The Annals of Statistics
1988, Vol. 16, No. 1, 198–217

BAHADUR EFFICIENCY OF RANK TESTS FOR THE CHANGE-POINT PROBLEM

BY JAAP PRAAGMAN

Eindhoven University of Technology

A sequence of independent random variables $X_1, X_2, \dots, X_N$ is said to have a change point if $X_1, X_2, \dots, X_n$ have a common distribution F and $X_{n+1}, \dots, X_N$ have a common distribution G, $G \neq F$. Consider the problem of testing the null hypothesis of no change against the alternative of a change $G < F$ at an unknown change point n. Two classes of statistics based upon two-sample linear rank statistics (max- and sum-type) are compared in terms of their Bahadur efficiency. It is shown that for every sequence of sum-type statistics a sequence of max-type statistics can be constructed with at least the same Bahadur slope at all possible alternatives. Special attention is paid to alternatives close to the null hypothesis.

1. Introduction. Suppose we have a sequence of independent random variables $X_1, X_2, \dots, X_N$, then the sequence is said to have a change point at n, if $X_1, X_2, \dots, X_n$ have a common distribution $F(x)$ and $X_{n+1}, \dots, X_N$ have a common distribution $G(x)$, $G \neq F$. We consider the problem of testing the null hypothesis $\mathscr{H}_0$ of no change against the alternative $\mathscr{H}_a$ of a one-sided change at an unknown change point n.

This change-point problem (c.p.p.) only differs from the well-known two-sample problem (t.s.p.) in that n is unknown. Both problems have the same null hypothesis, whereas the alternative for the c.p.p. can be conceived as the union of the alternatives for the $N-1$ t.s.p.'s with $n = 1, 2, \dots, N-1$, respectively. Therefore, most of the statistics used for the c.p.p. are generalizations of two-sample statistics.

Let $T_{N,k}$ denote a t.s.p. statistic (samples $X_1, \dots, X_k$ and $X_{k+1}, \dots, X_N$). Then there are two obvious ways to define a c.p.p. statistic: as a weighted sum, $S_N = \sum_{k=1}^{N-1} c_{N,k} T_{N,k}$ or a weighted maximum, $M_N = \max_{1 \le k < N} c_{N,k} T_{N,k}$, where the $c_{N,k}$ are weights. For the M_N statistic to make sense, suppose that under the null hypothesis, $ET_{N,k} = 0$ for all k.

Consider the case when $T_{N,k}$ is a linear rank statistic. Thus, let R_i denote the rank of X_i, $i = 1, \dots, N$, then $T_{N,k} = \sum_{i=1}^{k} a_N(R_i)$, with $a_N(i)$ the so-called scores. If testing is against location shift, or more generally against $G < F$, well-known examples are the median test ($a_N(i) = \operatorname{sign}((N+1)/2 - i)$) and the Wilcoxon test ($a_N(i) = N^{-1}((N+1)/2 - i)$). Note that $ET_{N,k} = 0$ under $\mathscr{H}_0$, implies that $\sum_{i=1}^{N} a_N(i) = 0$. It is assumed that F and G are continuous, so that ties among the observations occur with probability 0.

Putting $d_N(i) = \sum_{k=i}^{N} c_{N,k}$, the sum-type statistic may be rewritten as

$$S_N = \sum_{i=1}^{N} d_N(i)\, a_N(R_i), \tag{1.1}$$

Received February 1984; revised March 1987.
AMS 1980 *subject classifications*. Primary 62G20, 62G10.
Key words and phrases. Bahadur efficiency, linear rank test, change-point test.

a so-called simple linear rank statistic. Generally, for the c.p.p., the weights $c_{N,k}$ are assumed to be nonnegative, so that the regression constants $d_N(i)$ satisfy $d_N(1) \geq \cdots \geq d_N(N) = 0$.

For the max-type statistic, we have

$$(1.2) \qquad M_N = \max_{1 \leq k \leq N} c_{N,k} \sum_{i=1}^{k} a_N(R_i).$$

Various authors have investigated—special cases of—(1.1) and (1.2). Bhattacharyya and Johnson (1968) considered (1.1) in its general form and proved local average power optimality against specific translation alternatives, if a score function is used that is appropriate for the underlying F.

Max-type statistics were proposed by McGilchrist and Woodyer (1975), Pettitt (1979) and Wolfe and Schechtman (1984). The first author took median scores and equal weights ($c_{N,k} = 1$), whereas the others used Wilcoxon scores: Pettitt with equal weights, Wolfe and Schechtman with weights inversely proportional to the standard deviations of the $T_{N,k}$.

Only Bhattacharyya and Johnson reported some results on the Pitman asymptotic relative efficiency. They proved that the Pitman efficiency of two sum-type tests with the same weights is independent of that particular weight function. Similarly, for two statistics with the same score function but involving different weights, the ARE is independent of the scores.

In this paper, we study the asymptotic efficiency using Bahadur's approach. The exact Bahadur slope of both S_N and M_N will be derived under appropriate conditions on the limit behavior of the scores and weights. Unlike Pitman efficiency, Bahadur efficiency of tests with the same weights (the same score function) depends on those weights (scores). Also it will be shown, as the main purpose of this paper, that for each sum-type statistic, there is a max-type statistic, which is at least as efficient in the sense of Bahadur uniformly over all the c.p.p. alternatives. Related work can be found in Deshayes and Picard (1982) and Haccou, Meelis and van de Geer (1985), where the Bahadur efficiency of parametric tests—such as the likelihood-ratio test—for the change point problem is investigated.

For a description of the concepts of Bahadur slope and efficiency, refer to Bahadur (1967, 1971), or Groeneboom and Oosterhoff (1977). Our method of evaluating Bahadur slopes relies on the following theorem by Bahadur (1967).

THEOREM 1.1. *Let $\{T_N\}$ be a sequence of test statistics for testing $\mathscr{H}_0$: $\theta \in \Theta_0$ against $\mathscr{H}_a$: $\theta \in \Theta - \Theta_0 = \Theta_1$.*

If a real number $b(\theta)$ exists and a nonnegative function $h(t)$ continuous at $t = b(\theta)$, such that

$$\lim_{N \to \infty} N^{-1} T_N = b(\theta) \quad a.s.\ P_\theta,\ \theta \in \Theta_1,$$

and

$$\lim_{N \to \infty} -N^{-1} \log[\sup\{P_\theta(T_N \geq Nt);\ \theta \in \Theta_0\}] = h(t), \quad \text{for all } t \in \mathbb{R},$$

then the Bahadur slope of $\{T_N\}$ at θ equals $2h(b(\theta))$.

Following the method suggested by this theorem, we will derive large deviation results under $\mathscr{H}_0$ in Section 3 and almost-sure limits under a fixed alternative in Section 4, which yields the Bahadur slopes of S_N and M_N and our main theorem in Section 5. In the next section, however, we first present a slightly different set-up for our testing problem, which will allow us to write both S_N and M_N as special functions of empirical distribution functions.

In Section 6, some numerical results are given for the special case of statistics with median scores and in Section 7, the efficiency at alternatives close to the null hypothesis is considered.

2. S_N and M_N as functions of e.d.f.'s. The key observation in this section is that our statistics do not depend on the ranks of the X_i only, but on their order in time too; therefore, random variables Y_i are introduced to deal with this ordering in time.

Ruymgaart and van Zuylen (1978) used a similar construction to deal with the regression constants in linear rank statistics. Let (X, Y) be a random vector defined on $\mathbb{R} \times [0,1]$ with distribution H and consider the testing of

$$\tilde{\mathscr{H}}_0\colon H(x, y) = yF(x),$$

against

$$\tilde{\mathscr{H}}_a\colon H(x, y) = \begin{cases} yF(x), & \text{if } y \in [0, \lambda], \\ \lambda F(x) + (y - \lambda)G(x), & \text{if } y \in (\lambda, 1], \end{cases} \tag{2.1}$$

with F, G and $\lambda \in (0,1)$ unknown. Then, with observations (X_i, Y_i), $i = 1, \dots, N$, we have $X_i \sim F$ for all i with $Y_i \le \lambda$ and $X_i \sim G$ for all i with $Y_i > \lambda$. Testing $\tilde{\mathscr{H}}_0$ against $\tilde{\mathscr{H}}_a$ means testing of $\lambda = 1$ (or 0) against $\lambda \in (0,1)$ or, with D_i the anti-rank of Y_i, $i = 1, \dots, N$, it is the testing of $\mathscr{H}_0$ against $\mathscr{H}_a$ for the sequence $X_{D_1}, \dots, X_{D_N}$.

Introduce the functions $\phi_N\colon (0,1] \to \mathbb{R}$, which are defined by

$$\phi_N(u) = a_N(i), \quad \text{for } u \in \left(\frac{i-1}{N}, \frac{i}{N}\right], \qquad i = 1, \dots, N.$$

In the same way, ψ_N and γ_N are defined based upon $d_{N,i}$ and $c_{N,i}$, respectively, and with $\gamma_N(0) := 0$. Define $H_N(x, y)$ as the empirical distribution function of (X_i, Y_i), $i = 1, \dots, N$, and let $H_{N,x}(x) = H_N(x, 1)$ and $H_{N,y}(y) = H_N(\infty, y)$ be the marginals. Then (unless otherwise stated, integration is over $[0,1]$)

$$S_N = N \int \int J_N(u, v)\, dH_N\big(H_{N,x}^{-1}(u), H_{N,y}^{-1}(v)\big), \tag{2.2}$$

with

$$J_N(u, v) = \phi_N(u)\psi_N(v), \tag{2.3}$$

where the equality in (2.2) should be understood as having the same distribution under $\mathscr{H}_0$ $(\tilde{\mathscr{H}}_0)$ and $\mathscr{H}_a$ $(\tilde{\mathscr{H}}_a)$. In the same way,

$$M_N = \sup_{0 \le \rho \le 1} N \int\int J_{N,\rho}(u, v)\, dH_N\big(H_{N,x}^{-1}(u), H_{N,y}^{-1}(v)\big), \tag{2.4}$$

with

$$J_{N,\rho}(u, v) = \gamma_N(\rho)\phi_N(u) 1_{[0,\rho]}(v). \tag{2.5}$$

3. Large deviations. Formulating our statistics as functions of the empirical distribution functions makes it possible to use a theorem on large deviations by Groeneboom, Oosterhoff and Ruymgaart (1979). After some preliminary definitions, we will state that theorem.

Let H_0 be the distribution function of (X_i, Y_i), $i = 1, 2, \dots,$ and D the space of two-dimensional distribution functions $H(x, y)$ endowed with the topology τ of convergence on all Borel sets. This is the smallest topology on D such that the functions T_f: $H \to \int f\, dH$, $H \in D$, are continuous for each bounded measurable function $f: \mathbb{R}^2 \to \mathbb{R}$. No metrization of this topology exists. For $H \in D$, $K(H, H_0)$ is the Kullback–Leibler information number of H with respect to H_0.

Take a function T: $D \to \overline{\mathbb{R}}$ and, for $t \in \mathbb{R}$, define

$$\Omega_{T,t} = \{H \in D | T(H) \ge t\}$$

and

$$K(\Omega_{T,t}, H_0) = \inf\{K(H, H_0) | H \in \Omega_{T,t}\}.$$

THEOREM 3.1 [Groeneboom, Oosterhoff and Ruymgaart (1979)]. *Let T: $D \to \overline{\mathbb{R}}$ be τ-continuous at every H with $K(H, H_0) < \infty$, and suppose that the function $t \to K(\Omega_{T,t}, H_0)$, $t \in \mathbb{R}$, is continuous from the right at $t = r$ and $\{u_N\}$ is a sequence of real numbers such that $\lim_{N\to\infty} u_N = 0$. Then*

$$\lim_{N\to\infty} -N^{-1}\log \Pr\{T(H_N) \ge r + u_N\} = K(\Omega_{T,r}, H_0).$$

3.1. *Sum-type statistics.* For $H(x, y) \in D$, with marginals $H_x(x)$ and $H_y(y)$, define $\overline{H}(u, v) = H(H_x^{-1}(u), H_y^{-1}(v))$ and consider, for some fixed value of M, S_{JM}: $D \to \overline{\mathbb{R}}$ defined by

$$S_{JM}(H) = \int\int J_M(u, v)\, d\overline{H}(u, v).$$

It will be shown that this S_{JM} meets the conditions of Theorem 3.1 (Lemmas 3.2 and 3.3) and then the desired large deviation result for S_N will follow, since, when M is large, $|N^{-1}S_N - S_{JM}(H_N)|$ will be arbitrarily small a.s. for N sufficiently large (Theorem 3.4).

LEMMA 3.2. *S_{JM} is τ-continuous at every continuous H.*

PROOF. Since the topology τ is finer than the weak topology, it suffices to prove that S_{JM} is continuous with respect to the latter topology. Furthermore, J_M is piecewise constant, hence, the weak continuity follows if $H_i \to H$ for any sequence $\{H_i\}$ in D implies that $\overline{H}_i \to \overline{H}$. Thus it is only necessary to show that $\overline{H}_i(u, v) \to \overline{H}(u, v)$ for each continuity point (u, v) of $\overline{H}$.

Let $\{H_i\}$ be a sequence in D with $H_i \to H$, then the marginals $H_{ix} \to H_x$ and $H_{iy} \to H_y$ also, with H_x and H_y continuous. Consequently, $H_{ix}^{-1}(u) \to H_x^{-1}(u)$ and $H_{iy}^{-1}(v) \to H_y^{-1}(v)$ except on at most countable sets D_x and D_y of discontinuity points of H_x^{-1} and H_y^{-1}, respectively. Consider (u, v) with $u \notin D_x$, $v \notin D_y$. Fix $\varepsilon > 0$, then, an $I \in \mathbb{N}$ exists such that for all $i \geq I$, $|H_x^{-1}(u) - H_{ix}^{-1}(u)| \leq \varepsilon$ and $|H_y^{-1}(v) - H_{iy}^{-1}(v)| \leq \varepsilon$. Thus, for all $i \geq I$, $H_i(H_x^{-1}(u) - \varepsilon, H_y^{-1}(v) - \varepsilon) \leq \overline{H}_i(u, v) \leq H_i(H_x^{-1}(u) + \varepsilon, H_y^{-1}(v) + \varepsilon)$ and using the pointwise convergence of $H_i \to H$ and the continuity of H, we get $\overline{H}_i(u, v) \to \overline{H}(u, v)$. Next consider arbitrary (u_0, v_0) and choose $u_1, u_2 \in (0,1)/D_x$ and $v_1, v_2 \in (0,1)/D_y$ such that $\overline{H}(u_0, v_0) - \varepsilon < \overline{H}(u_1, v_1) < \overline{H}(u_0, v_0) < \overline{H}(u_2, v_2) < \overline{H}(u_0, v_0) + \varepsilon$. Then, for all i, $\overline{H}_i(u_1, v_1) \leq \overline{H}_i(u_0, v_0) \leq \overline{H}_i(u_2, v_2)$. Thus

$$\overline{H}(u_1, v_1) \leq \liminf_{i \to \infty} \overline{H}_i(u_0, v_0) \leq \limsup_{i \to \infty} \overline{H}_i(u_0, v_0) \leq \overline{H}(u_2, v_2)$$

and it follows that $\overline{H}_i(u_0, v_0) \to \overline{H}(u_0, v_0)$. □

Introduce $\Omega = \{H \in D | H$ has a density h; $\int h(u, v)\, du = \int h(u, v)\, dv = 1\}$, and define for every J: $(0,1] \times (0,1] \to \mathbb{R}$ for which $|\int\int J(u, v)\, du\, dv| < \infty$,

$$\Omega(t, J) := \left\{H \in \Omega \Big| \int\int J(u, v)\, dH(u, v) \geq t\right\},$$

$$I(t; J) := \inf\left\{\int\int h \log h\, du\, dv \,|H \in \Omega(t, J)\right\},$$

$$t^-(J) = \int\int J(u, v)\, du\, dv \quad \text{and} \quad t^+(J) = \sup\left\{\int\int J(u, v)\, dH(u, v) | H \in \Omega\right\}.$$

LEMMA 3.3. *Let* J: $(0,1] \times (0,1] \to \mathbb{R}$, $|\int\int J(u, v)\, du\, dv| < \infty$. *Then*:

(i) $I(t; J)$ *is nonnegative, nondecreasing and convex (hence continuous) in* t, *for* $t^-(J) < t < t^+(J)$.

(ii) *If* J_ε: $(0,1] \times (0,1] \to \mathbb{R}$ *is such that*

$$\sup_{H \in \Omega} \left| \int\int (J_\varepsilon - J) h\, du\, dv \right| < \varepsilon$$

and

$$t^-(J) < t - \varepsilon < t + \varepsilon < t^+(J),$$

then

$$I(t - \varepsilon; J) \leq I(t; J_\varepsilon) \leq I(t + \varepsilon; J)$$

and

$$t^-(J_\varepsilon) \leq t^-(J) + \varepsilon < t < t^+(J) - \varepsilon \leq t^+(J_\varepsilon).$$

(iii) *Suppose* H_0 *is the uniform distribution on the unit square. Then for* $t^-(J_M) < t < t^+(J_M)$,

$$K\left(\Omega_{S_{JM}, t}, H_0\right) = I(t; J_M).$$

PROOF. For (i) and (ii) see Woodworth (1970). For the proof of (iii) we use the same line of argument as in the proof of Theorem 2.2.1 in Groeneboom (1979). Take $H \in D$, such that $K(H, H_0) < \infty$ (thus $H \ll H_0$), suppose that h and $\bar{h}$ are the densities of H and $\bar{H}$, and that $h_x(x)$ and $h_y(y)$ are the densities of the marginals $H_x(x)$ and $H_y(y)$. Then

$$\begin{aligned} K(H, H_0) &= \int\int h(x, y) \log h(x, y)\, dx\, dy \\ &= \int\int \frac{h\big(H_x^{-1}(u), H_y^{-1}(v)\big)}{h_x\big(H_x^{-1}(u)\big) h_y\big(H_y^{-1}(v)\big)} \log h\big(H_x^{-1}(u), H_y^{-1}(v)\big)\, du\, dv \\ &= \int\int \bar{h}(u, v) \log \bar{h}(u, v)\, du\, dv \\ &\quad + \int\int \bar{h}(u, v) \log\big\{h_x\big(H_x^{-1}(u)\big) h_y\big(H_y^{-1}(v)\big)\big\}\, du\, dv \\ &\geq K(\bar{H}, H_0). \end{aligned}$$

Also, since $\bar{H} \in \Omega$ for every $H \in D$ with $K(H, H_0) < \infty$ and $\bar{\bar{H}} = \bar{H}$, it follows that

$$\begin{aligned} K\big(\Omega_{S_{JM}, t}, H_0\big) &= \inf\Big\{K(\bar{H}, H_0) | H \in D, \int\int J_M(u, v)\, d\bar{H}(u, v) \geq t\Big\} \\ &= \inf\{K(H, H_0) | H \in \Omega(t, J_M)\} = I(t; J_M) \qquad \square \end{aligned}$$

THEOREM 3.4. *Let (X_i, Y_i), $i = 1, \ldots, N$, be i.i.d. with continuous distribution $H(x, y) = yF(x)$, $x \in \mathbb{R}$, $y \in [0,1]$. S_N is defined by* (2.2) *and suppose a function J: $(0,1] \times (0,1] \to \mathbb{R}$ exists such that*

$$\sup_{H \in \Omega} \left| \int\int \big(J_N(u, v) - J(u, v)\big) h(u, v)\, du\, dv \right| \to 0. \tag{3.1}$$

If $\{u_N\}$ is a sequence of real numbers such that $u_N \to 0$, then for all t, with $t^-(J) < t < t^+(J)$,

$$\lim_{N \to \infty} -N^{-1} \log \Pr\big(N^{-1} S_N \geq t + u_N\big) = I(t; J).$$

PROOF. For every continuous distribution F, the distribution of S_N is the same, so without loss of generality we can assume that $H = H_0$, the uniform distribution on the unit square.

Fix $\varepsilon > 0$, such that $t^-(J) < t - 2\varepsilon < t + 2\varepsilon < t^+(J)$, then there is an M, such that for all $N \geq M$,

$$\sup_{H \in \Omega} \left| \int\int \big(J(u, v) - J_N(u, v)\big) h(u, v)\, du\, dv \right| < \varepsilon/4 \tag{3.2}$$

and

$$\sup_{H \in \Omega} \left| \int\int \big(J_N(u, v) - J_M(u, v)\big) h(u, v)\, du\, dv \right| < \varepsilon/2. \tag{3.3}$$

First, it is shown that for a sufficiently large N,

(3.4) $$|N^{-1}S_N - S_{JM}(H_N)| < \varepsilon, \quad \text{with probability 1.}$$

Introduce $\bar{h}_N(u, v) = N$ for $(u, v) \in ((R_i - 1)/N, R_i/N] \times ((Q_i - 1)/N, Q_i/N]$ and 0 elsewhere, with $Q_1, \dots, Q_N$ the ranks of $Y_1, \dots, Y_N$. Then

$$\begin{aligned}|N^{-1}S_N - S_{JM}(H_N)| &= \left|\int\int (J_N(u, v) - J_M(u, v))\, d\overline{H}_N(u, v)\right| \\ &= \sum_{i=1}^{N} \left|J_N\left(\frac{R_i}{N}, \frac{Q_i}{N}\right) - J_M\left(\frac{R_i}{N}, \frac{Q_i}{N}\right)\right| N^{-1} \\ &\leq \int\int |J_N(u, v) - J_M(u, v)| \bar{h}_N(u, v)\, du\, dv \\ &\quad + \frac{2M}{N} \max\left|J_M\left(\frac{k}{M}, \frac{l}{M}\right) - J_M\left(\frac{k'}{M}, \frac{l'}{M}\right)\right|,\end{aligned}$$

where the max is taken over all k, l, k' and l' with $1 \leq k, l, k', l' \leq M$, $|k - k'| \leq 1$ and $|l - l'| \leq 1$.

For N large, the second term will be smaller than $\varepsilon/2$, hence with (3.3), we get (3.4). But then

(3.5) $$\begin{aligned}&\lim_{N\to\infty} - N^{-1}\log \Pr(S_{JM}(H_N) \geq t + u_N - \varepsilon) \\ &\leq \liminf_{N\to\infty} - N^{-1}\log \Pr(S_N \geq N(t + u_N)) \\ &\leq \limsup_{N\to\infty} - N^{-1}\log \Pr(S_N \geq N(t + u_N)) \\ &\leq \lim_{N\to\infty} - N^{-1}\log \Pr(S_{JM}(H_N) \geq t + u_N + \varepsilon),\end{aligned}$$

and using Theorem 3.1 and Lemmas 3.2 and 3.3, the first and last limits are $I(t - \varepsilon; J_M)$ and $I(t + \varepsilon; J_M)$.

Furthermore, from (3.2) and Lemma 3.3(ii), we know that $I(t - 2\varepsilon; J) \leq I(t - \varepsilon; J_M)$ and $I(t + \varepsilon; J_M) \leq I(t + 2\varepsilon; J)$. Thus, using the continuity of $I(r, J)$ at $r = t - 2\varepsilon$ and $t + 2\varepsilon$ [Lemma 3.3(i)] along with (3.5), the proof is completed. □

This result has already been proved by Woodworth (1970). Yet we gave a proof, because Woodworth used a different approach, and the techniques used in our proof will be needed again further on in this paper. For our special case of S_N, we have $J_N(u, v) = \phi_N(u)\psi_N(v)$. Woodworth also pointed out that (3.1) holds with $J(u, v) = \phi(u)\psi(v)$, if the functions ϕ_N and ψ_N converge in L_2 to the functions ϕ and ψ. The weight functions for the c.p.p. often will be bounded and then, for the score functions, convergence in L_1 only is sufficient.

3.2. *Max-type statistics.* For the max-type statistics defined by (2.3) the same approach can be used; therefore define $M_{JM}\colon D \to \overline{\mathbb{R}}$, by

$$M_{JM}(H) := \sup_{0 \leq \rho \leq 1} \int\int J_{M,\rho}(u, v)\, d\overline{H}(u, v),$$

with

$$J_{M,\rho}(u,v) = \gamma_M(\rho)\phi_M(u)1_{[0,\rho]}(v).$$

LEMMA 3.5. *M_{JM}: $D \to \overline{\mathbb{R}}$ is τ-continuous at every continuous H.*

PROOF. Take $\rho \in (0,1]$ and fix k such that $(k-1)/M < \rho \le k/M$ and consider $\tilde{\rho}$, $(k-1)/M < \tilde{\rho} \le \rho$. Then $\gamma_M(\tilde{\rho}) = \gamma_M(\rho)$. Thus

$$\left|\int\int J_{M,\rho}(u,v)\,d\overline{H} - \int\int J_{M,\tilde{\rho}}(u,v)\,d\overline{H}\right| = \left|\gamma_M(\rho)\int_0^1\int_{\tilde{\rho}}^{\rho}\phi_M(u)\,d\overline{H}(u,v)\right|$$
$$\le |\gamma_M(\rho)|\max_u|\phi_M(u)|(\rho-\tilde{\rho}) < \varepsilon/3,$$

for $\rho - \tilde{\rho}$ small. Consequently, there is a neighborhood B_ρ of ρ, such that for all $\tilde{\rho} \in B_\rho$,

$$\left|\int\int J_{M,\tilde{\rho}}\,d\overline{H} - \int\int J_{M,\rho}\,d\overline{H}\right| < \varepsilon/3, \quad \text{for all } H \in D.$$

In a similar way a neighborhood B_0 for $\rho = 0$ can be found too. Since $[0,1]$ is compact, the cover $\bigcup_\rho B_\rho$ has a finite subcover $\{B_{\rho_i}\}$, $i = 1,\dots,q$, such that for every $\rho \in B_{\rho_i}$, $|\int\int J_{M,\rho}\,d\overline{H} - \int\int J_{M,\rho_i}\,d\overline{H}| < \varepsilon/3$ for all H, $i = 1,\dots,q$. Furthermore, Lemma 3.2 insures that for every ρ, $H \to \int\int J_{M,\rho}\,d\overline{H}$ is τ-continuous for every continuous H. So, for every ρ_i, $i = 1,\dots,q$, there is a τ-neighborhood $B_i(H)$ of H, such that $|\int\int J_{M,\rho_i}\,d\overline{H} - \int\int J_{M,\rho_i}\,d\overline{G}| \le \varepsilon/3$ for all $G \in B_i(H)$.

Now take $B(H) = \bigcap_{i=1,\dots,q} B_i(H)$; then, for each ρ, a ρ_i exists such that for all $G \in B(H)$,

$$\left|\int\int J_{M,\rho}\,d\overline{G} - \int\int J_{M,\rho}\,d\overline{H}\right| \le \left|\int\int J_{M,\rho}\,d\overline{G} - \int\int J_{M,\rho_i}\,d\overline{G}\right|$$
$$+\left|\int\int J_{M,\rho_i}\,d\overline{G} - \int\int J_{M,\rho_i}\,d\overline{H}\right|$$
$$+\left|\int\int J_{M,\rho_i}\,d\overline{H} - \int\int J_{M,\rho}\,d\overline{H}\right|$$
$$\le \varepsilon.$$

Since this holds for each ρ, $\sup_\rho \int\int J_{M,\rho}\,d\overline{H}$ is continuous at H. □

LEMMA 3.6. *Let J_ρ: $[0,1]\times[0,1] \to \mathbb{R}$, $|\int\int J_\rho(u,v)\,du\,dv| < \infty$, for every $\rho \in [0,1]$. Then*

$\inf_\rho I(t;\, J_\rho)$ is nonnegative, nondecreasing and continuous in t,
for $0 < t < \sup t^+(J_\rho)$.

Furthermore, let $\tilde{J}_\rho$: $[0,1]\times[0,1] \to \mathbb{R}$, $|\int\int \tilde{J}_\rho(u,v)\,du\,dv| < \infty$ for every $\rho \in [0,1]$; if a strictly increasing, continuous mapping q of $[0,1]$ onto itself exists such that

$$\sup_{H\in\Omega}\ \sup_{\rho\in[0,1]}\left|\int\int\left(\tilde{J}_{q(\rho)}(u,v) - J_\rho(u,v)\right)h(u,v)\,du\,dv\right| < \varepsilon,$$

then, for all t with $0 < t - \varepsilon < t + \varepsilon < \sup t^+(J_\rho)$,

$$\inf_\rho I(t-\varepsilon; J_\rho) \le \inf_\rho I\big(t; \tilde{J}_\rho\big) \le \inf_\rho I(t+\varepsilon; J_\rho)$$

and

$$\sup_\rho t^+(J_\rho) < \sup_\rho t^+\big(\tilde{J}_\rho\big) + \varepsilon.$$

PROOF. Directly from Lemma 3.3(i) and (ii). □

THEOREM 3.7. *Let* (X_i, Y_i), $i = 1, \dots, N$, *be i.i.d. with a continuous distribution* $H(x, y) = yF(x)$, $x \in \mathbb{R}$, $y \in [0,1]$. M_N *is defined by* (2.4) *and* (2.5), *and suppose there is a function* ϕ: $[0,1] \to \mathbb{R}$, *with* $\int \phi(u)\,du = 0$, $\int \phi^2(u)\,du < \infty$ *and* $\int(\phi_N(u) - \phi(u))^2\,du \to 0$, *and a function* γ: $[0,1] \to \mathbb{R}^+$ *such that the* γ_N *converge to* γ *in the Skorohod topology on every closed interval* $[\tau, 1-\tau]$; $0 < \tau < \frac{1}{2}$, $\gamma(0) = 0$ *and* $\gamma_N(v)(v(1-v))^{1/2}$ *is bounded on* $(0,1)$, *uniformly in N.*

If $\{u_N\}$ *is a sequence of real numbers such that* $u_N \to 0$, *then for all t*, $0 < t < \sup_\rho t^+(J_\rho)$,

$$\lim_{N\to\infty} -N^{-1}\log \Pr\big(N^{-1}M_N \ge t + u_N\big) = \inf_{0\le\rho\le 1} I(t; J_\rho),$$

with

$$J_\rho(u, v) = \phi(u)\gamma(\rho)1_{[0,\rho]}(v).$$

PROOF. Again, suppose without loss of generality, that $H = H_0$, the uniform distribution on the unit square. Fix $\varepsilon > 0$, such that $0 < t - 2\varepsilon < t + 2\varepsilon < \sup_\rho t^+(J_\rho)$. We first will show that strictly increasing continuous mappings q_N of $[0,1]$ onto itself exist, such that $q_N(\rho) \to \rho$ uniformly in ρ and

$$\sup_{H\in\Omega}\sup_{0\le\rho\le 1}\left|\int\!\!\int\big(J_{N,q_N(\rho)}(u,v) - J_\rho(u,v)\big)h(u,v)\,du\,dv\right| \to 0. \tag{3.6}$$

Fix τ, $0 < \tau < \frac{1}{2}$, and consider $\rho \in [\tau, 1-\tau]$. For these ρ, the left-hand side of (3.6) is smaller than

$$\max_{\tau\le\rho\le 1-\tau} \gamma_N(q_N(\rho))\int|\phi_N(u) - \phi(u)|\,du$$
$$+ \max_{\tau\le\rho\le 1-\tau}\left\{\int\Big[\gamma_N(q_N(\rho))1_{(0,q_N(\rho)]}(v) - \gamma(\rho)1_{(0,\rho]}(v)\Big]^2 dv\right\}^{1/2}$$
$$\times\left\{\int\phi^2(u)\,du\right\}^{1/2},$$

and the limit behavior of the ϕ_N together with the Skorohod convergence of $\gamma_N \to \gamma$ on $[\tau, 1-\tau]$ guarantees that both terms are smaller than $\varepsilon/2$ for large N.

For $\rho \in (0, \tau)$, take $q_N(\rho) = \rho$ and using the Cauchy–Schwarz inequality it follows that for these ρ (3.6) is smaller than

$$\sup_{H\in\Omega}\sup_{0<\rho<\tau}\gamma_N(\rho)\rho^{1/2}\left(\int(\phi_N(u)-\phi(u))^2\,du\right)^{1/2}$$

$$+\sup_{H\in\Omega}\sup_{0<\rho<\tau}|\gamma_N(\rho)-\gamma(\rho)|\rho^{1/2}\left\{\int\phi^2(u)\left(\int_0^\rho h(u,v)\,dv\right)du\right\}^{1/2}.$$

And since $\gamma_N(\rho)\rho^{1/2}$ is bounded for $\rho < \tau$, $\int(\phi_N(u) - \phi(u))^2\,du \to 0$ and $\{\int\phi^2(u)\int_0^\rho h(u, v)\,dv\,du\} \to 0$ uniformly in h if $\rho \downarrow 0$, both terms can be made smaller than $\varepsilon/2$ by first taking τ small enough and then N large enough. Finally, for $\rho \in (1 - \tau, 1)$ the result follows similarly using

$$\int_0^1\int_0^\rho \phi_N(u)h(u,v)\,du\,dv = -\int_0^1\int_\rho^1 \phi_N(u)h(u,v)\,du\,dv$$

[since $\int\phi_N(u)\,du = 0$] and thus the proof of (3.6) is complete.

Hence there is an $M \in \mathbb{N}$, such that for all $N \ge M$,

$$\sup_{H\in\Omega}\sup_{0\le\rho\le1}\left|\int\int\left(J_{N,q_N(\rho)}(u,v)-J_\rho(u,v)\right)h(u,v)\,du\,dv\right|<\varepsilon/2 \tag{3.7}$$

and

$$\sup_{H\in\Omega}\sup_{0\le\rho\le1}\left|\int\int\left(J_{M,q_M(\rho)}(u,v)-J_{N,q_N(\rho)}(u,v)\right)h(u,v)\,du\,dv\right|<\varepsilon/2. \tag{3.8}$$

Then it can be shown in the same way as in the proof of Theorem 3.4 that for a sufficiently large N, with probability 1,

$$M_{JM}(H_N)-\varepsilon<N^{-1}M_N<M_{JM}(H_N)+\varepsilon. \tag{3.9}$$

Due to its continuity (Lemma 3.5), a large deviation result for M_{JM} can be derived from Theorem 3.1, provided that the mapping $t \to \inf\{\int\int h\log h|H \in D, M_{JM} \ge t\}$ is continuous from the right. Now $\{H \in D|M_J(H) \ge t\} = \bigcup_\rho\{H \in D|\int\int J_{M,\rho}\,d\bar{H} \ge t\}$ and thus, with Lemma 3.3(iii), we get

$$\inf\left\{\int\int h\log h|H\in D,\ M_{JM}(H)\ge t\right\}=\inf_\rho I(t;J_{M,\rho}),$$

and the continuity required is given by the first part of Lemma 3.6. Thus, Theorem 3.1 may be applied to (3.9) in order to give

$$\begin{aligned}\inf_\rho I(t-\varepsilon;J_{M,\rho})&\le\liminf_{N\to\infty}-N^{-1}\log\Pr(M_N\ge N(t+u_N))\\&\le\limsup_{N\to\infty}-N^{-1}\log\Pr(M_N\ge N(t+u_N))\\&\le\inf_\rho I(t+\varepsilon;J_{M,\rho}).\end{aligned}\tag{3.10}$$

However $\inf_\rho I(t - 2\varepsilon;\ J_\rho) \le \inf_\rho I(t - \varepsilon;\ J_{M,\rho})$ and $\inf_\rho I(t + \varepsilon;\ J_{M,\rho}) \le$

$\inf_\rho I(t + 2\varepsilon; J_\rho)$ [(3.7) and the second part of Lemma 3.6], hence, with the continuity of $\inf_\rho I(t; J_\rho)$ (Lemma 3.6), the proof can be completed. □

COROLLARY 3.8.

$$\inf_\rho I(t;\, J_\rho) = \inf\left\{\int\!\!\int h \log h | H \in \Omega^*(t)\right\},$$

with

$$\Omega^*(t) := \Big\{ H \in \Omega | h(u,v) = f(u)1_{[0,\rho]}(v) + g(u)1_{(\rho,1]}(v),$$

$$\textit{and} \int\!\!\int J_\rho h\, du\, dv \geq t, \textit{ for some } \rho \in [0,1]\Big\}.$$

PROOF. Consider an arbitrary $H \in \Omega$, such that $\int\!\int J_\rho(u,v)\, dH(u,v) \geq t$ for some ρ, and define

$$f(u) := \rho^{-1}\int_0^\rho h(u,v)\, dv, \qquad g(u) := (1-\rho)^{-1}\int_\rho^1 h(u,v)\, dv$$

and

$$h_\rho(u,v) = \begin{cases} f(u), & 0 \leq v \leq \rho, \\ g(u), & \rho < v \leq 1. \end{cases}$$

Then by the information inequality [Kullback (1959), Theorem 2.4.2]

$$\begin{aligned}\int\!\!\int & h(u,v)\log h(u,v)\, du\, dv \\ &\geq \int_0^1 \rho f(u)\log f(u)\, du + \int_0^1 (1-\rho)g(u)\log g(u)\, du \\ &= \int\!\!\int h_\rho(u,v)\log h_\rho(u,v)\, du\, dv\end{aligned}$$

and

$$\begin{aligned}\int\!\!\int J_\rho(u,v)h_\rho(u,v)\, du\, dv &= \int_0^1 \gamma(\rho)\left(\int_0^\rho h_\rho(u,v)\, dv\right)\phi(u)\, du \\ &= \int_0^1 \gamma(\rho)\left(\int_0^\rho h(u,v)\, dv\right)\phi(u)\, du \\ &= \int\!\!\int J_\rho(u,v)h(u,v)\, du\, dv \geq t. \qquad \square\end{aligned}$$

4. Almost-sure limits. Our next concern is the derivation of the almost-sure limits under a fixed alternative for statistics of both types. Every alternative may be characterized by the triple (F, G, λ): the distributions (continuous) before and after the change point and $\lambda \in (0,1)$ for the change point itself. Again, the statistics (2.2) and (2.4) will be used, but now we will find the almost-sure limits under a general two-dimensional distribution H first. Afterward, the results will be specialized to H as defined in (2.1).

4.1. *Sum-type statistics.*

THEOREM 4.1. *Let* $(X, Y) \sim H$ *and let* S_N *be defined by* (2.2). *If the functions* J_N *satisfy* (3.1) *for some* J: $[0,1] \times [0,1] \to \mathbb{R}$, *then*

$$\lim_{N\to\infty} N^{-1}S_N = \int\int J(u,v)\, d\overline{H}(u,v) \quad a.s.\ P_H.$$

PROOF. Take $\varepsilon > 0$ and choose $\tilde{N}$ such that

$$\sup_{H\in\Omega}\int\int |J_{\tilde{N}}(u,v) - J(u,v)| h(u,v)\, du\, dv \le \varepsilon/3 \tag{4.1}$$

and, for all $N > \tilde{N}$,

$$\sup_{H\in\Omega}\int\int |J_{\tilde{N}}(u,v) - J_N(u,v)| h(u,v)\, du\, dv \le \varepsilon/3. \tag{4.2}$$

Then

$$\begin{aligned}\left|N^{-1}S_N - \int\int J(u,v)\, d\overline{H}(u,v)\right| &\le \int\int |J_{\tilde{N}}(u,v) - J_N(u,v)|\, d\overline{H}_N \\ &\quad + \left|\int\int J_{\tilde{N}}(u,v)\, d(\overline{H}_N - \overline{H})\right| \\ &\quad + \int\int |J_{\tilde{N}}(u,v) - J(u,v)|\, d\overline{H}.\end{aligned}$$

According to the Glivenko–Cantelli theorem the empirical distribution function H_N converges to H in the supremum metric and, just as in the proof of Lemma 3.2, we get $\overline{H}_N(u,v) \to \overline{H}(u,v)$ for every $(u,v) \in [0,1] \times [0,1]$. Since $J_{\tilde{N}}$ is piecewise constant, the second term will be smaller than $\varepsilon/3$ for a sufficiently large N. Furthermore, the last term is smaller than $\varepsilon/3$ according to (4.1) and it has already been shown in the proof of Theorem 3.4 that $\int\int |J_{\tilde{N}} - J_N|\, d\overline{H}_N < \varepsilon/3$ for large N, as a consequence of (4.2); this completes the proof. □

The next corollary states the result for the special H, defined by (2.1), starting with an F, G and λ. Introduce $\overline{F}_\lambda(u) = F(H_x^{-1}(u))$ and $\overline{G}_\lambda(u) = G(H_x^{-1}(u))$.

COROLLARY 4.2. *If* $H(x, y)$ *is defined by* (2.1) *and* $J(u,v) = \phi(u)\psi(v)$, *then*

$$\begin{aligned}\lim_{N\to\infty} N^{-1}S_N &= \int_0^\lambda \psi(v)\, dv \int_0^1 \phi(u)\, d\overline{F}_\lambda(u) \\ &\quad + \int_\lambda^1 \psi(v)\, dv \int_0^1 \phi(u)\, d\overline{G}_\lambda(u) \quad a.s.\ P_H.\end{aligned}$$

4.2. *Max-type statistics.*

THEOREM 4.3. *Let* $(X, Y) \sim H$ *and let* M_N *be defined by* (2.4) *and* (2.5). *If the functions* $J_{N,\rho}$ *satisfy* (3.6), *then*

$$\lim_{N\to\infty} N^{-1}M_N = \sup_\rho \int\int J_\rho(u,v)\, d\overline{H}(u,v) \quad a.s. P_H.$$

Proof. From Theorem 4.1, we know that for every $\rho \in [0,1)$,

$$\lim_{N\to\infty} \iint J_{N,q_N(\rho)}(u,v)\,d\overline{H}_N = \iint J_\rho(u,v)\,d\overline{H} \quad \text{a.s. } P_H.$$

Therefore, we only have to prove that this convergence is uniform in ρ. Fix $\varepsilon > 0$ and take $\tilde{N}$ such that

$$\sup_{H\in\Omega}\sup_{\rho} \iint \left| J_{\tilde{N},q_{\tilde{N}}(\rho)}(u,v) - J_\rho(u,v)\right| h(u,v)\,du\,dv \le \varepsilon/3, \tag{4.3}$$

and, for all $N > \tilde{N}$,

$$\sup_{H\in\Omega}\sup_{\rho} \iint \left| J_{\tilde{N},q_{\tilde{N}}(\rho)}(u,v) - J_{N,q_N(\rho)}(u,v)\right| h(u,v)\,du\,dv \le \varepsilon/3. \tag{4.4}$$

Consider an arbitrary $\rho \in [0,1]$ and write

$$\begin{aligned}
&\left|\iint J_{N,q_N(\rho)}\,d\overline{H}_N - \iint J_\rho\,d\overline{H}\right| \\
&\quad\le \iint \left|J_{N,q_N(\rho)} - J_{\tilde{N},q_{\tilde{N}}(\rho)}\right| d\overline{H}_N + \left|\iint J_{\tilde{N},q_{\tilde{N}}(\rho)}\,d(\overline{H}_N - \overline{H})\right| \\
&\quad\quad + \iint \left|J_{\tilde{N},q_{\tilde{N}}(\rho)} - J_\rho\right| d\overline{H}.
\end{aligned}$$

Just as in the proof of Theorem 3.7, it follows from (4.4) that the first term is smaller than $\varepsilon/3$, uniformly in ρ. And because of (4.3), the same holds true for the last term. Hence it only remains to show that $|\iint J_{\tilde{N},q_{\tilde{N}}(\rho)}\,d(\overline{H}_N - \overline{H})| \le \varepsilon/3$, uniformly in ρ, for large N.

Take $N_2 = N_2(\rho)$, such that $|\iint J_{\tilde{N},q_{\tilde{N}}(\rho)}\,d(\overline{H}_N - \overline{H})| \le \varepsilon/9$ for all $N > N_2$. As in the proof of Lemma 3.5, and using $q_N(\rho) \to \rho$ uniformly in ρ, we can find a neighborhood B_ρ, such that for all $\tilde{\rho} \in B_\rho$,

$$\left|\iint \left(J_{\tilde{N},q_{\tilde{N}}(\rho)} - J_{\tilde{N},q_{\tilde{N}}(\tilde{\rho})}\right) d\overline{H}\right| \le \varepsilon/9, \quad \text{for all } H \in D.$$

Hence we get for all $\tilde{\rho} \in B_\rho$ and $N > N_2(\rho)$,

$$\left|\iint J_{\tilde{N},q_{\tilde{N}}(\tilde{\rho})}\,d\overline{H}_N - \iint J_{\tilde{N},q_{\tilde{N}}(\tilde{\rho})}\,d\overline{H}\right| \le \varepsilon/3.$$

Using the compactness of the ρ-interval, the completion of the proof is straightforward. □

Again, this theorem holds true for all continuous $H \in D$. The next corollary is suitable for the special distributions that satisfy (2.1).

Introduce the functions b_λ: $[0,1] \to \mathbb{R}$,

$$b_\lambda(\rho) = \begin{cases} \rho \int \phi(u)\, d\bar{F}_\lambda(u), & \text{if } \rho \in [0, \lambda], \\ \lambda \int \phi(u)\, d\bar{F}_\lambda(u) + (\rho - \lambda) \int \phi(u)\, d\bar{G}_\lambda(u), & \text{if } \rho \in (\lambda, 1]. \end{cases}$$

COROLLARY 4.4. *If $H \in D$ satisfies* (2.1), *then* $\lim_{N \to \infty} N^{-1} M_N = \sup_\rho \gamma(\rho) b_\lambda(\rho)$ *a.s.* P_H.

5. The efficiency of S_N with respect to M_N. Using Theorem 1.1, the results of the foregoing sections enable us to find the exact Bahadur slope of S_N and M_N, and to compare the two types of statistic, leading to the main result of this paper (Theorem 5.1).

Define S_N as before, $\phi_N \to \phi$ and $\psi_N \to \psi$ in L_2 again.

Unlike in Sections 2 and 3, here, the monotonicity of the ψ_N will be used. Furthermore, the ϕ-function is assumed to be nonconstant and nonincreasing. For the alternatives under consideration, this assumption is not very restrictive.

Introduce $\Gamma(v) = \int_0^v \psi(s)\, ds$ and define γ: $[0,1] \to \mathbb{R}$,

$$\gamma(v) = \frac{\Gamma(v) - v\Gamma(1)}{v(1-v)}, \quad \text{for } v \in (0,1),\ \gamma(0) = 0. \tag{5.1}$$

Similarly, Γ_N and γ_N are defined based upon ψ_N. Note that the monotonicity of ψ_N (ψ) implies that $\gamma_N(v) \ge 0$, $[\gamma(v) \ge 0]$.

THEOREM 5.1. *Let $\{S_N\}$ be a sequence of statistics defined by* (2.2) *with $\int \phi_N(u)\, du = 0$ and nonincreasing ψ_N. Suppose there is a nonconstant and nonincreasing function ϕ: $[0,1] \to \mathbb{R}$ such that $\int (\phi_N(u) - \phi(u))^2\, du \to 0$ and a function ψ: $[0,1] \to \mathbb{R}^+$ such that $\int (\psi_N(v) - \psi(v))^2\, dv \to 0$. Then, for each alternative (F, G, λ), the Bahadur slope of $\{S_N\}$ is less than or equal to the Bahadur slope of the sequence $\{M_N\}$, defined by* (2.4) *and with the same score functions ϕ_N, and weight functions γ_N defined according to* (5.1).

Using the Cauchy–Schwarz inequality, it is easily demonstrated that the L_2 convergence of the ψ_N guarantees both the uniform boundedness of $\gamma_N(v) v(1-v))^{1/2}$ and the Skorohod convergence of $\gamma_N \to \gamma$. Therefore, Theorem 3.7 and Lemma 4.4 are in force.

For the proof of Theorem 5.1 the following lemma will be useful.

LEMMA 5.2. *Suppose an alternative (F, G, λ) holds. Then, for the $\{S_N\}$ and the corresponding $\{M_N\}$ as given in Theorem* 5.1,

$$\lim_{N \to \infty} N^{-1} S_N = \lim_{N \to \infty} N^{-1} M_N = \gamma(\lambda) b_\lambda(\lambda) \quad a.s.\ P_{F,G,\lambda}.$$

Proof. From Corollary 4.4 we see that $\lim_{N\to\infty} N^{-1}M_N = \max_\rho \gamma(\rho)b_\lambda(\rho)$. For $\rho \in [0, \lambda]$,

$$\frac{d}{d\rho}\gamma(\rho)b_\lambda(\rho) = \frac{(1-\rho)\psi(\rho) + \Gamma(\rho) - \Gamma(1)}{(1-\rho)^2}\int \phi(u)\, d\bar{F}_\lambda(u).$$

But ψ is nonincreasing thus $\Gamma(1) - \Gamma(\rho) = \int_\rho^1 \psi(v)\, dv \le \psi(\rho)(1-\rho)$. Furthermore, $F > G$ and, consequently, for all x, $F(x) \ge \lambda F(x) + (1-\lambda)G(x) = H_x(x)$. And since ϕ is nonincreasing,

$$\int \phi(u)\, d\bar{F}_\lambda(u) = \int_{\mathbf{R}} \phi(H_x(x))\, dF(x) \ge \int_{\mathbf{R}} \phi(F(x))\, dF(x) = \int \phi(u)\, du = 0.$$

Thus the derivative $d/d\rho\, (\gamma(\rho)b_\lambda(\rho))$ is nonnegative for $\rho \in [0, \lambda)$. In a similar way it can be seen that this derivative is nonpositive for $\rho \in [\lambda, 1]$, and, since $\gamma(\rho)b_\lambda(\rho)$ is continuous, it follows that $\max_\rho \gamma(\rho)b_\lambda(\rho) = \gamma(\lambda)b_\lambda(\lambda)$. On the other hand, according to Corollary 4.2, we get

$$\lim_{N\to\infty} N^{-1}S_N = \Gamma(\lambda)\int \phi(u)\, d\bar{F}_\lambda(u) + (\Gamma(1) - \Gamma(\lambda))\int \phi(u)\, d\bar{G}_\lambda(u),$$

and, since

$$\int \phi(u)\, du = 0 \quad \text{and} \quad \lambda \bar{F}_\lambda(u) + (1-\lambda)\bar{G}_\lambda(u) = u,$$

this equals

$$\frac{\Gamma(\lambda) - \lambda\Gamma(1)}{1-\lambda}\int \phi(u)\, d\bar{F}_\lambda(u),$$

which is just $\gamma(\lambda)b_\lambda(\lambda)$. □

Proof of Theorem 5.1. According to Lemma 5.2, the almost-sure limit of the corresponding max- and sum-type statistic is the same, so that a difference in slope will occur only when there is a difference in the exponential error.

Refer to Theorem 3.7 and Corollary 3.8 to write

$$\lim_{N\to\infty} -N^{-1}\log \Pr\left(N^{-1}M_N \ge t + u_N\right)$$

$$= \inf\Big\{ \int\int h \log h \,\Big|\, \int h\, du = \int h\, dv = 1,$$

$$h(u, v) = f(u)1_{[0,\rho]}(v) + g(u)1_{(\rho,1]}(v),$$

$$\int\int J_\rho(u, v)h(u, v)\, du\, dv \ge t,\ \rho \in [0, 1]\Big\}.$$

Here

$$\begin{aligned}
&\int\int J_\rho(u,v)h(u,v)\,du\,dv \\
&= \frac{\Gamma(\rho)-\rho\Gamma(1)}{\rho(1-\rho)}\int_0^1\int_0^\rho \phi(u)h(u,v)\,dv\,du \\
&= \rho^{-1}\Gamma(\rho)\int_0^1 \phi(u)\left(\int_0^\rho h(u,v)\,dv\right)du \\
&\quad -(1-\rho)^{-1}(\Gamma(1)-\Gamma(\rho))\int_0^1\phi(u)\left(-\int_\rho^1 h(u,v)\,dv\right)du \\
&= \int_0^\rho \psi(v)\,dv\int_0^1\phi(u)f(u)\,du + \int_\rho^1\psi(v)\,dv\int_0^1\phi(u)g(u)\,du \\
&= \int\int \phi(u)\psi(v)h(u,v)\,du\,dv.
\end{aligned}$$

When comparing this with the result of Theorem 3.4 for $J(u,v)=\phi(u)\psi(v)$, it appears that for the M_N, the infimum of $\int\int h\log h$ has to be taken over a subset of the set that is considered for the S_N. □

EXAMPLE 5.1. Let $S_N = 2N^{-1}\Sigma_{k=1}^N T_{N,k}$, then $d_N(i) = 2(N-i+1)N^{-1}$, thus $\psi(v)=2(1-v)$ and $\Gamma(v)=v(2-v)$. Hence $\gamma(v)=(v(2-v)-v)/(v(1-v))=1$ and the corresponding max-type statistic is $M_N=\max_{1\le k\le N}T_{N,k}$.

EXAMPLE 5.2. Let $S_N = k^{-1}NT_{N,k}$ with $k/N\to\theta$. Thus S_N is a two-sample statistic. Then

$$\psi(v)=\begin{cases}\theta^{-1}, & 0\le v\le\theta,\\ 0, & \theta<v\le 1,\end{cases}$$

and

$$\gamma(v)=\begin{cases}(1-\theta)\theta^{-1}(1-v)^{-1}, & 0\le v\le\theta,\\ v^{-1}, & \theta<v\le 1.\end{cases}$$

Thus the statistics M_N with this γ-function have Bahadur efficiency ≥ 1 with respect to the two-sample statistics; but in this case, the converse holds too, because

$$\int\int\phi(u)\psi(v)h(u,v)\,du\,dv = \int\phi(u)\left(\theta^{-1}\int_0^\theta h(u,v)\,dv\right)du;$$

i.e., to find the large deviation for the S_N, we may restrict ourselves to the set of distributions with uniform marginals and $h(u,v)=f(u)$ for $0\le v\le\theta$, and $h(u,v)=g(u)$ for $\theta\le v\le 1$, which is a subset of the set over which the infimum has to be taken for the M_N. Therefore for all alternatives (F,G,λ), the Bahadur efficiency of $\{T_{N,k}\}$ with respect to this $\{M_N\}$ is equal to 1.

From the last example it may be concluded that the statistics $\{S_N\}$ need not be inadmissible in the sense of Bahadur efficiency with respect to the corresponding $\{M_N\}$.

6. An example. In case of the M_N-statistic proposed by McGilchrist and Woodyer (1975), with $c_{N,k} = 1$ and median scores, an explicit expression for $\inf I(t; J_\rho)$ can be derived using Theorem 4 in Woodworth (1970),

$$\inf_\rho I(t; J_\rho) = I(t; J_{1/2})$$
$$= \tfrac{1}{2}(1 + 2t)\log(1 + 2t) + \tfrac{1}{2}(1 - 2t)\log(1 - 2t), \qquad 0 < t \leq \tfrac{1}{2}.$$

And, for any alternative (F, G, λ), we have

$$\lim_{N \to \infty} N^{-1}M_N = \lambda \int \phi(u)\, d\bar{F}_\lambda(u) = \lambda\left(2\bar{F}_\lambda(\tfrac{1}{2}) - 1\right) = \lambda(2F(m) - 1),$$

with m the median of the mixture distribution $\lambda F + (1 - \lambda)G$. Thus the slope of M_N can be calculated for every alternative.

Some results are presented in Figure 1, where the efficiency is depicted of the McGilchrist–Woodyer statistic with respect to the optimal likelihood-ratio statistic [Bahadur (1971) and Raghavachari (1970)] when F is normal or double exponential and $G(x) = F(x - \delta)$. It can be seen that, for small δ, the McGilchrist–Woodyer statistic is a good competitor of the likelihood-ratio test if F is double exponential and λ is close to 0.5.

Contrary to what might be expected from Figure 1, the efficiency of the McGilchrist–Woodyer statistic tends to 0 for all λ when δ tends to ∞, both for the normal distribution as well as the double exponential one. For F normal and a fixed λ, this efficiency increases at first with increasing δ; then decreases to 0. The δ at which the efficiency is maximal depends upon λ; the closer λ is to 0.5, the greater the δ for which the efficiency is maximal. For $\lambda = 0.5$, the maximum is attained for $\delta \simeq 2.9$.

Next, consider the corresponding sum-type statistics with median scores (see Example 5.1)

$$S_N = \frac{2}{N} \sum_{k=1}^{N-1} \left(\sum_{i=1}^{k} a_N(R_i) \right) = \frac{2}{N} \sum_{i=1}^{N} (N - i + 1) a_N(R_i).$$

Hence $\psi(v) = 2(1 - v)$ and $\phi(u) = \text{sign}(1 - 2u)$.

Then, at the alternative (F, G, λ), $\lim_{N \to \infty} N^{-1}S_N = \lambda(2\bar{F}_\lambda(\frac{1}{2}) - 1)$ also. Furthermore, under the null hypothesis this S_N has the same distribution as $W_N = 2N^{-1}\sum_{i=1}^{N} d_{N,i} a_N^1(R_i)$ with $d_{N,i} = 1$ if $i \leq [\frac{1}{2}N]$, $d_{N,i} = -1$ if $i > [\frac{1}{2}N]$ and $a_N^1(i) = N - i + 1$. In fact, only the roles of the weights and scores have been interchanged.

But W_N is Wilcoxon's two-sample statistic for two samples of equal size, thus large-deviation probabilities for S_N can be derived from Table 2c in Woodworth (1970), which gives the almost-sure limit and Bahadur slope for W_N at various alternatives. With respect to the McGilchrist–Woodyer statistic, the efficiency of this S_N-statistic turns out to be about 0.75, almost regardless of the values of δ and λ.

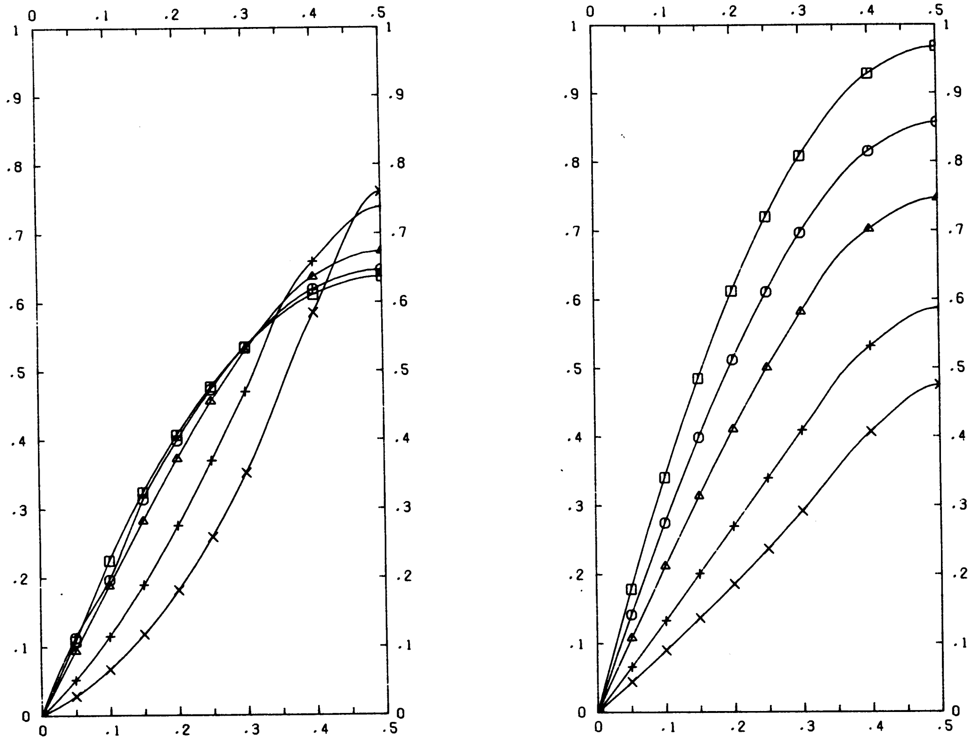

FIG. 1. *Bahadur efficiency of the McGilchrist–Woodyer statistic w.r.t. Raghavachari's upper bound.* (*a*) X_i *normal;* (*b*) X_i *double exponential.* □ $\delta = 0.1$; ○ $\delta = 0.5$; △ $\delta = 1$; + $\delta = 2$; × $\delta = 3$.

7. Efficiency at local alternatives. For alternatives near the null hypothesis, i.e., $G(x) = F(x - \delta)$ and δ small, Woodworth's Theorem 4 can be used again to get approximate expressions for the slopes of S_N and M_N.

Denote $\sigma_\phi^2 := \int \phi^2(u)\,du$ and $\sigma_\psi^2 := \int (\psi(v) - \int \psi(v')\,dv')^2\,dv$ and introduce for every F with density f, $\phi(u; f) = -f'(F^{-1}(u))/f(F^{-1}(u))$. Then, under the same assumptions on scores and weights as before, the slope of $\{S_N\}$ equals

$$\delta^2 \frac{\left(\int_0^\lambda \psi(v)\,dv - \lambda \int \psi(v)\,dv\right)^2}{\sigma_\psi^2} \frac{\left(\int \phi(u)\phi(u; f)\,du\right)^2}{\sigma_\phi^2} + o(\delta^2), \qquad \delta \to 0,$$

and the slope of M_N equals

$$\delta^2 \frac{\left(\sup_\rho\{\gamma(\rho)(\rho(1-\lambda) \wedge \lambda(1-\rho))\}\right)^2}{\sup_\rho \rho(1-\rho)\gamma^2(\rho)} \frac{\left(\int \phi(u)\phi(u; f)\,du\right)^2}{\sigma_\phi^2} + o(\delta^2), \quad \delta \to 0.$$

Thus, at local alternatives the slope of both S_N and M_N is the product of a factor that depends upon the scores (ϕ) only and a factor depending on the weights (ψ or γ) and the position of the change point (λ).

So, the mutual efficiency of two statistics of the same type and with the same score function ϕ does not depend upon that particular ϕ. This independence holds true even for the efficiency of an S_N- (with respect to an M_N-) test with the same scores, because the score-dependent factor is the same for S_N and M_N.

On the other hand, the mutual efficiency of two S_N- (or two M_N-) statistics with the same weight function is independent of the particular weights involved as well as λ. Thus the efficiency equals that of the corresponding two-sample statistics $T_{N,n}$.

Bhattacharyya and Johnson (1968) used Pitman's approach in order to derive asymptotic efficiencies of the sum-type statistics at local alternatives. Their results are equivalent to our local Bahadur efficiencies. For the max-type statistics it is shown in Praagman (1986) that the Pitman efficiencies also have a score- and a weight-dependent factor. Again, this score-dependent factor equals the score-dependent factor that has been found using Bahadur's approach, but, since no explicit expression has been found for the weight-dependent factor, no decisive answer can be given, whether the two approaches coincide in this case also.

Consider $S_N = \sum_{i=1}^N d_{N,i} a_N(R_i)$ and $M_N = \max_{1 \le k < N} c_{N,k} \sum_{i=1}^k a_N(R_i)$ with the same scores $a_N(i)$ and limit weight functions ψ and γ such that, $\gamma(v) = (\Gamma(v) - v\Gamma(1))/(v(1-v))$, where $\Gamma(v) = \int_0^v \psi(s)\,ds$, as in Section 5. From the expressions for the local slopes, it follows easily that the Bahadur efficiency of S_N with respect to M_N at (δ, λ), as $\delta \to 0$, equals $\sup\{\rho^{-1}(1-\rho)^{-1}(\int_0^\rho \psi(v)\,dv - \rho \int \psi(v)\,dv)^2\}\sigma_\psi^{-2}$, which as usual is independent of the scores, but does not depend on λ either. Furthermore, apply the Cauchy–Schwarz inequality to obtain that, for every ρ, $(\int_0^\rho \psi(v)\,dv - \rho\int\psi(v)\,dv)^2 \le \rho(1-\rho)\sigma_\psi^2$, with equality if and only if $\psi(v) - \int\psi(v)\,dv = 1_{[0,\rho]}(v) - \rho$. Thus at local alternatives the Bahadur efficiency of S_N with respect to the corresponding M_N is strictly smaller than 1, uniformly for $\lambda \in (0,1)$; unless S_N is, in fact, a two-sample statistic $T_{N,k}$ for some k, then the efficiency is 1 (see Example 5.2).

Finally, consider the M_N-statistics.

THEOREM 7.1. *If $G(x) = F(x - \delta)$, $\delta \to 0$, then the M_N-statistic with score function $\phi(u) = -\phi(u; f)$ and weight function $\gamma(\rho) = (\rho(1-\rho))^{-1/2}$ is optimal in the sense of Bahadur for all $\lambda \in (0,1)$.*

PROOF. It is easy to show that Raghavachari's upper bound [Raghavachari (1970)] equals $\delta^2\lambda(1-\lambda)\int\phi^2(u; f)\,du + o(\delta^2)$, $\delta \to 0$, for the local alternative (F, G, λ). Then optimality follows, since the slope of the M_N-statistic equals this upper bound. □

This agrees with the result of Hájek (1974), who proved the Bahadur optimality of $\sum_{i=1}^n a_N(R_i)$ for the alternative (F, G, λ), with $n/N \to \lambda$ when the score function is $\phi(u) = \log(\bar{f}/\bar{g})$. In general, this score function depends on λ, however, for the local alternatives we get $\log(\bar{f}/\bar{g}) = \delta\phi(u; f)$, which is independent of λ and thus leads to statistics that are optimal for all λ. Wolfe and Schechtman (1984) proposed an M_N-statistic with the weight function of Theorem 7.1 and Wilcoxon scores. These scores are optimal for logistic distributions; hence the Wolfe and Schechtman statistic is locally Bahadur optimal for the c.p.p. if F is logistic.

Acknowledgments. The author wishes to thank the referees and an Associate Editor for their valuable comments and suggestions.

REFERENCES

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*. SIAM, Philadelphia.

BHATTACHARYYA, G. K. and JOHNSON, R. A. (1968). Nonparametric tests for shift at an unknown time point. *Ann. Math. Statist.* **39** 1731–1743.

DESHAYES, J. and PICARD, D. (1982). Tests of disorder of regression: Asymptotic comparison. *Theory Probab. Appl.* **27** 100–115.

GROENEBOOM, P. (1979). *Large Deviations and Asymptotic Efficiencies*. Centre for Mathematics and Computer Science, Amsterdam.

GROENEBOOM, P. and OOSTERHOFF, J. (1977). Bahadur efficiency and probabilities of large deviations. *Statist. Neerlandica* **31** 1–24.

GROENEBOOM, P., OOSTERHOFF, J. and RUYMGAART, F. H. (1979). Large deviation theorems for empirical probability measures. *Ann. Probab.* **7** 553–586.

HACCOU, P., MEELIS, E. and VAN DE GEER, S. (1985). On the likelihood ratio test for a change-point in a sequence of independent exponentially distributed random variables. Report MS-R8507, Centre for Mathematics and Computer Science, Amsterdam.

HÁJEK, J. (1974). Asymptotic sufficiency of the vector of ranks in the Bahadur sense. *Ann. Statist.* **2** 75–83.

KULLBACK, S. (1959). *Information Theory and Statistics*. Wiley, New York.

MCGILCHRIST, C. A. and WOODYER, K. D. (1975). Note on a distribution-free cusum technique. *Technometrics* **17** 321–325.

PETTITT, A. N. (1979). A non-parametric approach to the change-point problem. *Appl. Statist.* **28** 126–135.

PRAAGMAN, J. (1986). Efficiency of change-point tests. Ph.D. thesis, Eindhoven Univ. Technology.

RAGHAVACHARI, M. (1970). On a theorem of Bahadur on the rate of convergence of test statistics. *Ann. Math. Statist.* **41** 1695–1699.

RUYMGAART, F. H. and VAN ZUYLEN, M. C. A. (1978). Asymptotic normality of multivariate linear rank statistics in the non-i.i.d. case. *Ann. Statist.* **6** 588–602.

WOLFE, D. A. and SCHECHTMAN, E. (1984). Nonparametric statistical procedures for the change point problem. *J. Statist. Plann. Inference* **9** 389–396.

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

FACULTY OF INDUSTRIAL ENGINEERING
UNIVERSITY OF TECHNOLOGY
P.O. Box 513
5600 MB EINDHOVEN
NETHERLANDS