The Annals of Probability
1992, Vol. 20, No. 2, 987–1003

STRONG MODERATE DEVIATION THEOREMS

By Tadeusz Inglot,[1] Wilbert C. M. Kallenberg and Teresa Ledwina[2]

Technical University of Wrocław, University of Twente and Technical University of Wrocław

Strong moderate deviation theorems are concerned with relative errors in the tails caused by replacing the exact distribution function by its limiting distribution function. A new approach for deriving such theorems is presented using strong approximation inequalities. In this way a strong moderate deviation theorem is obtained for statistics of the form $T(\alpha_n)$, where T is a sublinear functional and α_n is the empirical process. The basic theorem is also applied on linear combinations of order statistics, leading to a substantial improvement of previous results.

1. Introduction. Let $\{T_n\}$ be a sequence of random variables (r.v.'s) for which it is difficult or even impossible to evaluate its exact distribution. If it is possible to show that T_n has a limiting distribution, the asymptotic distribution may be used to approximate the exact distribution. This simple principle has been very successful, leading to approximations which are so accurate that Kass (1988) writes: "Our finite world seems tied to asymptopia."

The mathematical theorems underlying the preceding reasoning are usually concerned with *differences* between the exact distribution function of T_n and its limiting distribution function. Two comments can be made at this point. First, such theorems are addressed to the middle part of the distribution. In the tails we should consider the *relative* error and not the absolute error in order to give a sound mathematical basis for the application of the approximation. Second, although the classical asymptotic theory gives only information about the absolute error, it turns out that in many cases the approximation remains of practical value in a large part of the tail of the distribution in the sense of relative error. Paraphrasing Kass we might say that our finite world is more tied to asymptopia than classical theorems state. Therefore one may expect the existence of a lot of theorems on vanishing relative error in the tails. They are called moderate deviation theorems. (Speaking here on moderate deviation theorems we include also the so called Cramér-type large deviation theorems.) They enlarge the mathematical basis of the approximations, showing that approximating the distribution function of T_n at a point x_n by

Received September 1989; revised January 1991.

[1]Supported by Grant CPBP 01.02.
[2]Part of this research was performed when T. Ledwina was visiting the University of Twente.

AMS 1980 *subject classifications*. 60F10, 62G30.
Key words and phrases. Moderate deviations, Cramér type large deviations, strong approximation, sublinear functional, seminorm, empirical process, linear combinations of order statistics, goodness-of-fit tests.

its limiting distribution function at x_n is appropriate for a much larger range of points x_n than provided by the classical theorems.

On the other hand, the limitations of the approximations also become clear, since usually in the far tails of the distribution the relative error will not tend to zero. Therefore moderate deviation theorems give a demarcation of the applicability of the approximation of exact distributions by limiting distributions.

Moderate deviation theorems are applied to statistics, for example, in comparing tests or estimators; see, for instance, Rubin and Sethuraman (1965b), Johnson and Truax (1974, 1978), Kallenberg (1983a, b), Jurečková, Kallenberg and Veraverbeke (1988).

In the situation where T_n is a sum of n independent r.v.'s, moderate deviation theorems were initiated by Cramér (1938) and refined by Petrov (1954, 1975), Book (1976), Rubin and Sethuraman (1965a) and Amosova (1972). Also for a lot of other statistics, moderate deviation theorems are obtained. We mention Malevich and Abdalimov (1979) and Vandemaele (1982) on U-statistics, Vandemaele and Veraverbeke (1982) on L-statistics, Jurečková, Kallenberg and Veraverbeke (1988) on M-estimators and Kallenberg (1982), Seoh, Ralescu and Puri (1985), Wu (1986) on rank statistics. Finally we mention the general results of Chaganty and Sethuraman (1985, 1986, 1988).

A very useful new tool in deriving the limiting distribution of T_n is the theory of strong approximations. As stated before, moderate deviation theorems enlarge the range on which the exact distribution may be replaced by the limiting distribution. In this paper the powerful tool of strong approximations is used to establish this enlargement. In a way this means that not only the results but also the proofs based on strong approximations of classical limit theorems are more powerful than presumed.

Theorems on the relative error caused by replacing the exact distribution of T_n by its limiting distribution are sometimes called *strong* large or moderate deviation theorems to distinguish them from first order results on log $P(T_n > x_n)$, where $\{T_n > x_n\}$ is a large or moderate deviation event. Moderate deviation theorems of the latter type, using the strong approximation method, are given in Inglot and Ledwina (1990, 1989).

The method of proof consists of three steps. First, T_n is replaced by T_n^*; T_n^* has the same distribution as T_n and is close to W_n, which has the limiting distribution of T_n (strong approximation). Second, the well-known Slutsky argument is applied. Finally we use a moderate deviation theorem for W_n. This general approach is made more precise in Section 2. In Section 3 the method is applied on statistics T_n of the form $T(\alpha_n)$, where T is a sublinear functional, continuous w.r.t. the uniform norm and α_n the empirical process. Specific examples are, for example, (weighted) Kolmogorov–Smirnov statistics, (generalized) Cramér–von Mises statistics and chi-square statistics. Other applications are given in Section 4, including, for example, the Cressie–Read class of multinomial goodness-of-fit tests.

The basic theorem of Section 2 may also be applied to obtain strong moderate deviation theorems for statistics T_n, which are close to other statistics W_n, for which already strong moderate deviation theorems exist. (For instance, W_n may be a sum of i.i.d. r.v.'s.) In Section 5, we use this approach to derive strong moderate deviation theorems for L-statistics. The natural bound $o(n^{1/6})$ for the x-range is obtained, thus improving previous results of Vandemaele and Veraverbeke (1982) and Seoh, Ralescu and Puri (1985).

2. Basic theorem. The three basic steps of the method are: strong approximation, Slutsky's argument and a moderate deviation result for the limiting distribution (cf. also Remark 2.1). The steps are worked out in this section.

Let $\{T_n\}$ be a sequence of r.v.'s. Suppose that there exists a probability space on which we have a sequence of r.v.'s $\{T_n^*\}$ and a sequence of r.v.'s $\{W_n\}$ such that T_n^* has the same distribution as T_n and for positive constants $c_1, \dots, c_4$,

$$P\big(n^{1/2}|T_n^* - W_n| > c_1 \log n + x\big) \le c_2 e^{-c_3 x} \tag{2.1}$$

for all $0 < x < c_4 n^{1/3}$ and $n \ge n_0$. This is the strong approximation part.

Further assume that there exist positive constants a and $c_5 > c_1 + \frac{1}{2}ac_3^{-1}$ such that

$$P(W_n > x) = \exp\{-\tfrac{1}{2}ax^2 + g(x)\}, \tag{2.2}$$

with

$$\lim_{n\to\infty} x_n^{-2} g(x_n) = 0 \tag{2.3}$$

and

$$\limsup_{n\to\infty} \{g(x_n - \varepsilon_n) - g(x_n)\} \le 0 \tag{2.4}$$

if $x_n \to \infty$, $x_n = o(n^{1/6})$ as $n \to \infty$ and $\varepsilon_n = n^{-1/2}c_5 \max\{\log n, x_n^2\}$.

Note that the distribution of W_n does not depend on n and represents the limiting distribution of T_n (cf., however, also Remark 2.1). First it will be shown that (2.2), (2.3) and (2.4) imply

$$\lim_{n\to\infty} \{g(x_n + \eta_n) - g(x_n)\} = 0 \tag{2.5}$$

if $x_n \to \infty$, $x_n = o(n^{1/6})$ as $n \to \infty$ and $|\eta_n| \le \varepsilon_n = n^{-1/2}c_5 \max(\log n, x_n^2)$. So (2.2), (2.3) and (2.5) describe the moderate deviation result for the limiting distribution. To prove (2.5), consider sequences $\{x_n\}$, $\{\eta_n\}$ satisfying the conditions in the line below (2.5). We have

$$\begin{aligned} 1 &\le \liminf_{n\to\infty} \frac{P(W_n > x_n - |\eta_n|)}{P(W_n > x_n)} \le \limsup_{n\to\infty} \frac{P(W_n > x_n - \varepsilon_n)}{P(W_n > x_n)} \\ &= \limsup_{n\to\infty} \exp\left\{ax_n\varepsilon_n - \frac{1}{2}a\varepsilon_n^2 + g(x_n - \varepsilon_n) - g(x_n)\right\} \le 1, \end{aligned}$$

since $x_n\varepsilon_n \to 0$, $\varepsilon_n^2 \to 0$ as $n \to \infty$ and (2.4) holds. Therefore

$$\lim_{n\to\infty} \frac{P(W_n > x_n - |\eta_n|)}{P(W_n > x_n)} = 1$$

and hence

$$\lim_{n\to\infty} \{g(x_n - |\eta_n|) - g(x_n)\} = 0. \tag{2.6}$$

Replacing now x_n by $\max(x_n, x_n + \eta_n)$ in (2.6) yields (2.5). [Note that the sequences $\{\max(x_n, x_n + \eta_n)\}$, $\{\eta_n\}$ satisfy the conditions required for (2.6).]

Using Slutsky's argument we will now derive the basic result. Let, as above, $\{x_n\}$ be a sequence of real numbers satisfying

$$\lim_{n\to\infty} x_n = \infty, \qquad \lim_{n\to\infty} x_n n^{-1/6} = 0.$$

Let $\varepsilon_n = n^{-1/2}c_5 \max\{\log n, x_n^2\}$. In view of (2.1) and (2.2), we have for all $n \ge n_0$,

$$\begin{aligned} &\frac{P(|T_n^* - W_n| > \varepsilon_n)}{P(W_n > x_n)} \\ &\quad \le c_2 \exp\left\{-c_3\left(\varepsilon_n n^{1/2} - c_1 \log n\right) + \frac{1}{2}ax_n^2 - g(x_n)\right\}, \end{aligned} \tag{2.7}$$

which tends to zero, since $c_3\varepsilon_n n^{1/2}$ dominates the other terms as $n \to \infty$. Combining

$$\begin{aligned} &P(W_n > x_n + \varepsilon_n) - P(|T_n^* - W_n| > \varepsilon_n) \\ &\quad \le P(T_n > x_n) \le P(W_n > x_n - \varepsilon_n) + P(|T_n^* - W_n| > \varepsilon_n) \end{aligned}$$

with (2.2), (2.3), (2.5) and (2.7) now yields

$$\lim_{n\to\infty} \frac{P(T_n > x_n)}{P(W_n > x_n)} = 1$$

and thus the following theorem has been proved.

THEOREM 2.1. *Let $\{T_n\}$ be a sequence of r.v.'s satisfying* (2.1)–(2.4), *where T_n^* and W_n are as defined above. Let $\{x_n\}$ be a sequence of real numbers satisfying* $\lim_{n\to\infty} x_n = \infty$, $\lim_{n\to\infty} x_n n^{-1/6} = 0$. *Then*

$$\lim_{n\to\infty} \frac{P(T_n > x_n)}{P(W_n > x_n)} = 1. \tag{2.8}$$

Note that (2.2)–(2.4) hold in the important case that W_n has a standard normal distribution, since

$$1 - \Phi(x) = (2\pi)^{-1/2} e^{-x^2/2} x^{-1}(1 + o(1)) \quad \text{as } x \to \infty, \tag{2.9}$$

where Φ denotes the standard normal distribution function.

REMARK 2.1. A straightforward generalization of the basic theorem is obtained if in (2.2)–(2.4), g is replaced by g_n. In this case the distribution of W_n does not represent the limiting distribution of T_n, its distribution may depend on n. Now W_n is simply close to T_n in the sense of inequality (2.1) (not necessarily implied by a direct application of a standard strong approximation theorem) and for W_n, (2.2)–(2.4) hold with g replaced by g_n. The latter usually follows from the fact that for W_n, already a strong moderate deviation theorem is established. For instance, W_n may be a (standardized) sum of i.i.d. r.v.'s. Then (2.2)–(2.4) with g replaced by g_n follow immediately from, for example, Theorem 1 of Feller [(1971), page 549] and (2.9). Applications of this kind of the modified basic theorem are given in Section 5.

3. Sublinear functionals, seminorms. Let T be a sublinear functional on $D[0,1]$ (i.e., $T(x+y) \le T(x) + T(y)$, $T(\lambda x) = \lambda T(x)$ for all $\lambda \ge 0$ and $x, y \in D[0,1]$), continuous w.r.t. the uniform norm. Note that every seminorm is a sublinear functional. Consider the statistic $T_n = T(\alpha_n)$, where α_n is the uniform empirical process. Denote by B a Brownian bridge process. It is well known that T_n converges in distribution to $T(B)$. The following theorem shows that this approximation remains valid in the tails of the distribution.

THEOREM 3.1. *For each sequence $\{\rho(n)\}$ with $\lim_{n\to\infty} \rho(n) = 0$ we have, uniformly in the region $0 \le x \le \rho(n) n^{1/6}$,*

$$P(T_n > x) = P(T(B) > x)\{1 + o(1)\} \tag{3.1}$$

as $n \to \infty$.

PROOF. Since T is a sublinear functional and continuous w.r.t. the uniform norm, there exists a constant $c_6 > 0$ such that

$$|T(x) - T(y)| \le c_6 \sup_{0 \le t \le 1} |x(t) - y(t)| \tag{3.2}$$

for all $x, y \in D[0,1]$. In view of the KMT-inequality [cf. Theorem 4.4.1 on p. 133 of Csörgő and Révész (1981)] there exist a probability space, sequences of processes $\{\alpha_n^*\}$ and Brownian bridges $\{B_n\}$ defined on it such that α_n^* has the same distribution as α_n and for some positive constants c_7, c_8, c_9 and all n and x,

$$P\Big(n^{1/2} \sup_{0 \le t \le 1} |\alpha_n^*(t) - B_n(t)| > c_7 \log n + x\Big) \le c_8 e^{-c_9 x}. \tag{3.3}$$

Combination of (3.2) and (3.3) yields (2.1) with $T_n^* = T(\alpha_n^*)$ and $W_n = T(B_n)$. Application of Theorem 5.2 in Borell (1975) gives (2.2) and (2.3). Denoting $F(x) = P(T(B) \le x)$, it follows by Ehrhard (1983) that $\Phi^{-1}(F(x))$ is concave. Therefore there exists a positive constant c_{10} such that

$$\Phi^{-1}(F(x + \varepsilon)) \le c_{10}\varepsilon + \Phi^{-1}(F(x)) \tag{3.4}$$

for all $x \ge x_0 > 0$ and $\varepsilon > 0$. Let $\{x_n\}$ be a sequence of real numbers satisfying

$\lim_{n\to\infty} x_n = \infty$, $\lim_{n\to\infty} x_n n^{-1/6} = 0$, and let $\varepsilon_n = n^{-1/2} c_{11} \max\{\log n, x_n^2\}$ for some constant $c_{11} > 0$. Writing $z_n = \Phi^{-1}(F(x_n))$ and inserting $x = x_n - \varepsilon_n$, $\varepsilon = \varepsilon_n$ in (3.4), we obtain

$$\frac{1 - F(x_n - \varepsilon_n)}{1 - F(x_n)} \le \frac{1 - \Phi(z_n - c_{10}\varepsilon_n)}{1 - \Phi(z_n)}. \tag{3.5}$$

Since $x_n \to \infty$, $\varepsilon_n \to 0$ and $\varepsilon_n x_n \to 0$, (2.3) implies $\varepsilon_n z_n \to 0$ as $n \to \infty$. Moreover, $z_n \to \infty$ and hence

$$\frac{1 - \Phi(z_n - c_{10}\varepsilon_n)}{1 - \Phi(z_n)} \to 1 \quad \text{as } n \to \infty. \tag{3.6}$$

On the other hand,

$$\limsup_{n\to\infty} \log\left\{\frac{1 - F(x_n - \varepsilon_n)}{1 - F(x_n)}\right\} = \limsup_{n\to\infty} \{g(x_n - \varepsilon_n) - g(x_n)\}$$

and therefore (3.5) and (3.6) imply (2.4). Application of Theorem 2.1, together with the convergence in distribution of T_n to $T(B)$, completes the proof. □

REMARK 3.1. By inspection of the proof of Theorem 3.1 and (2.5), it is seen that

$$\lim_{n\to\infty} \frac{P(T(B) > x_n + \eta_n)}{P(T(B) > x_n)} = 1$$

for all sequences $\{x_n\}$, $\{\eta_n\}$ with $x_n \to \infty$, $x_n = o(n^{1/6})$ as $n \to \infty$ and $|\eta_n| \le n^{-1/2} c_{12} \max\{\log n, x_n^2\}$ for some constant $c_{12} > 0$. This result is used in the proof of Proposition 4.1.

Next we present some examples where Theorem 3.1 can be applied immediately, thus showing that in these cases the replacement of the exact distribution by the limiting distribution is valid in the sense of vanishing relative error in a much larger range than provided by the convergence in distribution. *Tacitly it is assumed that the possible weights occurring in the examples make the functionals not identically zero.* In all of the examples it is easily seen that T is a sublinear functional on $D[0, 1]$ [except for (b) all the functionals are seminorms], continuous w.r.t. the uniform norm. So the moderate deviation result (3.1) holds in all of the examples.

(a) (Weighted Kolmogorov–Smirnov statistics)

$$T_{\text{WKS}}(x) = \sup_{0 \le t \le 1} \{w(t)|x(t)|\},$$

where w is a nonnegative and bounded function.

(b) (One-sided weighted Kolmogorov–Smirnov statistics)

$$T_{\text{OWKS}}(x) = \sup_{0 \le t \le 1} \{w(t)x(t)\},$$

where w is a nonnegative and bounded function.

(c) (Generalized Cramér–von Mises statistics)

$$T_{\mathrm{GCvM}}(x) = \left\{ \int_0^1 |x(t)|^r w(t)\, dt \right\}^{1/r},$$

where $r \geq 1$ and w is a nonnegative, Lebesgue-measurable function, which is integrable on (0, 1).

(d) (Chi-square statistics)

$$T_\chi(x) = \left\{ \sum_{i=1}^{k} \frac{[x(a_i) - x(a_{i-1})]^2}{a_i - a_{i-1}} \right\}^{1/2},$$

or more generally

$$T_{\mathrm{G}\chi}(x) = \left\{ \sum_{i=1}^{k} w_i |x(a_i) - x(a_{i-1})|^r \right\}^{1/r},$$

where $r \geq 1$, $w_i \geq 0$, $i = 1, \dots, k$, and $0 = a_0 < a_1 < \cdots < a_{k-1} < a_k = 1$.

(e) (Watson statistic)

$$T_{\mathrm{W}}(x) = \left\{ \int_0^1 \left[x(t) - \int_0^1 x(v)\, dv \right]^2 dt \right\}^{1/2},$$

or more generally

$$T_{\mathrm{GW}}(x) = \left\{ \int_0^1 w(t) |x(t) - \int_0^1 x(v)\, dv|^r\, dt \right\}^{1/r},$$

where $r \geq 1$ and w is a nonnegative, Lebesgue-measurable function, which is integrable on (0, 1).

(f) (Quadratic statistics)

$$T_{\mathrm{Q}}(x) = \left\{ \sum_{i=1}^{\infty} \lambda_i [L_i(x)]^2 \right\}^{1/2},$$

or more generally

$$T_{\mathrm{GQ}}(x) = \left\{ \sum_{i=1}^{\infty} \lambda_i |L_i(x)|^r \right\}^{1/r},$$

where $r \geq 1$, $\lambda_i \geq 0$, $i = 1, 2, \dots$, $\{L_i\}$ is a sequence of bounded (w.r.t. the uniform norm) linear functionals with norms $\{\|L_i\|\}$ and $\sum_{i=1}^{\infty} \lambda_i \|L_i\|^r < \infty$. In view of the Bessel inequality, for T_{Q}, the latter condition may be replaced by $\sup_i \lambda_i < \infty$ if $L_i(x) = \int_{(0,1)} x(t) \phi_i(t)\, dt$ with $\phi_1, \phi_2, \dots$ an orthonormal basis in $L_2(0, 1)$. Note that $T_{\mathrm{G}\chi}$ in (d) is a special case of T_{GQ}, choosing $\lambda_i = w_i$ and $L_i(x) = x(a_i) - x(a_{i-1})$ for $i = 1, \dots, k$ and $\lambda_i = 0$ for $i > k$. Further special cases are Neyman's smooth tests for uniformity, taking T_{Q} with $\lambda_i = 1$ for $i = 1, \dots, k$, $\lambda_i = 0$ for $i > k$ and $L_i(x) = \int_{(0,1)} x(t) \pi_i'(t)\, dt$ with $\{\pi_i\}$ the normalized Legendre polynomials and Neuhaus' (1988) quadratic goodness-of-fit tests, taking T_Q with $L_i(x) = \int_{(0,1)} x(t)\sqrt{2}\,\sin(i\pi t)\,dt$ and $\lambda_i = \{\int_{-1}^{1} K(t)\cos(\pi iat)\,dt\}\pi^2 i^2$ with kernel K and bandwidth a, provided that $\lambda_i \geq 0$ and $\sup_i \lambda_i < \infty$ (which holds, e.g., for the recommended Parzen-2 kernel K and all bandwidths $a \in (0,1]$).

4. Further applications and generalizations. In this section we generalize the approach of the previous section to statistics, which are close to a continuous sublinear functional.

EXAMPLE 4.1 (Multinomial goodness-of-fit tests). Consider

$$\sum_{i=1}^{k} np_i f\left(\frac{\alpha_n(a_i) - \alpha_n(a_{i-1})}{\sqrt{n}\, p_i}\right) - nf(0), \tag{4.1}$$

where $0 = a_0 < a_1 < \cdots < a_{k-1} < a_k = 1$, $p_i = a_i - a_{i-1}$, $i = 1, \dots, k$, f is a measurable function defined on $(-1, \infty)$ such that for some $\delta > 0$, $f \in C^2$ on $[-\delta, \delta]$ with $f''(0) > 0$ and such that f''' exists and is bounded on $[-\delta, \delta]$. Note that (4.1) contains the Cressie–Read (1984) class, which class in turn includes Pearson's chi-square statistic, the likelihood ratio statistic, the Freeman–Tukey statistic and Neyman's modified chi-square statistic.

The statistic (4.1) is of the following form:

$$V_n(\alpha_n) = \{T(\alpha_n)\}^2 + R_n(\alpha_n), \tag{4.2}$$

where T is a sublinear functional on $D[0,1]$, continuous w.r.t. the uniform norm and $|R_n(\alpha_n)| \leq c_{13} n^{-1/2} \|\alpha_n\|^3$ on the event $\{\|\alpha_n\| \leq c_{14} n^{1/6}\}$ for some positive constants c_{13}, c_{14}. Here $\|\cdot\|$ denotes the supremum norm. [That (4.1) is of the form (4.2) follows by Taylor expansion of f around 0 and inserting T_χ of Example (d) of Section 3 for T.]

PROPOSITION 4.1. *If $V_n(\alpha_n)$ is of the form* (4.2), *then for each sequence* $\{\rho(n)\}$ *with* $\lim_{n\to\infty} \rho(n) = 0$ *we have, uniformly in the region* $0 \leq x \leq \rho(n) n^{1/6}$,

$$P\big(V_n(\alpha_n) > x^2\big) = P(T(B) > x)\{1 + o(1)\} \tag{4.3}$$

as $n \to \infty$.

PROOF. Let $\{x_n\}$ be a sequence of real numbers satisfying $\lim_{n\to\infty} x_n = \infty$, $\lim_{n\to\infty} x_n n^{-1/6} = 0$ and let $\varepsilon_n = n^{-1/2} c_{15} \max\{\log n, x_n^2\}$ for some sufficiently large constant $c_{15} > 0$. We have for $n \geq n_0$,

$$\frac{P\big(V_n(\alpha_n) > x_n^2\big)}{P(T(B) > x_n)} \leq \frac{P(T(\alpha_n) > x_n - \varepsilon_n)}{P(T(B) > x_n)} + \frac{P(R_n(\alpha_n) > x_n \varepsilon_n)}{P(T(B) > x_n)}. \tag{4.4}$$

Theorem 3.1 and Remark 3.1 now imply

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{P(T(\alpha_n) > x_n - \varepsilon_n)}{P(T(B) > x_n)} \\
(4.5)\qquad &= \lim_{n\to\infty} \frac{P(T(\alpha_n) > x_n - \varepsilon_n)}{P(T(B) > x_n - \varepsilon_n)} \lim_{n\to\infty} \frac{P(T(B) > x_n - \varepsilon_n)}{P(T(B) > x_n)} = 1.
\end{aligned}
$$

Define α_n^* and B_n as in the proof of Theorem 3.1; then we obtain

$$
\begin{aligned}
&\frac{P(R_n(\alpha_n) > x_n\varepsilon_n)}{P(T(B) > x_n)} \\
&= \frac{P\left(R_n(\alpha_n^*) > x_n\varepsilon_n, \|\alpha_n^* - B_n\| \le c_{16}, \|B_n\| < (1/2)c_{14}n^{1/6}\right)}{P(T(B) > x_n)} + o(1),
\end{aligned}
$$

where c_{16} is a positive constant. Hence

$$
\begin{aligned}
&\limsup_{n\to\infty} \frac{P(R_n(\alpha_n) > x_n\varepsilon_n)}{P(T(B) > x_n)} \\
(4.6)\qquad &= \limsup_{n\to\infty} \frac{P\left(\|B_n\| > \left\{(2c_{13})^{-1}x_n\varepsilon_n n^{1/2}\right\}^{1/3}\right)}{P(T(B) > x_n)} = 0,
\end{aligned}
$$

since $(x_n\varepsilon_n n^{1/2})^{2/3}$ dominates x_n^2 if c_{15} is chosen large enough. Combination of (4.4), (4.5) and (4.6) yields

$$
\limsup_{n\to\infty} \frac{P\left(V_n(\alpha_n) > x_n^2\right)}{P(T(B) > x_n)} \le 1.
$$

Similarly one proves

$$
\liminf_{n\to\infty} \frac{P\left(V_n(\alpha_n) > x_n^2\right)}{P(T(B) > x_n)} \ge 1.
$$

The proof of (4.3) is completed by noting that $V_n(\alpha_n)$ converges in distribution to $T(B)$. □

Example 4.1 (Continued). By application of Proposition 4.1, the strong moderate deviation result holds for the multinomial goodness-of-fit tests given by (4.1) and hence in particular for the Cressie–Read class.

5. *L*-statistics. Linear combinations of order statistics (or L-statistics) are statistics of the form

$$
(5.1)\qquad V_n = \sum_{i=1}^{n} c_{in} X_{i:n},
$$

where the weights c_{in}, $i = 1, \dots, n$; $n = 1, 2, \dots,$ are real numbers and where $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ are the order statistics of a sequence $X_1, \dots, X_n$

of i.i.d. r.v.'s with common d.f. F. Here we concentrate on L-statistics with weights

$$c_{in} = \int_{(i-1)/n}^{i/n} J(s)\, ds, \tag{5.2}$$

where J is a real-valued function on $(0, 1)$. For L-statistics with weights close to the weights given by (5.2), we refer to Remark 5.4.

Our first theorem on the large deviation behaviour of L-statistics concerns untrimmed weight functions J, while our second theorem deals with weight functions J vanishing outside some interval (a, b), $0 < a < b < 1$.

THEOREM 5.1. *Let V_n be given by* (5.1) *and* (5.2), *where J is Lipschitz of order* 1 *on* $(0, 1)$. *Let* $E \exp(t|X_1|^{1/2}) < \infty$ *for some* $t > 0$. *Define*

$$\mu = \int_0^1 J(s) F^{-1}(s)\, ds, \tag{5.3}$$

where

$$F^{-1}(s) = \inf\{x\colon F(x) \ge s\} \tag{5.4}$$

and let

$$\sigma^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(F(x))J(F(y)) \times\{\min(F(x), F(y)) - F(x)F(y)\}\, dx\, dy > 0. \tag{5.5}$$

Then uniformly in the region $-A \le x \le \rho(n) n^{1/6}$ *with* $A > 0$ *and* $\lim_{n\to\infty} \rho(n) = 0$,

$$P\big(n^{1/2}(V_n - \mu)/\sigma > x\big) = \{1 - \Phi(x)\}\{1 + o(1)\} \tag{5.6}$$

as $n \to \infty$.

Before proving the result, we compare it with earlier results of this type in the literature, given by Vandemaele and Veraverbeke (1982) and Seoh, Ralescu and Puri (1985). In both papers the natural bound $o(n^{1/6})$ for the x-range is only obtained for bounded r.v.'s X_i. In Theorem 5.1 the restrictive condition of boundedness is replaced by the much weaker classical condition for sums of i.i.d. r.v.'s, that is, the existence of the moment generating function of $|X_1|^{1/2}$ at some $t > 0$. [Note that the case of sums of i.i.d. r.v.'s is a special case of Theorem 5.1, obtained by choosing $J(s) \equiv 1$.] So, the natural bound $o(n^{1/6})$ for the x-range is established under a natural condition. After the present paper was submitted for publication, R. Norvaiša informed us about a paper of Bentkus and Zitikis, proving independently Theorem 5.1 in a different way.

PROOF OF THEOREM 5.1. Following Helmers (1978, 1981) we have

$$W_{n-} \le V_n \le W_{n+} \quad \text{a.e.}$$

with

$$W_{n+} = n^{-1} \sum_{i=1}^{n} h(U_i) + Kn^{-1} \int_0^1 \alpha_n^2(s)\, dF^{-1}(s),$$

$$W_{n-} = n^{-1} \sum_{i=1}^{n} h(U_i) - Kn^{-1} \int_0^1 \alpha_n^2(s)\, dF^{-1}(s),$$

where

$$(5.7)\quad h(U_i) = \mu + \int_{(0, U_i)} sJ(s)\, dF^{-1}(s) - \int_{[U_i, 1)} (1-s)J(s)\, dF^{-1}(s),$$

where K is a positive constant, $U_1, U_2, \dots$ independent uniform $(0,1)$ r.v.'s and α_n the uniform empirical process. By boundedness of J, it follows that $|h(u)| \le c_{17} + c_{18}|F^{-1}(u)|$ for some constants c_{17} and c_{18} and hence the moment generating function of $|h(U_i)|^{1/2}$ is finite at some $t > 0$. Since $Eh(U_i) = \mu$ and var $h(U_i) = \sigma^2$ it now follows by Linnik's theorem [cf., e.g., Petrov (1975), page 251] that

$$(5.8)\quad P\left(n^{-1/2}\left\{\sum_{i=1}^{n} (h(U_i) - \mu)/\sigma\right\} > x\right) = \{1 - \Phi(x)\}\{1 + o(1)\}$$

uniformly in the region $-A \le x \le \rho(n)n^{1/6}$. Next consider the functional

$$T(x) = \left\{\int_0^1 x^2(s)\, dF^{-1}(s)\right\}^{1/2} = \left\{\int_0^1 \left\{\frac{x(s)}{w(s)}\right\}^2 w^2(s)\, dF^{-1}(s)\right\}^{1/2},$$

where

$$w(s) = h(s(1-s)) \quad \text{and} \quad h(s) = \log^{-\delta}(1/s) \quad \text{for some } \delta > 1.$$

By the Markov inequality we have for all $x > 0$, $t > 0$,

$$P(|X_1| \ge x) = P\left(\exp(t|X_1|^{1/2}) \ge \exp(tx^{1/2})\right) \le e^{-tx^{1/2}} E e^{t|X_1|^{1/2}}.$$

Since $Ee^{t|X_1|^{1/2}} < \infty$ for some $t > 0$, it follows that

$$|\log\{F(x)[1 - F(x)]\}| \ge c_{19} + c_{20}|x|^{1/2} \quad \text{for all } x \in \mathbb{R},$$

for some constants c_{19}, c_{20} with $c_{20} > 0$. Hence

$$(5.9)\quad \int_0^1 w^2(s)\, dF^{-1}(s) < \infty$$

for all $\delta > 1$. In view of (5.9), T satisfies the weighted Lipschitz condition (1.2) of Inglot and Ledwina (1989), that is, for some $c > 0$

$$|T(x) - T(y)| \le c \sup_{0 \le t \le 1} \{|x(t) - y(t)|/w(t)\}.$$

Application of Proposition 3.3 of Inglot and Ledwina (1989) now yields

$$\lim_{n \to \infty} (nx_n^2)^{-1} \log P\big(T(\alpha_n) \ge x_n\sqrt{n}\big) = -\frac{a}{2} \tag{5.10}$$

for some $a > 0$ and $x_n = O(n^{-\gamma})$ with $\gamma > (\delta - 1)/(2\delta - 1)$. Choosing $\delta < 2$ we may take $\gamma = \frac{1}{3}$ and hence $x_n\sqrt{n} = O(n^{1/6})$. Since

$$\left| n^{1/2}(V_n - \mu)/\sigma - n^{-1/2}\left\{ \sum_{i=1}^{n} (h(U_i) - \mu) \right\}/\sigma \right| \le n^{-1/2}(K/\sigma)T^2(\alpha_n),$$

(5.10) implies (2.1) if we define

$$T_n^* = n^{1/2}(V_n - \mu)/\sigma, \qquad W_n = n^{-1/2}\left\{ \sum_{i=1}^{n} [h(U_i) - \mu] \right\}/\sigma.$$

By (5.8), we have (2.2)–(2.4) with g replaced by g_n and hence (5.6) follows from the modified version of Theorem 2.1, given in Remark 2.1, taking $T_n = T_n^*$. □

REMARK 5.1. The condition of the finiteness of the moment generating function of $|X_i|^{1/2}$ at some $t > 0$ is used two times in the proof of Theorem 5.1: First to ensure the finiteness of the moment generating function of $|h(U_i)|^{1/2}$ at some $t > 0$ and second to prove (5.9) for some $\delta < 2$. For the latter in fact we only need the condition

$$Ee^{t|X|^\alpha} < \infty \quad \text{for some } t > 0 \text{ and } \alpha > \tfrac{1}{4}. \tag{5.11}$$

This can be seen as follows. By the Markov inequality we have for all $x > 0$,

$$P(|X_1| \ge x) \le \exp(-tx^\alpha)E\{\exp(t|X_1|^\alpha)\},$$

and hence

$$|\log\{F(x)[1 - F(x)]\}| \ge c_{21} + c_{22}|x|^\alpha \quad \text{for all } x \in \mathbb{R}$$

for some constants c_{21}, c_{22} with $c_{22} > 0$. Therefore (5.9) holds if $2\alpha\delta > 1$. Since $\alpha > \frac{1}{4}$, (5.9) thus holds for some $\delta < 2$. So, if the moment generating function of $|h(U_i)|^{1/2}$ is finite at some $t > 0$ (for instance if h is bounded), then the condition on the moment generating function of $|X_i|^{1/2}$ in Theorem 5.1 may be replaced by (5.11).

In case of trimmed weight functions J we do not even need the condition on the moment generating function as is seen in the following theorem.

THEOREM 5.2. *Let V_n be given by* (5.1) *and* (5.2), *where J equals* 0 *outside* $[a, b]$ *and is Lipschitz of order* 1 *on* (a, b), $0 < a < b < 1$. *Let μ and $\sigma^2 > 0$ be as in Theorem* 5.1. *Assume*

$$F^{-1}(a+\varepsilon) - F^{-1}(a-\varepsilon) = O(\varepsilon), \quad F^{-1}(b+\varepsilon) - F^{-1}(b-\varepsilon) = O(\varepsilon) \tag{5.12}$$

as $\varepsilon \to 0$. *Then uniformly in the region* $-A \le x \le \rho(n)n^{1/6}$ *with* $A > 0$ *and* $\lim_{n\to\infty} \rho(n) = 0$,

$$P\big(n^{1/2}(V_n - \mu)/\sigma > x\big) = \{1 - \Phi(x)\}\{1 + o(1)\} \tag{5.13}$$

as $n \to \infty$.

PROOF. Inspection of Helmers' (1978, 1981) construction shows that V_n may be written as

$$V_n = n^{-1}\sum_{i=1}^{n} h(U_i) - \int_0^1 \int_s^{\Gamma_n(s)} \{J(t) - J(s)\}\,dt\,dF^{-1}(s), \tag{5.14}$$

where $\Gamma_n(s)$ denotes the empirical distribution function based on i.i.d. r.v.'s $U_1, \dots, U_n$ with a uniform (0, 1) distribution and h is given by (5.7). Since J vanishes outside $[a, b]$ and is bounded, it follows that h is bounded and hence (5.8) holds, for example, by Theorem 1 of Feller [(1971), page 549].

In view of (5.12), there exists $\varepsilon_0 > 0$ such that

$$F^{-1}(a+\varepsilon) - F^{-1}(a-\varepsilon) \le c_{23}\varepsilon, \quad F^{-1}(b+\varepsilon) - F^{-1}(b-\varepsilon) \le c_{23}\varepsilon \tag{5.15}$$

for some constant $c_{23} > 0$ and all $0 < \varepsilon < \varepsilon_0$. Without loss of generality, let $\varepsilon_0 < \min\{a, 1 - b, \frac{1}{2}(b - a)\}$. Let

$$L_n = \sup_{0 \le s \le 1} |\Gamma_n(s) - s|. \tag{5.16}$$

On the event $\{L_n < \varepsilon_0\}$ it will be shown that for some $c_{24} > 0$,

$$\int_0^1 \int_s^{\Gamma_n(s)} |J(t) - J(s)|\,dt\,dF^{-1}(s) \le c_{24} L_n^2. \tag{5.17}$$

Therefore we split up the interval (0, 1) in five parts: $(0, a - L_n)$, $[a - L_n, a + L_n]$, $(a + L_n, b - L_n)$, $[b - L_n, b + L_n]$, $(b + L_n, 1)$. If $s \in (0, a - L_n)$ or $s \in (b + L_n, 1)$, then both s and $\Gamma_n(s)$ are outside $[a, b]$ and hence $J(s) = 0$ and $J(t) = 0$ for t between s and $\Gamma_n(s)$. In view of (5.15) and the boundedness of J, we have

$$\int_{a-L_n}^{a+L_n} \int_s^{\Gamma_n(s)} |J(t) - J(s)|\,dt\,dF^{-1}(s) \le 2 \sup_{0<s<1} |J(s)| L_n c_{25} L_n = c_{26} L_n^2.$$

Similarly, the integral over $[b - L_n, b + L_n]$ is estimated. On $(a + L_n, b - L_n)$ both s and $\Gamma_n(s)$ are in (a, b) and therefore the Lipschitz condition may be applied, leading to

$$\int_{a+L_n}^{b-L_n}\int_s^{\Gamma_n(s)}|J(t) - J(s)|\,dt\,dF^{-1}(s) \le \int_a^b c_{27}\int_s^{\Gamma_n(s)}|t - s|\,dt\,dF^{-1}(s) \le c_{28}L_n^2$$

for some constants $c_{27}, c_{28} > 0$. This completes the proof of (5.17). Defining $T_n^* = n^{1/2}(V_n - \mu)/\sigma$, $W_n = n^{-1/2}\{\sum_{i=1}^n[h(U_i) - \mu]\}/\sigma$, the Dvoretzky–Kiefer–Wolfowitz inequality implies

$$P\big(n^{1/2}|T_n^* - W_n| > c_1 \log n + x\big)$$
$$\le P(L_n \ge \varepsilon_0) + P\Big(n^{1/2}L_n > \big\{\sigma c_{24}^{-1}(c_1 \log n + x)\big\}^{1/2}\Big) \le c_2 e^{-c_3 x}$$

for some positive constants $c_1, \dots, c_4$, all $0 < x < c_4 n^{1/3}$ and $n \ge n_0$, thus establishing (2.1). The modified version of Theorem 2.1, given in Remark 2.1, now yields the result. □

REMARK 5.2. A related result on trimmed L-statistics is given in Callaert, Vandemaele and Veraverbeke (1982). Due to a different approach their conditions are not quite comparable with the conditions of Theorem 5.2.

REMARK 5.3. Note that in Theorem 5.1 and 5.2, it is not required that the distribution function of X_i is continuous. The results even hold in the discrete case.

REMARK 5.4. Theorems 5.1 and 5.2 continue to hold if c_{in} is not exactly of the form (5.2) but close to it. For instance, let

$$V_n^* = \sum_{i=1}^n d_{in} X_{i:n} \tag{5.18}$$

with

$$\max_{1\le i\le n}\left| nd_{in} - n\int_{(i-1)/n}^{i/n} J(s)\,ds\right| = O(n^{-1/2}) \quad \text{as } n \to \infty. \tag{5.19}$$

[Condition (5.19) is written in this form to compare it with condition (∗) in Vandemaele and Veraverbeke (1982). Their $c_{in} = nd_{in}$, implying that condition (5.19) is weaker than their condition (∗).] Assume that the conditions of Theorem 5.1 hold. Define

$$J_n(s) = J(s) + nd_{in} - n\int_{(i-1)/n}^{i/n} J(t)\,dt, \qquad \frac{i-1}{n} \le s < \frac{i}{n}, \qquad i = 1, \dots, n.$$

Then

$$V_n^* = \sum_{i=1}^n \int_{(i-1)/n}^{i/n} J_n(s)\,ds\,X_{i:n}$$

and

$$\sup_{0<s<1} |J_n(s) - J(s)| = O(n^{-1/2}) \quad \text{as } n \to \infty.$$

In view of Helmers' (1978, 1981) construction V_n^* may be written as [cf. (5.14)]

$$V_n^* = \mu_n - \int_0^1 J_n(s)\{\Gamma_n(s) - s\}\, dF^{-1}(s)$$
$$- \int_0^1 \int_s^{\Gamma_n(s)} \{J_n(t) - J_n(s)\}\, dt\, dF^{-1}(s),$$

where

$$\mu_n = \int_0^1 J_n(s) F^{-1}(s)\, ds$$

and $\Gamma_n(s)$ is the empirical distribution function based on a sample of size n from the uniform (0, 1) distribution. Hence [cf. (5.1), (5.2) and (5.9)]

$$|(V_n^* - \mu_n) - (V_n - \mu)|$$
$$= \left| \int_0^1 \{J(s) - J_n(s)\}\{\Gamma_n(s) - s\}\, dF^{-1}(s) \right.$$
$$\left. + \int_0^1 \int_s^{\Gamma_n(s)} \{J(t) - J_n(t)\} - \{J(s) - J_n(s)\}\, dt\, dF^{-1}(s) \right|$$
$$\leq c_{29} n^{-1/2} \sup_{0<s<1} \left\{ |\Gamma_n(s) - s|\, |\log\{s(1-s)\}|^{\delta} \right\}$$

for some constant $c_{29} > 0$ and all $\delta > 1$. Defining $T_n^* = n^{1/2}(V_n^* - \mu_n)/\sigma$, $W_n = n^{1/2}(V_n - \mu)/\sigma$, application of Proposition 3.3 of Inglot and Ledwina (1989) now yields (2.1). By Theorem 5.1 we have (2.2)–(2.4) with g replaced by g_n and hence the modified version of Theorem 2.1, given in Remark 2.1, implies

$$P\left(n^{1/2}(V_n^* - \mu_n)/\sigma > x\right) = \{1 - \Phi(x)\}\{1 + o(1)\} \quad \text{as } n \to \infty,$$

uniformly in the region $-A \leq x \leq \rho(n) n^{1/6}$ with $A > 0$ and $\lim_{n\to\infty} \rho(n) = 0$.

Acknowledgments. The authors thank T. Żak for stimulating discussions and the referee for helpful comments.

REFERENCES

AMOSOVA, N. N. (1972). On limit theorems for the probabilities of moderate deviations. *Vestnik Leningrad Univ. Math.* **13** 5–14 (in Russian).

BENTKUS, V. and ZITIKIS, R. (1990). Probabilities of large deviations for L-statistics. *Litovsk. Mat. Sb.* **30** 479–488.

BOOK, S. A. (1976). The Cramér–Feller–Petrov large deviation theorem for triangular arrays. Technical report, California State College.

BORELL, C. (1975). The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* **30** 207–216.

CALLAERT, H., VANDEMAELE, M. and VERAVERBEKE, N. (1982). A Cramér type large deviation theorem for trimmed linear combinations of order statistics. *Comm. Statist. A—Theory Methods* **11** 2689–2698.

CHAGANTY, N. R. and SETHURAMAN, J. (1985). Large deviation local limit theorems for arbitrary sequences of random variables. *Ann. Probab.* **13** 97–114.

CHAGANTY, N. R. and SETHURAMAN, J. (1986). Multi-dimensional large deviation local limit theorems. *J. Multivariate Anal.* **20** 190–204.

CHAGANTY, N. R. and SETHURAMAN, J. (1988). Strong large deviation and local limit theorems. Technical Report, Florida State Univ.

CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Sci. Ind.* **736** 5–23.

CRESSIE, N. and READ, T. R. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B* **46** 440–464.

CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. Academic, New York.

EHRHARD, A. (1983). Symétrisation dans l'espace de Gauss. *Math. Scand.* **53** 281–301.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

HELMERS, R. (1978). Edgeworth expansions for linear combinations of order statistics. Ph.D. dissertation, Math. Centrum, Amsterdam.

HELMERS, R. (1981). A Berry–Esseen theorem for linear combinations of order statistics. *Ann. Probab.* **9** 342–347.

INGLOT, T. and LEDWINA, T. (1989). Large and moderate deviations for some functionals of weighted empirical process. Technical Report, Technical Univ. Wrocław.

INGLOT, T. and LEDWINA, T. (1990). On probabilities of excessive deviations for Kolmogorov–Smirnov, Cramér–von Mises and chi-square statistics. *Ann. Statist.* **18** 1491–1495.

JOHNSON, B. R. and TRUAX, D. R. (1974). Asymptotic behavior of Bayes tests and Bayes risk. *Ann. Statist.* **2** 278–294.

JOHNSON, B. R. and TRUAX, D. R. (1978). Asymptotic behavior of Bayes procedures for testing simple hypotheses in multiparameter exponential families. *Ann. Statist.* **6** 346–361.

JUREČKOVÁ, J., KALLENBERG, W. C. M. and VERAVERBEKE, N. (1988). Moderate and Cramér-type large deviation theorems for M-estimators. *Statist. Probab. Lett.* **6** 191–199.

KALLENBERG, W. C. M. (1982). Cramér type large deviations for simple linear rank statistics. *Z. Wahrsch. Verw. Gebiete* **60** 403–409.

KALLENBERG, W. C. M. (1983a). Intermediate efficiency, theory and examples. *Ann. Statist.* **11** 170–182.

KALLENBERG, W. C. M. (1983b). On moderate deviation theory in estimation. *Ann. Statist.* **11** 498–504.

KASS, R. E. (1988). Comment on "Saddlepoint methods" by N. Reid. *Statist. Sci.* **3** 234–236.

MALEVICH, T. L., and ABDALIMOV, B. (1979). Large deviation probabilities for U-statistics. *Theory Probab. Appl.* **24** 215–219.

NEUHAUS, G. (1988). Addendum to: "Local asymptotics for linear rank statistics with estimated score functions." *Ann. Statist.* **16** 1342–1343.

PETROV, V. V. (1954). A generalization of Cramér's limit theorem. *Selected Translations Math. Statist. Probab.* **6** 1–8.

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York.

RUBIN, H. and SETHURAMAN, J. (1965a). Probabilities of moderate deviations. *Sankhyā Ser. A.* **27** 325–346.

RUBIN, H. and SETHURAMAN, J. (1965b). Bayes risk efficiency. *Sankhyā Ser. A.* **27** 347–356.

SEOH, M., RALESCU, S. S. and PURI, M. L. (1985). Cramér type large deviations for generalized rank statistics. *Ann. Probab.* **13** 115–125.

VANDEMAELE, M. (1982). On probabilities of large deviations for U-statistics. *Theory Probab. Appl.* **27** 614.

VANDEMAELE, M. and VERAVERBEKE, N. (1982). Cramér type large deviations and moderate deviations for linear combinations of order statistics. *Ann. Probab.* **10** 423–434.

WU, T. J. (1986). A large deviation result for signed linear rank statistics under the symmetry hypothesis. *Ann. Statist.* **14** 774–780.

T. INGLOT
T. LEDWINA
INSTITUTE OF MATHEMATICS
TECHNICAL UNIVERSITY OF WROCŁAW
50-370 WROCŁAW
WYBRZEŹE WYSPIAŃSKIEGO 27
POLAND

W. C. M. KALLENBERG
FACULTY OF APPLIED MATHEMATICS
UNIVERSITY OF TWENTE
P.O. BOX 217
THE NETHERLANDS