extension of it and some elaboration of the limit function, see the Appendix at the end of the present paper). They used it to give a new proof for Bártfai's solution to the stochastic geyser problem, and later on it played an important role when Komlós, Major and Tusnády (1974) and (1975) proved that their $\mathcal{O}(\log n)$-rate strong approximation of partial sums (by a Wiener process) cannot be improved.

Komlós and Tusnády (1975) gave a deep insight into this Erdős-Rényi (E-R) phenomenon, when investigating the almost sure behaviour and limit distributions of "first large blocks" and their indices. In the coin-tossing situation ($P\{Y_1 = 0\} = P\{Y_1 = 1\} = \frac{1}{2}$) Erdős and Révész (1976) proved interesting refinements of this special case of the original law, giving a full characterization of the length of the longest head-run. Chan, M. Csörgő and Révész proved common generalizations of the E-R law and the LIL for the increments of one- and multi-parameter Wiener

---

Received July 1977; revised April 1978.
AMS 1970 *subject classifications.* Primary 60F15, 60F10; secondary 62B10, 62G20, 62F20.
Key words and phrases. Strong limit theorems, Erdős-Rényi maxima, large deviations, Kullback-Leibler information number, test statistics, point estimators.

and Kiefer processes. The references of this direction can be found in Chan and Csörgő (1976).

In another series of papers Book proved various generalizations of the E-R law for weighted sums and nonidentically distributed rv's and also several (as he calls them) versions of this law for generalized averages, where, in the latter, the limits (or, equivalently, the lengths of the blocks) do not determine the distribution of the summands. A description of his own work and references are in Book (1976). In these "versions" of Book (1975a) and (1975b) the length of a block is approximately $(c \log n)^{\beta}$ with $\beta > 1$, so they are not E-R laws any more in the sense that the stochastic geyser problem (or the original E-R law) plays the role of the lower limit to the strong invariance principle. Thus it is entirely natural that the limit in these versions of Book is $\alpha(c) = (2/c)^{1/2}$, the E-R limit-function of $N(0, 1)$-summands, and in fact, these versions (at least in the i.i.d.-case) are consequences of the strong invariance principle of Komlós, Major and Tusnády.

Finally, Book and Truax (1976) developed the analogous E-R law for sample quantiles (see Example (B) in Section 3). It seems that this is the full story at present. (See also references added in proof).

One goal of the present paper is to extend the E-R law for other than the sum-function, of the underlying Y-sequence, and the other one is to develop such laws for functionals of empirical df's. Erdős and Rényi based their proof on the Bahadur-Ranga Rao refinement of Chernoff's large deviation theorem. But (as it can also be seen from the proof in Section 2 here) their result can be proved using only the original limit theorem of Chernoff without the Bernstein-Chernoff inequality (Theorem 3.1 in Bahadur (1971)), a first-order large deviation statement. This observation is the base for the following Theorems 1 and 2.

THEOREM 1. *For a fixed natural number k let $X_1^{(j)}, X_2^{(j)}, \cdots (j = 1, \cdots, k)$ be k independent sequences of independent* rv's, *taking values in an arbitrary measurable space $(\mathcal{X}, \mathcal{B})$. For each n let $h_n^{(j)}$ be an arbitrary $(\mathcal{X}^n, \mathcal{B}^n)$-measurable extended real-valued function, $j = 1, \cdots, k$, and construct the new (real)* rv's $T_{i,n}^{(j)} = h_n^{(j)}(X_{i+1}^{(j)}, \cdots, X_{i+n}^{(j)})$. *Furthermore, let H be a k-variate Borel-measurable function. If, in a neighbourhood of α*

$$(1.1) \qquad \lim_{n\to\infty} \tfrac{1}{n} \log P\{H(T_{0,n}^{(1)}, \cdots, T_{0,n}^{(k)}) \geq \alpha\} = -I(\alpha),$$

where $I(\alpha)$ is a positive function, strictly increasing in some neighbourhood of α, then

$$\max_{0 \leq i \leq n-l(n)} H(T_{i,l(n)}^{(1)}, \cdots, T_{i,l(n)}^{(k)}) \to_{\text{a.s.}} \alpha,$$

as $n \to \infty$, where $l(n) = [(I(\alpha))^{-1} \log n]$.

On choosing $\mathcal{X} = \mathcal{R}^1$, $k = 1$, $h_n(x_1, \cdots, x_n) = (x_1 + \cdots + x_n)/n$ and $H(x) = x$ we get the original E-R law from Theorem 1 via Chernoff's theorem.

Towards formulating the other aim, first consider $X_1, X_2, \cdots$, a sequence of independent d-dimensional random vectors with an arbitrary common df $F(x)$,

$x \in \mathfrak{R}^d$ and let $F_n(x)$, $x \in \mathfrak{R}^d$, be the empirical df of $X_1, \cdots, X_n$. If $Z_n(x) = n(F_n(x) - F(x))$, $Z_0 = 0$, then the E-R law says that for each fixed $x \in \mathfrak{R}^d$ and $0 < \alpha < 1$, such that $F(x) \leqslant 1 - \alpha$, we have

$$\max_{0 \leqslant i \leqslant n - l(n)} \frac{1}{l(n)}(Z_{i+l(n)}(x) - Z_i(x)) \to_{\text{a.s.}} \alpha,$$

where $l(n) = [(I^*(\alpha, F(x)))^{-1} \log n]$ and

$$(1.2) \quad I^*(\alpha, t) = (\alpha + t)\log\left(\frac{\alpha + t}{t}\right) + (1 - \alpha - t)\log\left(\frac{1 - \alpha - t}{1 - t}\right) \quad \text{for } 0 \leqslant t \leqslant 1 - \alpha,$$
$$= \infty \quad \text{for } t > 1 - \alpha.$$

Our noted second goal is to develop analogous strong laws for the appropriate maximums of $\mathcal{H}([c \log n]^{-1}(Z_{i+[c \log n]}(\cdot) - Z_i(\cdot)))$ for some reasonable class of functionals $\mathcal{H}$, or, more generally, for functionals of several such averages. The theory of first-order large deviations of functionals of empirical df's is rather unified, and the general formulation will also give nearly complete information about the real notion of the a.s. limits in E-R type laws.

In order to be in accordance with the existing generality in the literature, we formulate the result for empirical probability measures (pm's). Let $\mathbb{S}$ be a separable complete metric space and $\mathfrak{B}$ the σ-field of Borel sets in $\mathbb{S}$. Let Λ be the set of all pm's on $\mathfrak{B}$. For $P, Q \in \Lambda$

$$I(Q, P) = \textstyle\int_{\mathbb{S}} q \log q\,dP \quad \text{if } Q \ll P,$$
$$= \infty \quad \text{otherwise}$$

(with $q = dQ/dP$) denotes the Kullback-Leibler (K-L) information number (or I-divergence) of Q relative to P (specifically, if $\mathbb{S} = \mathfrak{R}^d$, F and G are d-dimensional df's with respective induced Borel measures μ_F and μ_G then, by definition

$$(1.3) \qquad I(G, F) = I(\mu_G, \mu_F)$$

is the K-L number of G realtive to F). Fix a natural number k. For $\mathbb{P} = (P^{(1)}, \cdots, P^{(k)})$, $\mathbb{Q} = (Q^{(1)}, \cdots, Q^{(k)}) \in \Lambda^k$ and $\zeta = (\zeta_1, \cdots, \zeta_k) \in (0, 1]^k$ with $\sum_{j=1}^k \zeta_j = 1$ define $I_\zeta(\mathbb{Q}, \mathbb{P}) = \sum_{j=1}^k \zeta_j I(Q^{(j)}, P^{(j)})$. For an extended real-valued function $\mathcal{H}$ on Λ^k and a real number α set

$$\Omega_\alpha = \Omega_\alpha(\mathcal{H}) = \{\mathbb{Q} \in \Lambda^k : \mathcal{H}(\mathbb{Q}) \geqslant \alpha\}$$

and define

$$I_\zeta(\alpha) = I_\zeta(\alpha, \mathcal{H}, \mathbb{P}) = \inf\{I_\zeta(\mathbb{Q}, \mathbb{P}) : \mathbb{Q} \in \Omega_\alpha\}.$$

Further, for each $1 \leqslant j \leqslant k$, let $X_1^{(j)}, \cdots, X_{n_j}^{(j)}$ be i.i.d. rv's taking values in $\mathbb{S}$ according to a pm $P^{(j)} \in \Lambda$, where we suppose that the k samples are also independent. Denote by $P_{n_j}^{(j)}$ the empirical pm of the jth sample, i.e., $n_j P_{n_j}^{(j)}(B) =$ the number of $X_1^{(j)}, \cdots, X_{n_j}^{(j)}$ with values in $B \in \mathfrak{B}$. Suppose that all the sample sizes n_j tend to infinity such that $n_j/N \to \zeta_j$, $1 \leqslant j \leqslant k$ where $N = n_1 + \cdots + n_k$.

Then, if $\mathcal{H}$ is a nice enough function, we find that

$$(1.4) \qquad \lim_{N\to\infty} \frac{1}{N} \log P\{\mathcal{H}(P_{n_1}^{(1)}, \cdots, P_{n_k}^{(k)}) \geqslant \alpha\} = -I_{\xi}(\alpha).$$

Motivated by Sanov (1957), Hoadley (1967) was the first to prove (1.4) in case of $\mathcal{S} = \mathcal{R}^1$ and when $P^{(1)}, \cdots, P^{(k)}$ are all nonatomic. With an extra rate condition on the convergence of n_j, he requires $\mathcal{H}$ to be uniformly continuous in the product topology on Λ^k induced by the usual supremum metric on Λ. In case of $k = 1$ (but $\mathcal{S} = \mathcal{R}^d$, $d \geqslant 1$) his result was further developed by Stone (1974). A result of Borovkov (1967) reduces uniform continuity of $\mathcal{H}$ to continuity in this one-sample case. Special functionals were treated before and after by a number of authors (see the examples in Section 3, and the references in Bahadur (1971); it was, in fact, Bahadur's exact slope that motivated this research in the most part). Finally, in a very substantial paper, Groeneboom, Oosterhoff and Ruymgaart (1976) proved (1.4) without any special restriction, assuming only that $\mathcal{H}$ is continuous in the k-product of the topology τ (on Λ) of convergence on all Borel sets, and I_{ξ} is continuous from the right in α. (Even this last assertion is only a special case of their Corollary 3.2).

Here we shall be concerned with equal sample sizes ($n_1 = \cdots = n_k = n$). In this case

$$I_{(1/k, \cdots, 1/k)}(\alpha) = \frac{1}{k} \inf\{\Sigma_{j=1}^k I(Q^{(j)}, P^{(j)}) : (Q^{(1)}, \cdots, Q^{(k)}) \in \Omega_\alpha\},$$

and we denote by $I_k(\alpha)$ this quantity. Then (1.4) becomes

$$(1.5) \qquad \lim_{n\to\infty} (kn)^{-1} \log P\{\mathcal{H}(P_n^{(1)}, \cdots, P_n^{(k)}) \geqslant \alpha\} = -I_k(\alpha).$$

Let $S_0^{(j)} = 0$, $S_n^{(j)} = nP_n^{(j)}$, and $U^{(j)}(i, n) = \frac{1}{n}(S_{i+n}^{(j)} - S_i^{(j)})$, $1 \leqslant j \leqslant k$. Now we can formulate the following

THEOREM 2. *Assume that a functional $\mathcal{H}$ satisfies* (1.5) *at each point in some neighbourhood of α and that α is a point of strict growth of I_k. Then*

$$(1.6) \qquad \max_{0 \leqslant i \leqslant n - l(n)} \mathcal{H}(U^{(1)}(i, l(n)), \cdots, U^{(k)}(i, l(n))) \to_{\text{a.s.}} \alpha,$$

as $n \to \infty$, where $l(n) = l(n, k, \alpha) = [(kI_k(\alpha))^{-1} \log n]$. (An alternative formulation is as follows. If (α_1, α_2) is the (possibly infinite) interval where $I_k(\alpha)$ is positive and strictly increasing, and c is any number in the range of $(kI_k(\alpha))^{-1}$, $\alpha \in (\alpha_1, \alpha_2)$, then (1.6) *holds with $l(n) = [c \log n]$, and the* a.s. *limit is $\alpha(c)$ for which $I_k(\alpha(c)) = (kc)^{-1}$.)*

We have just seen conditions under which (1.5) holds. Note that $I_k(\alpha)$ is always a nonnegative and monotonically nondecreasing function. The condition that I_k is strictly increasing in some neighbourhood of α is a very mild one. For example, if Ω_α is convex and $I_k(\alpha)$ is attained on some $\mathbb{Q} \in \Omega_\alpha$, then I_k is strictly increasing. On the other hand, $I_k(\alpha)$ is attained if Ω_α is closed in the topology of total variation as shown by Csiszár (1975), or closed in τ-topology (see Lemma 3.2 of Groeneboom et al. (1976)). In fact, it is hard to imagine a functional for which I_k would be constant on some interval. In typical cases the interval (α_1, α_2) is such that $I_k(\alpha_1 +) = 0$ and $I_k(\alpha_2 -) = L_k$ say, where this L_k can be ∞. Therefore $\alpha(c)$ is determined, in these cases, on $((kL_k)^{-1}, \infty)$, it is strictly decreasing and $\alpha(0 +) = \alpha_2$, $\alpha((kL_k)^{-1} -) = \alpha_1$.

Theorems 1 and 2 in their stated form do not follow from each other, but, of course, there is an intersection between them. For example, in case of real-valued identically distributed rv's and symmetrical $h_n^{(1)}, \cdots, h_n^{(k)}$ functions the assertion of Theorem 1 follows from that of Theorem 2. But this latter form could not cover some interesting examples. On the other hand, Theorem 1 could have been so formulated to contain Theorem 2. The only reason we did not do so was that it would have been unwise to hide the special form of Theorem 2 which allows us to fully describe the limits of the E-R laws. That the latter can be expressed in terms of the K-L information is of special interest; a fact which would not have been apparent in a more general form. Clearly then, the proof of the two statements is the same. In the next section we prove Theorem 2, but, mutatis mutandis, this is a proof of Theorem 1 as well. This proof, of course, follows the line of that of Erdős and Rényi in general. The difference is that we must be careful here when handling the remainder term of the large deviation.

The results of this exposition can play exactly the same role when best rates of strong approximations for functionals, other than sums of the underlying rv's, are to be constructed, as that played by the classical E-R law for sums (Komlós, Major and Tusnády (1974), (1975)). Since our E-R maximums can be viewed as strongly consistent estimates for the inverse of exact Bahadur slopes of test statistics and point estimators, these results might also be of some interest in statistics. Indeed, the just mentioned possibility of line of thought was the main motivation of the author to collect and work out most of the numerous corollaries of statistical nature in Section 3. These corollaries are E-R laws for sums of vectors, sample quantiles, the trimmed mean, classical test statistics like t, F, chi-square, likelihood ratio, all standard rank statistics, for the most frequently used functionals of one-, and two-sample empirical processes, and for maximum likelihood and other consistent point estimators. While working out these corollaries an attempt was also made to summarize the literature of large deviations for these just mentioned statistics. The aim, however, was not to skin the cat too many times, but rather to avoid the trap of proving so many E-R laws separately. The examples themselves seem to be important from the point of view of information theoretical statistics.

2. Proof of Theorem 2. It will be convenient to write (1.5) in the following form.

$$P\{\mathcal{H}(P_n^{(1)}, \cdots, P_n^{(k)}) \geqslant \alpha\} = \exp\{-nkI_k(\alpha) + a_n \log n\}, \tag{2.1}$$

where a_n is such a sequence that $(a_n \log n)/n \to 0$. Let $\varepsilon > 0$ be such a small number that (2.1) holds for $\alpha \pm \varepsilon$ in place of α, and also $\delta = \delta(\varepsilon) = k(I_k(\alpha + \varepsilon) - I_k(\alpha)) > 0$, $\zeta = \zeta(\varepsilon) = (I_k(\alpha) - I_k(\alpha - \varepsilon))/I_k(\alpha) > 0$. Let $b_n = a_{l(n)}$, and for a positive integer N and a positive number $L > N$ define $\Theta(L, N) = \max_{0 \leqslant i \leqslant L-N} \mathcal{K}(U^{(1)}(i, N), \cdots, U^{(k)}(i, N))$. Then the left hand side of (1.6) is $\Theta(n, l(n))$, and we first show that

$$\text{(2.2)} \qquad \limsup_{n\to\infty} \Theta(n, l(n)) \leqslant \alpha, \text{ a.s.}$$

If $L_N = \exp\{(N+1)kI_k(\alpha)\}$, then we have, by (2.1), that

$$\begin{aligned} P\{\Theta(L_N, N) \geqslant \alpha + \varepsilon\} &\leqslant \Sigma_{i=0}^{[L_N]-N} P\{\mathcal{K}(U^{(1)}(i, N), \cdots, U^{(k)}(i, N)) \geqslant \alpha + \varepsilon\} \\ &= \Sigma_{i=0}^{[L_N]-N} N^{a_N} \exp\{-NkI_k(\alpha + \varepsilon)\} \\ &\leqslant CN^{a_N} e^{-\delta N}, \end{aligned}$$

where $C = \exp\{kI_k(\alpha)\}$. Now

$$\frac{-\log(N^{a_N} e^{-\delta N})}{\log N} = \frac{N}{\log N}\left(\delta - \frac{a_N \log N}{N}\right) \to \infty,$$

as $N \to \infty$, whence $C \Sigma N^{a_N} e^{-\delta N} < \infty$ by the first-order logarithm-criterion. This means a.s. that only finitely many of the events $\{\Theta(L_N, N) \geqslant \alpha + \varepsilon\}$ occur. But, if $L_{N-1} \leqslant n < L_N$ then $l(n) = N$; hence $\{\Theta(n, l(n)) \geqslant \alpha + \varepsilon\} \subseteq \{\Theta(L_N, N) \geqslant \alpha + \varepsilon\}$, and, consequently, the same is true for the events $\{\Theta(n, l(n)) \geqslant \alpha + \varepsilon\}$. (2.2) then holds, since $\varepsilon > 0$ was arbitrarily small.

On the other hand, leaving out the overlapping blocks, and then using the independence of the samples and applying (2.1) again, we find that

$$\begin{aligned} P\{\Theta(n, l(n)) < \alpha - \varepsilon\} &\leqslant P\{\mathcal{K}(U^{(1)}(rl(n), l(n)), \cdots, U^{(k)}(rl(n), l(n))) \\ &\qquad < \alpha - \varepsilon, 0 \leqslant r \leqslant [n/l(n)] - 1\} \\ &\leqslant \Pi_{r=0}^{[n/l(n)]-1}(1 - P\{\mathcal{K}(U^{(1)}(rl(n), l(n)), \cdots, U^{(k)}(rl(n), l(n))) \geqslant \alpha - \varepsilon\}) \\ &= \left(1 - \mathcal{O}\left(\{(kI_k(\alpha))^{-1} \log n\}^{b_n} n^{-\tau}\right)\right)^{\frac{n}{l(n)}}, \end{aligned}$$

where $\tau = I_k(\alpha - \varepsilon)/I_k(\alpha) > 0$. For $j = 0, 2$ set $c_n(j) = (b_n - j)\log\{(kI_k(\alpha))^{-1} \log n\}/\log n$. Now $c_n(0) \to 0$, since $c_n(0) \backsim U_{l(n)}$, where $U_N = (a_N \log N)/((kI_k(\alpha))^{-1}N)$ and we know that $U_N \to 0$ as $N \to \infty$. Therefore $d_n = (c_n(0) - \tau)\log n \to -\infty$, which implies that $\{(kI_k(\alpha))^{-1} \log n\}^{b_n} n^{-\tau} = e^{d_n} \to 0$. This means (with $1 - \tau = (I_k(\alpha) - I_k(\alpha - \varepsilon))/I_k(\alpha) = \zeta$) that

$$P\{\Theta(n, l(n)) < \alpha - \varepsilon\} \leqslant v_n = \exp\left\{-\mathcal{O}\left(\{(kI_k(\alpha))^{-1} \log n\}^{b_n - 1} n^{\zeta}\right)\right\}.$$

Since $c_n(2)$ also converges to zero, we have that

$$\{(kI_k(\alpha))^{-1} \log n\}^{b_n - 2} n^{\zeta} = \exp\{(c_n(2) + \zeta)\log n\} \to \infty,$$

and, consequently, $-\log v_n/\log n \to \infty$. Hence $\Sigma P\{\Theta(n, l(n)) < \alpha - \varepsilon\} < \infty$ again by the logarithm-criterion, and, again applying the Borel-Cantelli lemma, we get

(2.3) $$\liminf_{n\to\infty}\Theta(n, l(n)) \geqslant \alpha, \text{ a.s.}$$

(2.2) and (2.3) together prove the theorem.

3. Examples. The general form of these examples is the following. Given $k(\geqslant 1)$ sequences $X_1^{(j)}, X_2^{(j)}, \cdots, j = 1, \cdots, k$, of rv's, for each i we construct from $X_{i+1}^{(j)}, \cdots, X_{i+l(n,k)}^{(j)}$, $j = 1, \cdots, k$, a new rv $T_{i,l(n,k)}$ for which, in many cases, we shall have

(3.1) $$\max_{0\leqslant i\leqslant n-l(n,k)} T_{i,l(n,k)} \to_{\text{a.s.}} \alpha,$$

as $n \to \infty$, where $l(n, k) = [(kI_k(\alpha))^{-1} \log n]$. The function $I_k(\alpha)$ is determined by the distribution of the terms of the X-sequences and the mode of construction of T. In case of $k = 1$ the corresponding quantities will be denoted simply by $\{X_n\}$, $l(n)$ and $I(\alpha)$. All the references below refer to the corresponding large deviation result, with the only exception in Examples (A) and (B).

EXAMPLE (A). *Sums of vectors.* Let $X_1, X_2, \cdots$ be i.i.d. d-dimensional random vectors with common df $F(x)$, $x \in \mathfrak{R}^d$, and moment-generating function $R(t) = \int_{\mathfrak{R}^d} \exp\{\langle t, x\rangle\} dF(x)$, $t \in \mathfrak{R}^d$, $\langle t, x\rangle = \Sigma_{i=1}^d t_i x_i$, assumed to be finite in some neighbourhood of the origin. Let $\zeta(x) = \inf\{\exp\{-\langle t, x\rangle\}R(t) : t \in \mathfrak{R}^d\}$ be the d-variate Chernoff function, and for $A \subset \mathfrak{R}^d$ put $\zeta(A) = \sup\{\zeta(x) : x \in A\}$. Let $h(x)$ be a d-variate continuous real function such that $\zeta(C_\alpha) > 0$ if $\alpha > h(EX_1)$, where $C_\alpha = \{x : h(x) > \alpha\}$. If

$$T_{i,n} = h\left(\frac{1}{n}(X_{i+1} + \cdots + X_{i+n})\right),$$

then we have (3.1) for $T_{i,l(n)}$ with $\alpha > h(EX_1)$ and $I(\alpha) = -\log \zeta(C_\alpha)$. This result is a generalization of the original E-R law and is stated by Bártfai (1977) as a consequence of his multidimensional generalization of Chernoff's theorem. It is worthwhile to remark here that the present $I(\alpha)$ also has a K-L information form, since, by Lemma 2 of Hoeffding (1965), for every d-dimensional Borel set B we have $-\log \zeta(B) = I(\Omega(B), F) = \inf\{I(G, F) \text{ of } (1.3) : G \in \Omega(B)\}$, where $\Omega(B) = \{G : \int_{\mathfrak{R}^d}(x_1, \cdots, x_d)dG(x_1, \cdots, x_d) \in B\}$.

EXAMPLE (B). *Sample quantiles.* Given i.i.d. rv's $X_1, X_2, \cdots$ with continuous, strictly increasing df $F(x)$, let $X_{i,n}(q)$ be the q-quantile of the subsequence $X_{i+1}, \cdots, X_{i+n}$, i.e., if $X_{i,n}(1) \leqslant \cdots \leqslant X_{i,n}(n)$ are the corresponding order statistics, then $X_{i,n}(q) = X_{i,n}(\overline{qn})$, where $\bar{y}$ here stands for the smallest integer $\geqslant y$. Denoting by G_n the empirical df of n uniformly distributed rv's on (0, 1), from the elementary identity $P\{X_{0,n}(q) > \alpha\} = P\{F(X_{0,n}(q)) > F(\alpha)\} = P\{G_n(1 - F(\alpha)) - (1 - F(\alpha)) > F(\alpha) - q\}$, $F(\alpha) > q$, it follows via Chernoff's theorem that $\lim_{n\to\infty} n^{-1} \log P\{X_{0,n}(q) > \alpha\} = -I^*(F(\alpha) - q, 1 - F(\alpha))$, $I^*(\cdot, \cdot)$ being that of (1.2), provided that $0 < q < F(\alpha) < 1$. Therefore (3.1) holds for $T_{i,l(n)} = X_{i,l(n)}(q)$ with $l(n) = [(I^*(F(\alpha) - q, 1 - F(\alpha)))^{-1} \log n]$. This is the E-R law of Book and Truax (1976) mentioned in the introduction.

EXAMPLE (C). *Trimmed mean.* Let $X_1, X_2, \cdots$ be i.i.d. rv's with df $F \in D$, where D is the set of all one-dimensional df's, and for each i, n let $X_{i,n}(1), \cdots, X_{i,n}(n)$ be the ordering of $X_{i+1}, \cdots, X_{i+n}$. For $0 < a < \frac{1}{2}$, the a-trimmed mean of the latter sample is

$$T_{i,n} = (n - 2[an])^{-1}\Sigma_{k=[an]+1}^{n-[an]} X_{i,n}(k).$$

By Theorem 5.3 of Groeneboom et al. (1976) we have (3.1) for $T_{i,l(n)}$ with $-\infty < \alpha < \infty$ and $I(\alpha) = \inf\{I(G, F)$ of (1.3) $: G \in \Omega_\alpha^a\}$, where $\Omega_\alpha^a = \{G \in D : \int_a^{1-a} G^{-1}(u)du \geqslant (1 - 2a)\alpha\}$. In the case when F is continuous Groeneboom et al. (1976) gave a more explicit formula for this $I(\alpha)$. Applying Theorems 5.1 and 5.2 of these authors, the present E-R law holds not only for the trimmed mean but for more general linear combinations of order statistics (L-estimators).

EXAMPLE (D). *t- and F-test statistics.* Let $X_1^{(j)}, X_2^{(j)}, \cdots$ be i.i.d. rv's with common normal $-N(m_j, \sigma^2)$ distribution, $j = 1, \cdots, k$. Set

$$\bar{X}_{i,n}^{(j)} = \frac{1}{n}\Sigma_{m=i+1}^{i+n} X_m^{(j)}, \qquad j = 1, \cdots, k,$$

and let $T_{i,n}$ be $(kn - k)^{-\frac{1}{2}}$ times the statistic in Student's k-sample t-test

$$T_{i,n} = \left(\Sigma_{j=1}^{k}\left(\bar{X}_{i,n}^{(j)} - m_j\right)\right) \Big/ \left\{\frac{k}{n}\Sigma_{j=1}^{k}\Sigma_{s=i+1}^{i+n}\left(X_s^{(j)} - \bar{X}_{i,n}^{(j)}\right)^2\right\}^{\frac{1}{2}},$$

and construct

$$\hat{F}_{i,n} = \left(\Sigma_{s=i+1}^{i+n}\left(X_s^{(1)} - \bar{X}_{i,n}^{(1)}\right)^2\right) \Big/ \left(\Sigma_{s=i+1}^{i+n}\left(X_s^{(2)} - \bar{X}_{i,n}^{(2)}\right)^2\right)$$

of Fisher's F-test. Also, with a fixed $r \geqslant 2$, form

$$\tilde{F}_{i,n} = \left(\Sigma_{s=i+1}^{i+r}\left(X_s^{(1)} - \bar{X}_{i,r}^{(1)}\right)^2\right) \Big/ \left(\Sigma_{j=2}^{k}\Sigma_{s=i+1}^{i+n}\left(X_s^{(j)} - \bar{X}_{i,n}^{(j)}\right)^2\right).$$

Then (3.1) holds for this $T_{i,l(n,k)}$ with $0 < \alpha < \infty$ and $I(\alpha, k) = (1/(2k))\log(1 + \alpha^2)$ (Bahadur (1971, page 13)). We also formulate, but only in the present situation, the equivalent statement (in all the other examples the analogous remark holds throughout): for all $c > 0$ we have, as $n \to \infty$,

$$\max_{0 \leqslant i \leqslant n-[c \log n]} T_{i,[c \log n]} \to_{\text{a.s.}} \left(\exp\left\{\frac{2}{c}\right\} - 1\right)^{\frac{1}{2}}. \tag{3.2}$$

For $\hat{F}_{i,l(n,2)}$ (3.1) is also valid with $1 < \alpha < \infty$ and $I(\alpha, 2) = (\frac{1}{2})\log((1 + \alpha)/(2\alpha^{\frac{1}{2}}))$ by Klotz (1967) or Killeen, Hettmansperger and Sievers (1972). For $\tilde{F}_{i,l(n,2)}$ we have (3.1) with $0 < \alpha < \infty$ and $I(\alpha, k) = ((k - 1)/(2k))\log(1 + \alpha)$. Naturally, all the above examples could have been formulated in nonstatistical language, i.e., writing simply appropriate chi-square variables instead of the figuring sample variances.

EXAMPLE (E). *Chi-square test statistics.* Let $X_1, X_2, \cdots$ be i.i.d. discrete rv's with possible values $a_1, \cdots, a_d$, $d \geqslant 2$, and $P\{X_1 = a_j\} = p_j$, where $p = (p_1, \cdots, p_d) \in \Theta = \{v = (v_1, \cdots, v_d)$: $v_j > 0, j = 1, \cdots, d, \Sigma_{j=1}^{d} v_i = 1\}$. For each n and $j = 1, \cdots, d$ let $nv_{i,n}^{(j)}$ = the number of indices m with $i + 1 \leqslant m \leqslant i$

$+ n$ and $X_m = a_j$. For $v, q \in \Theta$ set

(3.3) $$\delta(v, q) = \Sigma_{j=1}^{d}(v_j - q_j)^2/q_j.$$

With $\nu_{i,n} = (\nu_{i,n}^{(1)}, \cdots, \nu_{i,n}^{(d)})$ construct

(3.4) $$T_{i,n} = \delta(\nu_{i,n}, p)$$

of the chi-square test. Then (3.1) holds for $T_{i,l(n)}$ with $0 < \alpha < \max\{p_j^{-1} - 1 : 1 \leqslant j \leqslant d\}$ and $I(\alpha) = \inf\{I(v, p) : v \in \Omega_\alpha\}$, where

(3.5) $$I(v, p) = \Sigma_{j=1}^{d} v_j \log(v_j/p_j)$$

is the discrete K-L information, and $\Omega_\alpha = \{v : v \in \Delta, \delta(v, p) \geqslant \alpha\}$, where Δ is the closure of Θ (see Bahadur (1971, page 31)).

Now let $X_1^*, X_2^*, \cdots$ be i.i.d. rv's with absolutely continuous df $F(x)$. Assume that $F'(x) = f(x)$ exists for all x, and $f(x) > 0$ for $a < x < b$ and zero otherwise for $-\infty \leqslant a < b \leqslant \infty$. Let an integer d and constants $0 = q_0 < q_1 < \cdots < q_d = 1$ be given. For $F(b_j) = q_j, j = 1, \cdots, d-1, b_0 = a, b_d = b$, denote by $n\nu_{i,n}^{(j)}$ now the number of those indices m for which $i + 1 \leqslant m \leqslant i + n$ and $X_m \in (b_{j-1}, b_j]$. Then, redefining $T_{i,n}$ by (3.4) in terms of the new $\nu_{i,n}$ and $p = (p_1, \cdots, p_d)$, $p_j = q_j - q_{j-1}, j = 1, \cdots, d$, we have, of course, the same result. But, in the setting of the present paragraph, there is another interesting chi-square test proposed by Witting (see Sievers (1975), page 904)), when the cell boundaries are determined by order statistics.

Let $Y_{i,n}(1) \leqslant \cdots \leqslant Y_{i,n}(n)$ be the order statistics of $X_{i+1}^*, \cdots, X_{i+n}^*$. For all n let $n_j = [nq_j] + 1$ and $Z_{i,n}^{(j)} = Y_{i,n}(n_j)$, $j = 1, \cdots, d-1$, further $Z_{i,n} = (Z_{i,n}^{(0)}, Z_{i,n}^{(1)}, \cdots, Z_{i,n}^{(d-1)}, Z_{i,n}^{(d)})$, where $Z_{i,n}^{(0)} = a$, $Z_{i,n}^{(d)} = b$. For $z = (z_0, \cdots, z_d)$ let $g(z) = (g_1(z), \cdots, g_d(z))$ with $g_j(z) = F(z_j) - F(z_{j-1}), j = 1, \cdots, d$, and $p^{(n)} = (p_1^{(n)}, \cdots, p_d^{(n)})$ with $p_j^{(n)} = (n_j - n_{j-1})/(n+1), j = 1, \cdots, d, n_0 = 0, n_d = n + 1$. Define the statistic

$$T_{i,n}^* = \delta(g(Z_{i,n}), p^{(n)}).$$

Then (3.1) holds for $T_{i,l(n)}^*$ with α as before and $I(\alpha) = \inf\{I(p, v)$ of (3.4) $: v \in \Omega_\alpha\}$. Note that $I(p, v) \neq I(v, p)$, in general.

More generally, let h be a real function on $\mathcal{R}^{d-1}$ such that for the set $A_\alpha = \{x \in \mathcal{R}^{d-1} : h(x) \geqslant \alpha\}$ we have $\phi \neq A_\alpha \subset$ closure (interior(A_α)). Let

$$\tilde{T}_{i,n} = h(Z_{i,n}^{(1)}, \cdots, Z_{i,n}^{(d-1)}),$$

and p and $g(z)$ as in the preceeding paragraph. Then, by Sievers (1975, page 904), we have (3.1) for $\tilde{T}_{i,l(n)}$ with α in that interval where $\tilde{I}(\alpha) = \inf\{I(p, g(z))$ of (3.5): $z_0 = a, z_d = b, (z_1, \cdots, z_{d-1}) \in A_\alpha\}$ is strictly increasing.

EXAMPLE (F). *Likelihood ratio statistic.* Let $\mathcal{X}$ be a finite-dimensional Euclidean space, $\mathcal{B}$ be the field of its Borel sets, and $\mathcal{P}$ be a given set of probability measures on $\mathcal{B}$. It is assumed that $\mathcal{P}$ is a dominated set, i.e., there exist a σ-finite μ, and a family $\{f_P : P \in \mathcal{P}\}$ of $\mathcal{B}$-measurable functions, $0 \leqslant f_P \leqslant \infty$, such that, for each $P \in \mathcal{P}$, $dP/d\mu = f_P$ on $\mathcal{X}$. Fix one single measure P_0 in $\mathcal{P}$ and let $\mathcal{P}_1 = \mathcal{P} - \{P_0\}$. Let $X_1, X_2, \cdots$ be i.i.d. rv's with values in $\mathcal{X}$ and distributed according to P_0. For each $P \in \mathcal{P}$, i and n let $L_{i,n}(P) = \Pi_{j=i+1}^{i+n} f_P(X_j)$ and put $L_{i,n}(\mathcal{P}_1) = \sup\{L_{i,n}(P) : P \in \mathcal{P}_1\}$. For any $Q \in \mathcal{P}_1$, denote $I(Q, P_0) = \int_{\mathcal{X}} \log(f_Q/f_{P_0})\, dQ$, which equals the I-divergence of (1.3) ($\mu_G = Q, \mu_F = P_0$).

Form the *likelihood ratio statistic* of Neyman and Pearson

$$T_{i,n} = (1/n)\log(L_{i,n}(\mathcal{P}_1)/L_{i,n}(P_0)).$$

Suppose Assumptions 10.3 and 10.4 (10.1 and 10.2 are vacuous here) of Bahadur (1971, page 38) to hold. Then, by his Theorem 10.1, we have (3.1) again in every α in the interior of the set $K = \{I(Q, P_0) \text{ of } (1.3) : Q \in \mathcal{P}_1\}$ with $I(\alpha) = \alpha$, that is

$$\max\{T_{i,[\alpha^{-1}\log n]} : 0 \leqslant i \leqslant n - [\alpha^{-1}\log n]\} \to_{\text{a.s.}} \alpha.$$

In the special case when all the measures of $\mathcal{P}$ are concentrated on the set $\{a_1, \cdots, a_d\}$ (i.e., the setting of the first paragraph of Example (E)), P_0 with weights $p_1, \cdots, p_d$, then $K = (0, \max\{\log(1/p_j) : 1 \leqslant j \leqslant d\})$.

EXAMPLE (G). *Rank statistics.* Given, for each i and n, independent rv's $X_{i+1}, \cdots, X_{i+n}$, denote by $R_{i+1}, \cdots, R_{i+n}$ their ranks, and define the linear rank statistic

$$T_{i,n} = (1/n)\Sigma_{j=i+1}^{i+n} a_n(R_j/(n+1), j/(n+1)),$$

where $a_n(\cdot,\cdot)$ is a function on the unit square such that for each n, a_n is constant on rectangles $\{(i-1)/n \leqslant u < i/n,\ (j-1)/n \leqslant v < j/n\}$, $1 \leqslant i, j \leqslant n$, and there exists a function $a(\cdot,\cdot)$ such that $\sup\{|\int_0^1\int_0^1(a_n - a)w| : w \in W\} \to 0$, as $n \to \infty$, where $W = \{w(\cdot,\cdot) : w \geqslant 0, \int_0^1 w(u,v)\, du = 1 = \int_0^1 w(u,v)\, dv\}$ is the set of all bivariate densities on the unit square with uniform marginals (Property A of Woodworth (1970)). Assume that for all i and n the vector of ranks is equally likely to be any of the $n!$ permutations of $(1, \cdots, n)$. Then, by Theorem 1 of Woodworth (1970), (3.1) holds for $T_{i,l(n)}$ with $\alpha_1 = \int_0^1\int_0^1 a < \alpha < \sup\{\int_0^1\int_0^1 aw : w \in W\} = \alpha_2$ and $I(\alpha) = \inf\{\int_0^1\int_0^1 w \log w : \int_0^1\int_0^1 aw \geqslant \alpha,\ w \in W\}$, provided this $I(\alpha)$ is strictly increasing. Under a further mild condition Woodworth gives an explicit formula for $I(\alpha)$ in terms of integral equations.

The present general E-R law for linear rank statistics covers numerous concrete examples which could also be deduced from earlier, parallel and later large deviation statements. In the one-sample situation the statistic of Klotz (1965) (see also in Bahadur (1971, page 14)) is a good example, which is a common generalization of the sign- , Wilcoxon- , and normal scores-test statistics. The two-sample case covers all standard equal sample size Chernoff-Savage statistics (see Theorem 2 of Woodworth (1970), which, in turn, was augmented by Remark 1 of Hájek (1974)). In particular, the appropriate large deviation base for the E-R law for the two-sample Wilcoxon-Mann-Whitney statistic can be found in Hoadley (1967), Stone (1967), Sievers (1969) and Woodworth (1970), for the Fisher-Yates-Terry-Hoeffding normal scores statistic in Stone (1968), Woodworth (1970) and (together with absolute and quadratic scores statistics) in Hwang and Klotz (1975), for

Mood's median test statistic in Woodworth (1970) and (together with Mathisen's median test statistics) in Killeen, Hettmansperger and Sievers (1972). Several other concrete examples of E-R laws for linear rank statistics can be immediately written down utilizing other large deviation results in the cited papers.

There can also be mentioned E-R laws for nonlinear rank statistics. Such is the law for Kendall's tau statistic, based on its large deviation by Sievers (1969) and Woodworth (1970), and the laws for Hodges' and Blumen's bivariate sign test statistics, based on Killeen and Hettmansperger (1972), together with their own Wilcoxon-type signed rank statistic and the nonrank competitor of the latter three (for testing the location of a bivariate normal density), Hotelling's T^2.

EXAMPLE (H). *Functionals of the empirical process.* Let $X_1, X_2, \cdots$ be i.i.d. d-dimensional random vectors with common df $F(x)$, $x \in \mathfrak{R}^d$. Let $F_{i,n}(x)$ be the empirical df of $X_{i+1}, \cdots, X_{i+n}$, and consider the *Kolmogorov-Smirnov* and *Kuiper statistics* $D_{i,n}^{\pm} = \sup\{\pm(F_{i,n}(x) - F(x))\psi(F(x)) : x \in \mathfrak{R}^d\}$, $D_{i,n} = \max\{D_{i,n}^+, D_{i,n}^-\}$, $K_{i,n} = D_{i,n}^+ + D_{i,n}^-$.

If $\psi(t) \equiv 1$ and $T_{i,l(n)}$ is any of $D_{i,l(n)}, D_{i,l(n)}^+, D_{i,l(n)}^-, K_{i,l(n)}$ then for $0 < \alpha < 1$ we have (3.1) with $I(\alpha) = \inf\{I^*(\alpha, F(x)) : x \in \mathfrak{R}^d\}$, where $I^*(\alpha, t)$ is defined under (1.2). The corresponding first three large deviation results are due to Sethuraman (1964) (for $d = 1$ and continuous F see all four also in Bahadur (1971, page 15–17); the fourth was first derived by Abrahamson (1967)). But combining the methods of Sethuraman (1964, page 1311) and Bahadur (1971, page 17) we get the general result for Kuiper's statistic as well.

Now let $d = 1$ and assume F to be continuous. Let $\psi(t)$ be finite, positive and continuous function in (0, 1) such that $t\psi(t) \to 0$ as $t \to 0+$ and $(1-t)\psi(t) \to 0$ as $t \to 1-$. By Abrahamson (1967), if $T_{i,l(n)}$ is any of $D_{i,l(n)}, D_{i,l(n)}^+, D_{i,l(n)}^-$, then (3.1) holds true for $0 < \alpha < 1$ and $I(\alpha) = \inf\{I^*(\alpha/\psi(t), t) : 0 < t < 1\}$, where $I^*(\cdot, \cdot)$ is again that of (1.2). Note that the most interesting weight-function $\psi(t) = (t(1-t))^{-\frac{1}{2}}$ is included here.

Consider also the *Cramér-von Mises statistic*

$$w_{i,n}^2 = \int_{-\infty}^{\infty}(F_{i,n}(x) - F(x))^2\, dF(x).$$

By Hoadley's theorem we have (3.1) for $w_{i,l(n)}^2$ with $0 < \alpha < \frac{1}{3}$ and in case of $d = 1$, F continuous, the corresponding $I(\alpha)$ (of (1.5)) was computed by Mogul'skiĭ (1977). $I(\alpha) = \int_0^1 v(t)\log v(t)\, dt$, where the density function $v(t)$ is zero outside [0, 1] and otherwise is the solution of the equation $\int_{v(0)}^{v(t)}(2(1+u))^{-1}(\lambda(1+c-u-\log u))^{-\frac{1}{2}}\, du = t$, where the constants are such that $\lambda > 0$, $c > 0$, $v(1) < 1 < v(0)$, $v(0) - \log v(0) = v(1) - \log v(1) = 1 + c$, and $\int_0^1 \log v(t)\, dt = \alpha\lambda + c$. The solution is unique. Mogul'skiĭ also proved that $I(\alpha) = (\pi^2/2)\alpha + (\pi^4/24)\alpha^2 + \mathcal{O}(\alpha^{\frac{5}{2}})$ as $\alpha \to 0$. The E-R law also holds of course, for other integral-type statistics like those of Anderson-Darling and Watson, for instance. In these cases the explicit form of the I-divergences are not known, only their local behaviour ($\alpha \to 0$) as determined by Nikitin (1976).

EXAMPLE (I). *Functionals of two empirical* df's. Let $X_1^{(j)}, X_2^{(j)}, \cdots, j = 1, 2,$ be two independent sequences of i.i.d. d-dimensional random vectors with common df $F(x)$, $x \in \mathfrak{R}^d$. With the appropriate empirical df's (as in the former example), let $M_{i,n}(x) = |F_{i,n}^{(1)}(x) - F_{i,n}^{(2)}(x)|$, $M_{i,n}^{\pm}(x) = \pm(F_{i,n}^{(1)}(x) - F_{i,n}^{(2)}(x))$, and consider the (equal size) *two-sample Kolmogorov-Smirnov*, *Kuiper* and *Littell statistics*

$$M_{i,n}^{\pm} = \sup\{M_{i,n}^{\pm}(x) : x \in \mathfrak{R}^d\},\ M_{i,n} = \max(M_{i,n}^{+}, M_{i,n}^{-}),\ V_{i,n} = M_{i,n}^{+} + M_{i,n}^{-},\ U_{i,n} = \min(M_{i,n}^{+}, M_{i,n}^{-}).$$

Then the original E-R law says that if $T_{i,l(n,2)}$ is any of $M_{i,l(n,2)}^{\pm}(x)$, $M_{i,l(n,2)}(x)$, then at each fixed $x \in \mathfrak{R}^d$ we have (3.1) with $0 < \alpha < 1$ and $I(\alpha, 2, x) = (\frac{1}{2})\tilde{I}(\alpha, F(x))$, where (for $0 \leqslant t \leqslant 1$)

$$\tilde{I}(\alpha, t) = \alpha \log K(\alpha, t) - \log\{1 + (t(1-t)(K(\alpha, t) - 1)^2/K(\alpha, t))\},$$

$$K(\alpha, t) = B(t)\alpha/(1-\alpha) + \{(B(t)\alpha/(1-\alpha))^2 + (1+\alpha)/(1-\alpha)\}^{\frac{1}{2}},$$
$$B(t) = (1/(2t(1-t))) - 1,$$

computed by Chernoff's theorem. Suppose that F is continuous. Extending Abrahamson (1967) (she treated the case $d = 1$, and her expression is rather more complicated, since she dealt with unequal sample sizes) we find that the large deviation index of $M_{0,n}$, $M_{0,n}^{+}$, $M_{0,n}^{-}$, $V_{0,n}$ and $U_{0,n}$ is (again) the infimum of $\tilde{I}(\alpha, t)$, $0 \leqslant t \leqslant 1$. $\tilde{I}(\alpha, t)$ takes its minimum at $t = \frac{1}{2}$. The corresponding result even more simply follows in case of $d = 1$ from the known exact distribution of $M_{0,n}^{+}$. This was noted by Klotz (1967), and for $U_{0,n}$ by Littell (1972). Hence (3.1) is true for any of $M_{i,l(n,2)}^{+}$, $M_{i,l(n,2)}^{-}$, $M_{i,l(n,2)}$, $V_{i,l(n,2)}$, $U_{i,l(n,2)}$ with $0 < \alpha < 1$ and $I(\alpha, 2) = (\frac{1}{2})\{(1+\alpha)\log(1+\alpha) + (1-\alpha)\log(1-\alpha)\} = I^*(\alpha/2, \frac{1}{2})$, where $I^*(\cdot, \cdot)$ is that of (1.2). Note that in the alternative formulation c can only be in $(1/\log 4, \infty)$.

By Hoadley's theorem (3.1) also holds for the appropriate equal size *two-sample Cramér-von Mises statistic*, but the explicit form of the corresponding $I(\alpha, 2)$ (of (1.5)) is not known (at least to the author).

EXAMPLE (J). *A test on the circle.* Let $X_1, X_2, \cdots$ be independent uniformly distributed rv's on the circumference of a unit circle, and denote by $N_{i,n}$ the maximal number of points from $X_{i+1}, \cdots, X_{i+n}$ that can be covered by a suitably chosen semicircle. Set $N_{i,n}^* = (2N_{i,n}/n) - 1$. Here $n^{\frac{1}{2}}N_{0,n}^*$ is *Ajne's statistic* for testing uniformity on the circle. By Rao (1972), for $N_{i,l(n)}^*$ we have (3.1) with $0 < \alpha < 1$ and $I(\alpha) = I^*(\alpha/2, \frac{1}{2})$, i.e., the same as in the previous example.

EXAMPLE (K). *Consistent point estimators.* Let $(\mathfrak{X}, \mathfrak{B}, P_\theta)$ be an arbitrary probability space for each θ in an open interval Θ of the real line. Assume that for each $\theta \in \Theta$, P_θ admits a density function $f(x, \theta)$ with respect to a given σ-finite measure. Let $s = (x_1, x_2, \cdots)$ be a sequence of i.i.d. observations of $\mathfrak{X}$, which is then distributed according to $P_\theta = P_\theta^{(\infty)}$ on $(\mathfrak{X}^\infty, \mathfrak{B}^\infty)$ when θ obtains. Set $u(x_1, \theta)$

$= (\partial/\partial\theta)\log f(x_1, \theta)$, and let $\varphi(t, \theta, \alpha) = E_\theta \exp\{tu(x_1, \theta + \alpha)\}$ and $I_\theta = E_\theta(u(x_1, \theta))^2$ be the corresponding moment generating function and Fisher information. Introduce also $\zeta_1(\theta, \alpha) = \inf\{\varphi(t, \theta, \alpha) : t \geqslant 0\}$, and $\zeta_2(\theta, \alpha) = \inf\{\varphi(t, \theta, -\alpha) : t \leqslant 0\}$. Let $\hat{\theta}_{i,n} = \hat{\theta}_{i,n}(s) = \hat{\theta}_{i,n}(x_{i+1}, \cdots, x_{i+n})$ be the usual maximum likelihood estimator (MLE) of θ based on $x_{i+1}, \cdots, x_{i+n}$. Under the regularity conditions of Fu (1973) or those of Bahadur (see Remark 2 in Fu (1973), and Proposition 6 in Bahadur (1967)), if c is a sufficiently large positive number and $T_{i,n} = |\hat{\theta}_{i,n} - \theta|$, then

$$P_\theta\{\max_{1 \leqslant i \leqslant n-[c \log n]} T_{i,[c \log n]} \to \alpha(c, \theta)\} = 1, \tag{3.6}$$

where $\alpha(c, \theta)$ is such that $\max\{\zeta_1(\theta, \alpha(c, \theta)), \zeta_2(\theta, \alpha(c, \theta))\} = \exp\{-1/c\}$. Exactly the same result holds for $T_{i,[c \log n]} = |\theta^*_{i,[c \log n]} - \theta|$, where $\theta^*_{i,n}$ is the "maximum probability estimator with respect to a prior density" (a concept that generalizes MLE) of Fu (1973) under his regularity conditions. In both cases $\alpha(c, \theta) \sim (2/(cI_\theta))^{\frac{1}{2}}$ as $c \to \infty$. If $g(\theta)$ is a sufficiently smooth function on Θ and $\hat{T}_{i,n} = g(\hat{\theta}_{i,n})$ is the MLE of g, then (3.6) also holds for $T_{i,[c \log n]} = |\hat{T}_{i,[c \log n]} - g(\theta)|$, provided that the conditions above are in force again. The limit now is asymptotically $(2(dg(\theta)/d\theta)/(cI_\theta))^{\frac{1}{2}}$ as $c \to \infty$.

In the case of not necessarily maximum likelihood-like consistent estimators two other results can be mentioned based on Fu (1975). Let $t_{i,n} = t_{i,n}(x_{i+1}, \cdots, x_{i+n})$ be an estimator of θ based on $x_{i+1}, \cdots, x_{i+n}$, having a density function $f_n(t, \theta)$. Firstly, suppose that in each $\theta \in \Theta$ $r(t, \theta) = \lim_{n\to\infty} n^{-1} \log f_n(t, \theta)$ is such that $0 < I(\alpha, \theta) = \min(-r(t + \alpha, \theta), -r(t - \alpha, \theta))$ is continuous and strictly increasing in $(0, \alpha_2)$, where $0 < \alpha_2$ can be ∞. If the additional regularity conditions of Theorem 2.1 of Fu (1975) are satisfied, then (3.6) is true for $T_{i,[c \log n]} = |t_{i,[c \log n]} - \theta|$ with $c > (I(\alpha_2 - , \theta))^{-1}$, and the limit $\alpha(c, \theta)$ is such that $I(\alpha(c, \theta), \theta) = 1/c$.

Secondly, for the consistent estimator $t_{0,n}$ suppose that the log-likelihood ratio $L_n(t, \theta', \theta) = (1/n)\log(f_n(t, \theta')/f_n(t, \theta))$ is monotonic and, for every $\theta, \theta' \in \Theta$ there exists a positive constant $R(\theta', \theta)$ such that $L_n(t_{0,n}, \theta', \theta) \to R(\theta', \theta)$ a.s. mod $P_{\theta'}$. If $I^*(\alpha, \theta) = \inf_{\theta'}\{R(\theta', \theta) : |\theta' - \theta| > \alpha\}$ is positive and strictly increasing in α, then for all $c > 0$ we have (3.6) for the corresponding $T_{i,[c \log n]}$ with $I^*(\alpha(c, \theta), \theta) = 1/c$. Note that if $I(P_{\theta'}, P_\theta)$ is the K-L information number of (1.3), then $I^*(\alpha, \theta) \leqslant \inf_{\theta'}\{I(P_{\theta'}, P_\theta) : |\theta' - \theta| > \alpha\}$.

4. Appendix on the original E-R law. Let $Y_1, Y_2, \cdots$ be a sequence of i.i.d. nondegenerate real rv's, and let $Z_0 = 0$, $Z_n = Y_1 + \cdots + Y_n$, $R(t) = E\exp\{tY_1\}$, $A_- = \inf\{t : R(t) < \infty\}$, $A_+ = \sup\{t : R(t) < \infty\}$, $\psi(t) = (d/dt)\log R(t)$. For $x \in (EY_1, \lim_{t\nearrow A_+}\psi(t))$ let $\zeta_+(x) = \inf\{\exp\{-tx\}R(t) : 0 \leqslant t \leqslant A_+\}$, and for $x \in (\lim_{t\searrow A_-}\psi(t), EY_1)$ let $\zeta_-(x) = \inf\{\exp\{-tx\}R(t) : A_- \leqslant t \leqslant 0\}$ (the two pieces of Chernoff's function). Let us make the conventions: $-\log 0 = \infty$, $1/\infty = 0$. Set $I_\otimes(x) = -\log \zeta_\otimes(x)$, $\otimes = +, -$, and let $I_\otimes^{-1}(\cdot)$ be the function inverse to $I_\otimes(\cdot)$. Finally, let $m(n, c)$ be a sequence of positive integers such that $m(n, c)/\log n \to c$, as $n \to \infty$. The result cited in the first paragraph of the introduction is already a slight generalization of the original statement of Erdős and Rényi, and is due to P.

Bártfai. Using Chernoff's theorem and a closer analysis of the Chernoff function (the latter can be based on Borovkov (1964, page 255)), the following full form of the E-R theorem results from essentially the same proof as in Section 2.

THEOREM. *If* $0 < A_+$, *then for all* $c > 0$

$$\max_{0 \leqslant i \leqslant n - m(n,c)} \frac{Z_{i+m(n,c)} - Z_i}{m(n,c)} \to_{\text{a.s.}} \alpha_+(c),$$

where

$$\alpha_+(c) = \sup\left\{x : I_+(x) \leqslant \frac{1}{c}\right\} = I_+^{-1}(1/c), \quad \textit{for} \quad c \geqslant 1/-\log P\{Y_1 = \operatorname{esssup} Y_1\},$$
$$= \operatorname{esssup} Y_1, \qquad \textit{for} \quad c < 1/-\log P\{Y_1 = \operatorname{esssup} Y_1\}.$$

If $A_- < 0$, *then for all* $c > 0$

$$\min_{0 \leqslant i \leqslant n - m(n,c)} \frac{Z_{i+m(n,c)} - Z_i}{m(n,c)} \to_{\text{a.s.}} \alpha_-(c),$$

where

$$\alpha_-(c) = \inf\left\{x : I_-(x) \leqslant \frac{1}{c}\right\} = I_-^{-1}(1/c), \quad \textit{for} \quad c \geqslant 1/-\log P\{Y_1 = \operatorname{essinf} Y_1\},$$
$$= \operatorname{essinf} Y_1, \qquad \textit{for} \quad c < 1/-\log P\{Y_1 = \operatorname{essinf} Y_1\}.$$

Acknowledgment. I am grateful to a referee whose comments improved the presentation of this paper.

REFERENCES

ABRAHAMSON, I. G. (1967). The exact Bahadur efficiences for the Kolmogorov-Smirnov and Kuiper one- and two-sample statistics. *Ann. Math. Statist.* **38** 1475–1490.

BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324.

BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics*. Regional conference series in applied mathematics. SIAM, Philadelphia.

BÁRTFAI, P. (1977). On the multivariate Chernoff theorem. Preprint of the *Math. Inst. Hungar. Acad. Sci.*

BOOK, S. A. (1975a). An extension of the Erdős-Rényi new law of large numbers. *Proc. Amer. Math. Soc.* **48** 438–446.

BOOK, S. A. (1975b). A version of the Erdős-Rényi law of large numbers for independent random variables. *Bull. Inst. Math. Acad. Sinica* **3** 199–211.

BOOK, S. A. (1976). Large deviation probabilities and the Erdős-Rényi law of large numbers. *Canad. J. Statist.* **4** 185–209.

BOOK, S. A. and TRUAX, D. R. (1976). An Erdős-Rényi law for sample quantiles. *J. Appl. Probability* **13** 578–583.

BOROVKOV, A. A. (1964). Analysis of large deviations in boundary-value problems with arbitrary boundaries (in Russian). *Sibirsk. Math. Ž.* **5** 253–289.

BOROVKOV, A. A. (1967). Boundary-value problems for random walks and large deviations in function spaces (in Russian.) *Teor. Verojatnost. i Primenen.* **12** 635–654.

CHAN, A. H. C. and CSÖRGŐ, M. (1976). On the Erdős-Rényi increments and the P. Lévy modulus of continuity of a Kiefer process. In *Empirical Distributions and Processes.* Lecture Notes in Mathematics **566**, 17–32, Springer-Verlag.

CSISZÁR, I. (1975). I-divergence geometry of probability distributions and minimization problems. *Ann. Probability* **3** 146–158.

ERDŐS, P. and RÉNYI, A. (1970). On a new law of large numbers. *J. Analyse Math.* **23** 103–111.

ERDŐS, P. and RÉVÉSZ, P. (1975). On the length of the longest head-run. *Topics in Information Theory Keszthely (Hungary)* Coll. Math. Soc. János Bolyai **16** 219–228.

FU, J. C. (1973). On a theorem of Bahadur on the rate of convergence of point estimators. *Ann. Statist.* **1** 745–749.

FU, J. C. (1975). The rate of convergence of consistent point estimators. *Ann. Statist.* **3** 234–240.

GROENEBOOM, P., OOSTERHOFF, J. and RUYMGAART, F. H. (1976). Large deviation theorems for empirical probability measures. Preprint SW 49/76 of Math. Centrum, Amsterdam.

HÁJEK, J. (1974). Asymptotic sufficiency of the vector of ranks in the Bahadur sense. *Ann. Statist.* **2** 78–83.

HOADLEY, A. B. (1967). On the probability of large deviations of functions of several empirical cdf's. *Ann. Math. Statist.* **38** 360–382.

HOEFFDING, W. (1965). On probabilities of large deviations. *Proc. Fifth. Berkeley Symp. Math. Statist. Probability* **1** 203–219.

HWANG, T. Y. and KLOTZ, J. H. (1975). Bahadur efficiency of linear rank statistics for scale alternatives. *Ann. Statist.* **3** 947–954.

KLOTZ, J. H. (1965). Alternative efficiencies for signed rank tests. *Ann. Math. Statist.* **36** 1759–1766.

KLOTZ, J. H. (1967). Asymptotic efficiency of the Kolmogorov-Smirnov test. *J. Amer. Statist. Assoc.* **62** 932–938.

KILLEEN, T. J. and HETTMANSPERGER, T. P. (1972). Bivariate tests for location and their Bahadur efficiencies. *Ann. Math. Statist.* **43** 1507–1516.

KILLEEN, T. J., HETTMANSPERGER, T. P. and SIEVERS, G. L. (1972). An elementary theorem on the probability of large deviations. *Ann. Math. Statist.* **43** 181–192.

KOMLÓS, J. and TUSNÁDY, G. (1975). On sequences of "pure heads". *Ann. Probability* **3** 608–617.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1974). Weak convergence and embedding. *Limit theorems of probability theory, Keszthely (Hungary).* Coll. Math. Soc. János Bolyai **11** 149–165.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent rv's, and the sample df. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 111–131.

LITTELL, R. C. (1972). On the efficiency of a competitor of the two-sample Kolmogorov-Smirnov and Kuiper test. *Ann. Math. Statist.* **43** 1991–1992.

MOGUL'SKIĬ, A. A. (1977). Remarks on the large deviation for the ω^2-statistic (in Russian). *Teor. Verojatnost. i Primenen.* **23** 170–175.

NIKITIN, JA. JU. (1976). Relative asymptotic Bahadur efficiency of statistics based on an empirical distribution function (in Russian). *Dokl. Akad. Nauk SSSR* **231** 802–805.

RAO, J. S. (1972). Bahadur efficiencies of some tests for uniformity on the circle. *Ann. Math. Statist.* **43** 468–479.

SANOV, I. N. (1957). On the probability of large deviations of random variables (in Russian). *Mat. Sb.* **42** (84) 11–44. English translation in *Sel. Trans. Math. Statist. Prob.* **1** (1961) 213–244.

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

SIEVERS, G. L. (1969). On the probability of large deviations and exact slopes. *Ann. Math. Statist.* **40** 1908–1921.

SIEVERS, G. L. (1975). Multivariate probabilities of large deviations. *Ann. Statist.* **3** 897–905.

STONE, M. (1967). Extreme tail probabilities of the two-sample Wilcoxon statistic. *Biometrika* **54** 629–640.

STONE, M. (1968). Extreme tail probabilities for sampling without replacement and exact Bahadur efficiency of the two-sample normal scores test. *Biometrika* **55** 371–375.

STONE, M. (1974). Large deviations of empirical probability measures. *Ann. Statist.* **2** 362–366.

WOODWORTH, G. G. (1970). Large deviations and Bahadur efficiency of linear rank statistics. *Ann. Math. Statist.* **41** 251–283.

References added in proof.

CSÖRGŐ, S. (1979). Bahadur efficiency and Erdős-Rényi maxima. To be submitted to *Sankhyā Ser. A.*

DEO, C. M. (1976). A note on Erdős-Rényi law of large numbers. *Canad. Math. Bull.* **19** 497–500.

STEINEBACH, J. (1978a). On a necessary condition for the Erdős-Rényi law of large numbers. *Proc. Amer. Math. Soc.* **68** 97–100.

STEINEBACH, J. (1978b). A strong law of Erdős-Rényi type for cumulative processes in renewal theory. *J. Appl. Probability.* **15** 96–111.

STEINEBACH, J. (1979). Das Gesetz der Grossen Zahlen von Erdős-Rényi im wartemodell G/G/1. *Studia Sci. Math. Hungar.* To appear.

BOLYAI INSTITUTE
SZEGED UNIVERSITY
H-6720 SZEGED
ARADI VÉRTANUK TERE 1
HUNGARY