Kybernetika

Konrad Behnen; Georg Neuhaus

Likelihood ratio rank tests for the two-sample problem with randomly censored data

Kybernetika, Vol. 27 (1991), No. 2, 81--99

Persistent URL: http://dml.cz/dmlcz/124518

KYBERNETIKA — VOLUME 27 (1991), NUMBER 2

LIKELIHOOD RATIO RANK TESTS FOR THE TWO-SAMPLE PROBLEM WITH RANDOMLY CENSORED DATA*

KONRAD BEHNEN, GEORG NEUHAUS

In the two sample problem with random censorship likelihood ratio rank tests are constructed for testing randomness versus cones of alternatives which are generated by a finite number of suitable score functions. These nonlinear rank tests improve the range of power sensitivity of the corresponding linear rank tests. The applicability of the asymptotics is demonstrated by Monte Carlo simulation.

1. INTRODUCTION

The aim of the present paper is to construct a class of likelihood ratio rank tests for the two sample problem of testing randomness versus general stochastically larger alternatives under the assumption of random censorship. The new tests are designed to have larger ranges of sensitivity than the usual linear rank tests. In order to achieve this goal we replace the single direction of asymptotic alternatives corresponding to the optimal score function of a linear rank test by a cone which is generated by a finite number of suitable score functions. The actual choice of the proposed cone is based on the consideration of so-called generalized shift alternatives, where the amount of shift may depend on the value of the observation. This seems to be more realistic than the classical shift model, which assumes that the two treatments differ by constant shift only. In contrast, the family of generalized shift alternatives may be identified with the nonparametric alternative that the distribution of the first sample is stochastically larger than the distribution of the second sample.

The sample sizes will be m (first sample) and n (second sample), and $N = m + n$ is the pooled sample size. Let $(X_1, \Delta_1), \ldots, (X_m, \Delta_m)$ and $(X_{m+1}, \Delta_{m+1}), \ldots, (X_N, \Delta_N)$ be the observable random variables (rv's) of the respective first and second sample,

* Presented at the "Kolloquium über Mathematische Statistik im Rahmen der Wissenschaftlichen Kolloquien der Universität Hamburg und der Karls-Universität Prag", Hamburg, June 1989.

i.e., the value of X_i is the ith observation, where $\Delta_i = 1$ indicates an uncensored observation and $\Delta_i = 0$ means that the value of X_i results from censoring.

In order to specify the general model we use the unobservable random variables $X_{11}, \dots, X_{1N}$ for the potential measurements and the unobservable random variables $X_{21}, \dots, X_{2N}$ for the potential censoring points, i.e. the observable pairs (X_i, Δ_i) are defined as (for $i = 1, \dots, N$)

$$X_i = \min(X_{1i}, X_{2i}), \quad \Delta_i = 1(X_{1i} \leqq X_{2i}), \tag{1.1}$$

where $1(X_{1i} \leqq X_{2i})$ is the indicator function of the event $\{X_{1i} \leqq X_{2i}\}$.

We assume the first sample $X_{11}, \dots, X_{1m}$ of potential measurements to be i.i.d. with continuous df F_1, the second sample $X_{1m+1}, \dots, X_{1N}$ of potential measurements to be i.i.d. with continuous df F_2, and the censoring rv's $X_{21}, \dots, X_{2N}$ to be i.i.d. with continuous df G (*equal censoring*). Additionally we assume the censoring rv's $(X_{21}, \dots, X_{2N})$ and the measurement rv's $(X_{11}, \dots, X_{1N})$ to be stochastically independent (*random censoring*).

Our aim will be the construction of tests with suitable power properties for testing the *null hypothesis of randomness*

$$\mathscr{H}_0 \colon F_1 = F_2 \tag{1.2}$$

versus the nonparametric alternative

$$\mathscr{H}_1 \colon F_1 \leqq F_2, \quad F_1 \neq F_2 \tag{1.3}$$

that the underlying distribution F_1 of the first sample $X_{11}, \dots, X_{1m}$ is *stochastically larger* than the underlying distribution F_2 of the second sample $X_{1m+1}, \dots, X_{1N}$.

In fact, only the ranks $R_1, \dots, R_N$ of the pooled sample $X_1, \dots, X_N$ and the censoring indicators $\Delta_1, \dots, \Delta_N$ will be used in the definition of the proposed tests.

In order to achieve our goal we use asymptotic theory ($N \to \infty$) for suitable cones in $\mathscr{H}_1$:

For any total sample size $N \geqq 2$ we assume sample sizes $m = m_N$ and $n = n_N$ such that $m \to \infty$ and $n \to \infty$ as $N \to \infty$. For given dimension $r \in \mathbb{N}$ and any column vector $\vartheta \in \mathbb{R}_r$ we define *asymptotic generalized shift alternatives* at F according to

$$\mathscr{L}(X_{1i}) \sim F(x - c_{Ni}\vartheta^{\mathrm{T}} D(x)), \tag{1.4}$$

$$c_{Ni} = \sqrt{\left(\frac{nm}{N}\right)} \begin{cases} 1/m & \text{as} \quad 1 \leqq i \leqq m, \\ -1/n & \text{as} \quad m+1 \leqq i \leqq N, \end{cases} \tag{1.5}$$

where F is a given df with absolutely continuous density f and finite positive Fisher-information.

$$0 < I(f) = \int (f'/f)^2 \, \mathrm{d}F < \infty, \tag{1.6}$$

and where $D = (D_1, \dots, D_r)^{\mathrm{T}}$ is a given vector of bounded *generalized shift functions* $D_\varrho \colon \mathbb{R} \to (0, \infty)$ with bounded and continuous derivative d_ϱ, $\varrho = 1, \dots, r$.

Defining

$$T_t(x) := x - t\vartheta^{\mathrm{T}} D(x), \quad x \in \mathbb{R}, \tag{1.7}$$

we notice that the assumptions on D and $d = (D_1', \dots, D_r')^\mathrm{T}$ imply the function T_t: $\mathbb{R} \to \mathbb{R}$ to be bijective, strictly increasing, and continuously differentiable with derivative $T_t' = 1 - t\vartheta^\mathrm{T} d$, if the factor $t \in \mathbb{R}$ fulfils the condition

$$|t| \, \|\vartheta^\mathrm{T} d\|_\infty < 1 \tag{1.8}$$

with $\|\cdot\|_\infty$ being the supnorm. Thus, if N is sufficiently large (or if $|\vartheta|$ is sufficiently small), formula (1.4) defines a proper distribution with df $F \circ T_{c_{Ni}}$ and continuous density $f_{c_{Ni}}$ according to

$$f_t = (f \circ T_t)(1 - t\vartheta^\mathrm{T} d) . \tag{1.9}$$

Obviously the assumption $\vartheta \geqq 0$ (componentwise) implies

$$F_1(x) := F(x - c_{N1}\vartheta^\mathrm{T} D(x)) \leqq F(x - c_{NN}\vartheta^\mathrm{T} D(x)) =: F_2(x) \quad \forall x \in \mathbb{R} ,$$

i.e. the distribution of the first sample is stochastically larger than the distribution of the second sample. Conversely, if $F_1 \leqq F_2$ holds true then there exists some function D: $\mathbb{R} \to [0, \infty)$ with the property $F_1(x) = F_2(x - D(x))$ $\forall x$. Therefore arbitrary stochastic larger alternatives may be described by generalized shift functions.

Thus, testing the null hypothesis $\vartheta = 0$ versus the alternative hypothesis $\vartheta \geqq 0$, $\vartheta \neq 0$ under the model (1.4) (with fixed F and fixed vector $D = (D_1, \dots, D_r)^\mathrm{T}$ of generalized shift functions) can be viewed as a sub-problem of the general problem of testing $\mathscr{H}_0$ versus $\mathscr{H}_1$. Obviously $\vartheta \geqq 0$ specifies a cone in the general model $\mathscr{H}_0 \cup \mathscr{H}_1$ which corresponds to (F, D).

In the next section we prove the local asymptotic normality of the corresponding sequence of statistical experiments and derive asymptotic likelihood ratio tests for testing $\vartheta = 0$ versus $\vartheta \geqq 0$, $\vartheta \neq 0$.

The final version of the proposed asymptotic likelihood ratio test will be based on an r-vector $S_N = (S_{1N}, \dots, S_{rN})^\mathrm{T}$ of linear rank statistics $S_{\varrho N}$ of the form (for $\varrho = 1, \dots, r$)

$$S_{\varrho N} = \sum_{i=1}^{N} c_{Ni}(\Delta_i b^{(1)}_{\varrho N R_i} + (1 - \Delta_i)\, b^{(2)}_{\varrho N R_i}) , \tag{1.10}$$

where $b^{(1)}_{\varrho N1}, \dots, b^{(1)}_{\varrho NN}$ and $b^{(2)}_{\varrho N1}, \dots, b^{(2)}_{\varrho NN}$ are two sets of scores and where $R_1, \dots, R_N$ are the ranks of the observable sample $X_1, \dots, X_N$.

The resulting (conditional) rank test will be distribution free under the general null hypothesis of randomness $\mathscr{H}_0$, i.e. the (conditional) null distribution will be the same for any $F_1 = F_2$ and any censoring df G.

2. LOCAL ASYMPTOTIC NORMALITY AND THE ASYMPTOTIC LIKELIHOOD RATIO TEST

In addition to (1.4) and (1.6) we assume that the censoring df G has a Lebesgue (λ) density g and, naturally,

$$F\{x: G(x) < 1\} > 0 . \tag{2.1}$$

If ν denotes the counting measure on the set $\{0, 1\}$ and if $\mu = \lambda \times \nu$ denotes the product measure of λ and ν on $\mathbb{R} \times \{0, 1\}$ it's easily proved that $f(z; c_{Ni})$ according to

$$f(z;t) = \delta(1 - G(x)) f_t(x) + (1 - \delta)(1 - F \circ T_t(x))\, g(x),$$
$$z = (x, \delta) \in \mathbb{R} \times \{0, 1\}, \tag{2.2}$$

is a μ-density of the distribution of the observable rv $Z_i = (X_i, \Delta_i)$, cf. (1.7) to (1.9).

Under the assumptions (1.6) and (2.1) Fisher's information $J(f, g, \vartheta^{\mathrm{T}} D)$ of $f(\cdot\,; t)$ at $t = 0$ is finite and positive, if $(\vartheta^{\mathrm{T}} D)^2 > 0$ holds true:

$$J := J(f, g, \vartheta^{\mathrm{T}} D) = \mathsf{E}_{t=0}\left(\frac{\partial \log f(Z;t)}{\partial t}\bigg|_{t=0}\right)^2$$
$$= \int \left(-\frac{f'}{f}\,\vartheta^{\mathrm{T}} D - \vartheta^{\mathrm{T}} d\right)^2 (1 - G)\,\mathrm{d}F + \int \frac{f^2}{1 - F} (\vartheta^{\mathrm{T}} D)^2\,\mathrm{d}G \in (0, \infty).$$

Since $f^2(\vartheta^{\mathrm{T}} D)^2/(1 - F) > 0$ $[F]$, formula (2.1) implies $J > 0$. The first of the two integrals clearly is finite because of (1.6). The finiteness of the last integral follows from the boundedness of $\vartheta^{\mathrm{T}} D$ and $f^2/(1 - F)$, the latter following from $f(x) = -\int_x^\infty f'\,\mathrm{d}\lambda$ and

$$f^2(x) = \left(\int_x^\infty f'\,\mathrm{d}\lambda\right)^2 = \left(\int_x^\infty \left(\frac{f'}{f}\right) \mathrm{d}F\right)^2 \leqq I(f)\,(1 - F(x)) \tag{2.3}$$

by the Cauchy-Schwarz inequality.

Under the above assumptions the densities (1.9) are 2-differentiable at $t = 0$, more exactly,

$$\frac{2}{t}\left(f_t^{1/2} - f^{1/2}\right) \to f^{1/2}\left(-\frac{f'}{f}\,\vartheta^{\mathrm{T}} D - \vartheta^{\mathrm{T}} d\right) =: h \quad \text{as} \quad t \to 0 \tag{2.4}$$

in quadratic mean with respect to the Lebesgue measure λ on $\mathbb{R}$.

Proof of (2.4). If $\vartheta^{\mathrm{T}} d = 0$ we are in the classical shift situation for which (2.4) is well-known, see e.g. [2] p. 211/212.

Now assume $\vartheta^{\mathrm{T}} d \neq 0$. Since the derivatives d_ϱ are bounded we may assume that the condition (1.8) is fulfilled. Then the function T_t defined in (1.7) is bijective, strictly increasing, and continuously differentiable on $\mathbb{R}$. The same holds true for the inverse function T_t^{-1}. Therefore we have the following inequality for $s \neq 0$, $|s|\,\|\vartheta^{\mathrm{T}} d\|_\infty < 1$, cf. [2] p. 212,

$$\int_{\mathbb{R}} \left(\frac{2}{s}\left(f_s^{1/2} - f^{1/2}\right)\right)^2 \mathrm{d}\lambda \leqq \frac{4}{s}\int_0^s \left[\int_{\mathbb{R}} \left(\frac{\partial f_t^{1/2}}{\partial t}\right)^2 \mathrm{d}\lambda\right] \mathrm{d}t$$
$$= \frac{1}{s}\int_0^s \left[\int_{\mathbb{R}} \frac{((f' \circ T_t)(-\vartheta^{\mathrm{T}} D)(1 - t\vartheta^{\mathrm{T}} d) + (f \circ T_t)(-\vartheta^{\mathrm{T}} d))^2}{(f \circ T_t)(1 - t\vartheta^{\mathrm{T}} d)}\,\mathrm{d}\lambda\right] \mathrm{d}t$$
$$= \frac{1}{s}\int_0^s \left[\int_{\mathbb{R}} \left(\left(\frac{f'}{f}\right)(\vartheta^{\mathrm{T}} D \circ T_t^{-1}) + \frac{\vartheta^{\mathrm{T}} d \circ T_t^{-1}}{1 - t\vartheta^{\mathrm{T}} d \circ T_t^{-1}}\right)^2 f\,\mathrm{d}\lambda\right] \mathrm{d}t,$$

since $(T_t^{-1})' = 1/(1 - t\vartheta^T d \circ T_t^{-1})$. According to the Lebesgue convergence theorem, under the above assumptions on f and D, the integral in square brackets is continuous in t in a neighbourhood of $t = 0$. Consequently, the lim $\sup_{s \to 0}$ of the L.H.S. doesn't exceed $\int h^2 \, d\lambda$. Since the convergence in (2.4) holds true λ – a.s., Vitali's theorem implies the convergence (2.4) in quadratic mean. □

Without any further assumption on G the quadratic mean convergence (2.4) implies

$$\frac{2}{t}(f^{1/2}(\cdot;t) - f^{1/2}(\cdot;0)) \to h_1 \quad \text{in} \quad L_2(\lambda \times \nu) \quad \text{as} \quad t \to 0 \tag{2.5}$$

with

$$h_1(x,\delta) = \delta(1 - G(x))^{1/2}\, h(x) + (1 - \delta)\left(\frac{g(x)}{1 - F(x)}\right)^{1/2} f(x)\, \vartheta^T D(x),$$

which is equivalent to

$$\frac{2}{t}(f^{1/2}(\cdot,\delta;t) - f^{1/2}(\cdot,\delta;0)) \to h_1(\cdot,\delta) \quad \text{in} \quad L_2(\lambda) \quad \text{as} \quad t \to 0$$

for $\delta = 0$ and $\delta = 1$. For $\delta = 1$ the assertion immediately follows from (2.4). For $\delta = 0$ the pointwise convergence is clear, and Vitali's theorem is applicable, since we have

$$\frac{2}{t}(f^{1/2}(x,0;t) - f^{1/2}(x,0;0)) = g^{1/2}(x)\,(\vartheta^T D(x))\left(\frac{f^2}{1 - F} \circ T_t(x)\right)^{1/2}$$

for some suitable t_x between 0 and t, and since the R.H.S. is bounded by $g^{1/2}(x)$ . . $\|\vartheta^T D\|_\infty I^{1/2}(f)$, cf. inequality (2.3).

Recently [3] has derived (2.5) in a more general (not necessarily translation) setting.

In the sequel let $P_{N\vartheta}$ denote the *joint* distribution of the observable rv's $(X_1, \Delta_1), \ldots$ $\ldots, (X_N, \Delta_N)$ under the (local asymptotic) model (1.4), i.e. for any $N \geq 2$ and any $i = 1, \ldots, N$ the df of X_{1i} is $F \circ T_{c_{Ni}}$ and the df of X_{2i} is G.

Let H denote the df of X_i under the null hypothesis P_{N0}, i.e.,

$$H = 1 - (1 - F)(1 - G),$$

and define the random vector $T_N = (T_{1N}, \ldots, T_{rN})^T$ by (for $\varrho = 1, \ldots, r$)

$$T_{\varrho N} = \sum_{i=1}^{N} c_{Ni}(\Delta_i b_\varrho^{(1)}(H(X_i)) + (1 - \Delta_i)\, b_\varrho^{(2)}(H(X_i))) \tag{2.6}$$

with

$$\begin{aligned} b_\varrho^{(1)} &= \left(-\frac{f'}{f} \circ H^{-1}\right)(D_\varrho \circ H^{-1}) - (d_\varrho \circ H^{-1}), \\ b_\varrho^{(2)} &= \left(\frac{f}{1 - F} \circ H^{-1}\right)(D_\varrho \circ H^{-1}), \end{aligned} \tag{2.7}$$

where H^{-1} is the left-continuous inverse of H.

Using the *regression functions* p_j: $(0, 1) \to [0, 1]$ according to

$$p_1(u) = P_{N0}\{\Delta_1 = 1 \mid H(X_1) = u\}\,,$$
$$p_2(u) = 1 - p_1(u)\,, \quad 0 < u < 1\,, \tag{2.8}$$

we define an inner product $\langle\cdot, \cdot\rangle_0$ on the space $L_2^2 = L_2(0, 1) \times L_2(0, 1)$ by (for all $(a_1, a_2), (b_1, b_2) \in L_2^2$)

$$\langle(a_1, a_2), (b_1, b_2)\rangle_0 = \textstyle\int_0^1 a_1 b_1 p_1\, \mathrm{d}\lambda + \int_0^1 a_2 b_2 p_2\, \mathrm{d}\lambda\,,$$
$$= \textstyle\int (a_1 \circ H)(b_1 \circ H)(1 - G)\, \mathrm{d}F + \int (a_2 \circ H)(b_2 \circ H)(1 - F)\, \mathrm{d}G\,. \tag{2.9}$$

Finally, we assume that the $r \times r$-matrix

$$\Gamma = (\langle(b_\varrho^{(1)}, b_\varrho^{(2)}), (b_\sigma^{(1)}, b_\sigma^{(2)})\rangle_0)_{\varrho,\sigma=1,\ldots,r} \tag{2.10}$$

corresponding to the functions (2.7) exists and is non-singular. (Especially the proportional hazard assumption (3.3) will imply the existence of Γ.)

Under the above assumptions and notations the q.m.-differentiability (2.5) and [8] (Theorem 79.2 and Lemma 80.11) yield the following LAN-result.

Theorem 2.1. Assume $m_N/N \to \eta \in (0, 1)$ as $N \to \infty$. Then, for any given $\vartheta \in \mathbb{R}^r$, we have the limiting law

$$\mathscr{L}(T_N \mid P_{N\vartheta}) \to^{\mathscr{L}} \mathscr{N}(\Gamma\vartheta, \Gamma) \tag{2.11}$$

and the approximation

$$\frac{\mathrm{d}P_{N\vartheta}}{\mathrm{d}P_{N0}} = \exp\left(\vartheta^{\mathrm{T}} T_N - \tfrac{1}{4}\, \vartheta^{\mathrm{T}} \Gamma \vartheta + {}^{0}P_{N0}(1)\right). \tag{2.12}$$

Notice, because of (2.7) we have

$$\mathsf{E}_{\vartheta=0}(\Delta_i b_\varrho^{(1)}(H(X_i)) + (1 - \Delta_i)\, b_\varrho^{(2)}(H(X_i)))$$
$$= \textstyle\int (b_\varrho^{(1)} \circ H)(1 - G)\, \mathrm{d}F + \int (b_\varrho^{(2)} \circ H)(1 - F)\, \mathrm{d}G$$
$$= \textstyle\int fgD_\varrho\, \mathrm{d}\lambda - \int fgD_\varrho\, \mathrm{d}\lambda = 0\,.$$

Remark 2.2. For any $\varrho = 1, \ldots, r$ let $b_{\varrho N1}^{(1)}, \ldots, b_{\varrho NN}^{(1)}$ and $b_{\varrho N1}^{(2)}, \ldots, b_{\varrho NN}^{(2)}$ be two sets of scores such that the corresponding jump-functions

$$b_{\varrho N}^{(j)}(u) = \sum_{i=1}^{N} b_{\varrho Ni}^{(j)}\, 1\left(\frac{i-1}{N} \leqq u < \frac{i}{N}\right), \quad 0 < u < 1\,, \tag{2.13}$$

converge to $b_\varrho^{(j)}$ in $L_2(0, 1)$, $j = 1, 2$. Then the assertions of Theorem 2.1 remain true if T_N is replaced by the statistic S_N defined in formula (1.10), cf. [5] Theorem 3.3. □

Now we use Theorem 2.1 in order to construct an asymptotic likelihood ratio test for testing $\{P_{N0}\}$ versus $\{P_{N\vartheta}\colon \vartheta \geqq 0, \vartheta \neq 0\}$:

According to (2.12) and the remark the likelihood ratio $\mathrm{d}P_{N\vartheta}/\mathrm{d}P_{N0}$ is approximated by $\exp(\vartheta^{\mathrm{T}} S_N - \frac{1}{2}\vartheta^{\mathrm{T}}\Gamma\vartheta)$. Therefore, the asymptotic likelihood ratio test is based on the likelihood ratio statistic $LR(\Gamma, S_N)$ of the limiting model, i.e.

$$LR(\Gamma, x) = \sup_{\vartheta \geqq 0} (2\vartheta^{\mathrm{T}} x - \vartheta^{\mathrm{T}} \Gamma \vartheta), \quad x \in \mathbb{R}^r . \tag{2.14}$$

It's well-known how to evaluate the supremum $LR(\Gamma, x)$, cf. [4], [6], [7]: If $\Gamma = (\gamma_{\varrho\sigma})_{\varrho,\sigma=1,\dots,r}$ is positive definite then $(\forall x \in \mathbb{R}^r)$

$$LR(\Gamma,x) = \max \{x_J^{\mathrm{T}}(\Gamma_{J\times J})^{-1} x_J : (\Gamma_{J\times J})^{-1} x_J \geqq 0, \emptyset \neq J \subset R\} , \tag{2.15}$$

where $R = \{1, \dots, r\}$, $x_J = (x_i : i \in J)^{\mathrm{T}}$, and $\Gamma_{J\times J}$ is the $|J| \times |J|$ – matrix $(\gamma_{\varrho\sigma})_{\varrho,\sigma\in J}$. If the dimension r is small the evaluation of (2.15) is quite easy. In Section 3 we choose $r = 3$.

Another representation of $LR(\Gamma, x)$ is given by

$$LR(\Gamma, x) = (\max (0, \sup \{\vartheta^{\mathrm{T}} x : \vartheta^{\mathrm{T}} \Gamma \vartheta = 1\}))^2 . \tag{2.16}$$

Making use of the compactness of $\{\vartheta : \vartheta^{\mathrm{T}} \Gamma \vartheta = 1\}$ we can prove that the function $LR(\Gamma, x)$ is continuous in (Γ, x) on the set

$$\{(\Gamma, x) \in \mathbb{R}^{r\times r} \times \mathbb{R}^r : \Gamma \text{ positive definite}\} .$$

Using the continuity of $LR(\Gamma, x)$ in x the limiting law (2.11) and Remark 2.2 imply the convergence in distribution of $LR(\Gamma, S_N)$ under $\{P_{N\vartheta}\}$. In the presence of censoring these asymptotic results are of limited value, since S_N and $LR(\Gamma, \cdot)$ are constructed on the basis of fixed df's F and G. In practice, however, the actual df's F^* and G^*, say, are unknown. Therefore, even under the null hypothesis corresponding to F^* and G^*, the limiting distribution of $LR(\Gamma, S_N)$ will depend on F^* and G^*. Moreover, this dependence is too complicated as to estimate the unknown quantities of the limiting distribution by the data. Only in the case $F = F^*$, $G = G^*$ the limiting distribution will be quite simple, namely a mixture of χ^2-distributions. cf. [4], [6], [7].

We escape these difficulties by the conditioning device of [5]. The crucial fact is that for arbitrary continuous df's F^* and arbitrary continuous censoring df's G^* the rank vector $R = (R_1, \dots, R_N)$ and the vector $\Delta^{(\,)} = \Delta_N^{(\,)} = (\Delta^{(1)}, \dots, \Delta^{(N)})$ of censoring rv's corresponding to the ordered observations $(X^{(1)}, \dots, X^{(N)})$ of $(X_1, \dots, X_N)$ are *independent under the null hypothesis* $\mathscr{H}_0$. In this situation the distribution of R is the uniform distribution $\mathscr{U}_N$ on the set of all permutations of $(1, \dots, N)$, shortly: $R \sim \mathscr{U}_N$. Likewise, the vector of *antiranks*

$$D = (D_1, \dots, D_N) \quad \text{defined by} \quad R_{D_i} = i \quad \forall 1 \leqq i \leqq N$$

has the uniform distribution $\mathscr{U}_N$ under $\mathscr{H}_0$.

Thus, rewriting the definition (1.10) of $S_N = (S_{1N}, \dots, S_{rN})^{\mathrm{T}}$ according to $S_N(D, \Delta^{(\,)})$, where

$$S_{\varrho N} = S_{\varrho N}(D, \Delta^{(\,)}) = \sum_{i=1}^{N} c_{ND_i}(\Delta^{(i)} b_{\varrho Ni}^{(1)} + (1 - \Delta^{(i)})\, b_{\varrho Ni}^{(2)}) , \tag{2.17}$$

leads to the idea of conditioning the test statistic $S_N(D, \Delta^{(\,)})$ on the observed value

$\delta \in \{0, 1\}^N$ of $\Delta^{(\ )}$. The corresponding test function may be written in the form

$$\varphi_N(d, \delta) = \begin{cases} 1, & \text{if } LR(\Gamma, S_N(d, \delta)) \geqq c_N(\alpha, \delta), \\ 0, & \text{otherwise}, \end{cases} \tag{2.18}$$

where $d = (d_1, \ldots, d_N)$ is any permutation of $(1, \ldots, N)$ and where $c_N(\alpha, \delta)$ is the upper $(1 - \alpha)$-quantile of the distribution of $LR(\Gamma, S_N(D, \delta))$ under $D \sim \mathscr{U}_N$ and fixed $\delta \in \{0, 1\}^N$. Then, under the null hypothesis point corresponding to $F_1 = F_2 = F^*$ and censoring df G^* we have

$$\mathsf{E}_{F^*, G^*} \varphi_N(D, \Delta^{(\ )}) \leqq \alpha .$$

In the next theorem we prove the convergence

$$c_N(\alpha, \Delta_N^{(\ )}) \to c(\alpha, F^*, G^*) \quad \text{in probability} \tag{2.19}$$

under $F_1 = F_2 = F^*$, G^*, where $c(\alpha, F^*, G^*)$ is the upper $(1 - \alpha)$-quantile of the unconditional limiting distribution of $LR(\Gamma, S_N)$ under $F_1 = F_2 = F^*$, G^*. This implies the asymptotic equivalence of the test sequences $\{\varphi_N\}$ and $\{\psi_N\}$ where $\psi_N = 1(LR(\Gamma, S_N) \geqq c(\alpha, F^*, G^*))$. Under $F_1 = F_2 = F^*, G^*$ the random vector S_N converges in distribution to some normally distributed random vector S with mean zero and covariance matrix

$$\Gamma^* = (\langle (b_\varrho^{(1)}, b_\varrho^{(2)}), (b_\sigma^{(1)}, b_\sigma^{(2)}) \rangle_* - \langle (b_\varrho^{(1)}, b_\varrho^{(2)}), (1, 1) \rangle_* \langle (b_\sigma^{(1)}, b_\sigma^{(2)}), (1, 1) \rangle_*)_{\varrho, \sigma = 1, \ldots, r}, \tag{2.20}$$

where $\langle (\cdot, \cdot), (\cdot, \cdot) \rangle_*$ is defined in the same way as $\langle (\cdot, \cdot), (\cdot, \cdot) \rangle_0$ but with F, G, and $H = 1 - (1 - F)(1 - G)$ replaced by F^*, G^*, and H^*, cf. formula (2.9). This result is an immediate consequence of [5] Theorem 3.3.

If Γ and Γ^* are non-singular the df of $LR(\Gamma, S)$ $(\geqq 0)$ is continuous on the interval $(0, \infty)$, since (2.15) implies

$$P\{LR(\Gamma, S) = c\} \leqq \sum_{\emptyset \neq J \subset R} P\{S_J^{\mathrm{T}} (\Gamma_{J \times J})^{-1} S_J = c\} = 0 \quad \forall c > 0 .$$

Moreover, the df of $LR(\Gamma, S)$ is strictly increasing:

For any $0 < a < b$ we get

$$P\{LR(\Gamma, S) \in (a, b)\} = P\{S \in (LR)^{-1} (\Gamma, (a, b))\} > 0 ,$$

since $(LR)^{-1} (\Gamma, (a, b))$ is nonvoid and open, and since S has a nondegenerate normal distribution.

Theorem 2.3. Let $(\Omega, \mathscr{A}, P)$ be the basic probability space on which the rv's X_{1i}, X_{2i}, $X_i = \min(X_{1i}, X_{2i})$, $i \geqq 1$, are defined such that all X_{ji} are independent, and $X_{1i} \sim F^*$, $X_{2i} \sim G^*$ $\forall i$ with arbitrary continuous df's F^*, G^*.

For $\varrho = 1, \ldots, r$ and $N \geqq 1$ let $b_{\varrho N}^{(1)}$ and $b_{\varrho N}^{(2)}$ be score functions (2.13) which converge in $L_2(0, 1)$ to $b_\varrho^{(1)}$ and $b_\varrho^{(2)}$, respectively. Let $F_N(\delta, \cdot)$ be the df of $LR(\Gamma, S_N(D, \delta))$ under $D \sim \mathscr{U}_N$ and fixed $\delta \in \{0, 1\}^N$, and let F_0 be the df of $LR(\Gamma, S)$.

Assume Γ and Γ^* to be non-singular.

Then

$$A_N := \sup_{-\infty < t < \infty} |F_N(\Delta_N^{(\,)}, t) - F_0(t)| \to \quad \text{in probability}\,, \tag{2.21}$$

and also the convergence (2.19) holds true.

Proof. The proof closely follows the pattern of the proof of [5] Theorem 5.2. For any $\delta = (\delta_1, \ldots, \delta_N) \in \{0, 1\}^N$ and $\varrho = 1, \ldots, r$ define

$$\beta_{\varrho N}^*(i, \delta) = \delta_i\, b_{\varrho N}^{(1)}\left(\frac{i}{N+1}\right) + (1 - \delta_i)\, b_{\varrho N}^{(2)}\left(\frac{i}{N+1}\right) \tag{2.22}$$

and $\bar{\beta}_{\varrho N}^*(\delta) = 1/N \sum_{i=1}^{N} \beta_{\varrho N}^*(i, \delta)$. Putting $\beta_{\varrho N}^*(i) = \beta_{\varrho N}^*(i, \Delta_N^{(\,)})$ and $\bar{\beta}_{\varrho N}^* = \bar{\beta}_{\varrho N}^*(\Delta_N^{(\,)})$ we get in the same way as in Theorem 5.1 of [5] the convergence

$$\frac{1}{N} \sum_{i=1}^{N} (\beta_{\varrho N}^*(i) - \bar{\beta}_{\varrho N}^*)\,(\beta_{\sigma N}^*(i) - \bar{\beta}_{\sigma N}^*) \to \langle (b_\varrho^{(1)}, b_\varrho^{(2)}), (b_\sigma^{(1)}, b_\sigma^{(2)}) \rangle_* \tag{2.23}$$

in probability, $\varrho, \sigma = 1, \ldots, r$. Likewise,

$$\frac{1}{N} \max_{1 \leqq i \leqq N} |\beta_{\varrho N}^*(i) - \bar{\beta}_{\varrho N}^*| \to 0 \quad \text{in probability}\,. \tag{2.24}$$

In order to prove (2.21) take an arbitrary subsequence $\mathbb{N}_1$ of $\mathbb{N} = \{1, 2, \ldots\}$. Then, according to (2.23) and (2.24), there exists a further subsequence $\mathbb{N}_2 = \{r_1, r_2, \ldots\}$ of $\mathbb{N}_1$ with $r_1 < r_2 < \ldots$ and a set $M \in \mathscr{A}$ with $P(M) = 0$ such that the condition (N) of [2] p. 195 holds true for

$$\sum_{\varrho=1}^{r} \lambda_\varrho \beta_{\varrho r_\nu}^*(i, \Delta_{r_\nu}^{(\,)}(\omega))\,, \quad 1 \leqq i \leqq r_\nu,\ \nu \geqq 1\,, \quad \forall \lambda_1, \ldots, \lambda_r \in \mathbb{R}\,,$$

as long as $\omega \notin M$. According to the Cramér-Wold device and to Problem 8 in [2] p. 195 we get $S_{r_\nu}(D, \Delta_{r_\nu}^{(\,)}) \xrightarrow{\mathscr{D}} S$ as $\nu \to \infty$, if $\omega \notin M$. Using the continuity of $LR(\Gamma, \cdot)$ the continuity of F_0 on $(0, \infty)$, and $P\{S \in \partial\{x\colon LR(\Gamma, x) = 0\}\} = 0$ we get $A_{r_\nu}(\omega) \to 0$ as $\nu \to \infty$ $\forall \omega \notin M$. Since $\mathbb{N}_1 \subset \mathbb{N}$ is arbitrary and since convergence in probability is a metric convergence, the proof of (2.21) is concluded. Since convergence in distribution implies the convergence of all quantiles, if the limiting distribution function is strictly increasing, the above assumptions imply $c_{r_\nu}(\alpha, \Delta_{r_\nu}^{(\,)}(\omega)) \to$ $\to F_0^{-1}(1 - \alpha)$ as $\nu \to \infty$ $\forall \omega \notin M$. This proves the assertion (2.19). □

Now let's consider the conditional distribution of S_N (given the $\Delta_N^{(\,)}$ – vector) under the null hypothesis $F_1 = F_2 = F^*$ and G^*. In the first step we evaluate the covariance matrix $\Gamma_N(\delta)$ of $\mathscr{L}_{F^*,G^*}[S_N \mid \Delta_N^{(\,)} = \delta]$:

From (2.17) and (2.22) we get

$$S_{\varrho N} = \sum_{i=1}^{N} c_{Ni} \beta_{\varrho N}^*(R_i, \Delta_N^{(\,)})\,, \quad \varrho = 1, \ldots, r\,. \tag{2.25}$$

Therefore Theorem II.3.1.d of [2] implies (for all $\delta = (\delta_1, \ldots, \delta_N) \in \{0, 1\}^N$)

$$\Gamma_{N\varrho\sigma}(\delta) = \frac{1}{N-1} \sum_{i=1}^{N} (\beta_{\varrho N}^*(i, \delta) - \bar{\beta}_{\varrho N}^*(\delta)) (\beta_{\sigma N}^*(i, \delta) - \bar{\beta}_{\sigma N}^*(\delta)) . \tag{2.26}$$

Thus $\Gamma_N(\Delta_N^{()})$ does not depend on (F^*, G^*), and from (2.23) we have

$$\Gamma_N(\Delta_N^{()}) \to \Gamma^* \quad \text{in} \quad (F^*, G^*)\text{-probability} . \tag{2.27}$$

Therefore the proof of Theorem 2.3 implies the following theorem:

Theorem 2.4. Given the assumptions and notations of Theorem 2.3 let $F_N^*(\delta, \cdot)$ be the distribution function of the conditional distribution of $LR(\Gamma_N(\Delta_N^{()}), S_N)$ given $\Delta_N^{()} = \delta$ under the hypothesis F^* and G^* and let $F_0^*(\Gamma^*, \cdot)$ be the distribution of $LR(\Gamma^*, X)$, where $X \sim \mathcal{N}(0, \Gamma^*)$.

Then the convergences

$$\begin{aligned} &\sup_{-\infty < t < \infty} |F_N^*(\Delta_N^{()}, t) - F_0^*(\Gamma^*, t)| \to 0 , \\ &\sup_{-\infty < t < \infty} |F_0^*(\Gamma_N(\Delta_N^{()}), t) - F_0^*(\Gamma^*, t)| \to 0 . \end{aligned} \tag{2.28}$$

hold true in (F^*, G^*)-probability.

As a consequence we may use $F_0^*(\Gamma_N(\delta), \cdot)$ as an approximation of the conditional distribution of $LR(\Gamma_N(\Delta_N^{()}), S_N)$ given $\Delta_N^{()} = \delta$ under the null hypothesis. The distribution $F_0^*(\Gamma_N(\delta), \cdot)$ is a mixture of χ^2-distributions, cf. [4], [6], [7]. If the dimension r is smaller than 4 there is a simple explicit representation, cf. Section 3.2 of [1].

3. THREE REPRESENTATIVE SCORE FUNCTIONS AND THE CORRESPONDING LIKELIHOOD RATIO TEST

In the present section we discuss a sensible choice of the score functions $(b_\varrho^{(1)}, b_\varrho^{(2)})$, $\varrho = 1, \ldots, r$, see (2.7), $\varrho = 1, \ldots, r$, see (2.7), by choosing suitable generalized shift functions D_ϱ, $\varrho = 1, \ldots, r$, as well as a suitable df F and a suitable censoring df G.

From the variety of all possible generalized shifts we choose three special types according to the following considerations:

If F is the distribution of the standard treatment and if the new treatment mainly shows a positive reaction for values of x where $F(x)$ is small, then we call this a *lower shift* situation. If the new treatment mainly shows a positive reaction for values of x where $F(x)$ is near $\frac{1}{2}$, then we call this a *central shift* situation. Similarly, *upper shift* situations are defined.

Assumpting $F(0) = \frac{1}{2}$ we formalize these notions by the following selections of the shift functions D_ϱ:

$$\begin{aligned} D_1 &= 1 - F && \text{(lower shift)}, \\ D_2 &= 4F(1 - F) && \text{(central shift)}, \\ D_3 &= F && \text{(upper shift)}. \end{aligned} \tag{3.1}$$

In the sequel let $\vec{b}_\varrho = (b_\varrho^{(1)}, b_\varrho^{(2)})$, $\varrho = 1, 2, 3$, be defined by formula (2.7) with D_ϱ according to formula (3.1).

Notice, the cone $V = \{\sum_{\varrho=1}^{3} \vartheta_\varrho \vec{b}_\varrho : \vartheta \geqq 0\}$ spanned by $\vec{b}_1, \vec{b}_2, \vec{b}_3$, contains the constant-shift optimal score function

$$(b^{(1)}, b^{(2)}) = \left(-\left(\frac{f'}{f}\right) \circ H^{-1}, \left(\frac{f}{1 - F}\right) \circ H^{-1}\right),$$

c.f. formula (3.13) of [5].

Because of simplicity and since the corresponding cone V contains suitable approximations of the optimal score functions for other df's F (e.g. the normal df or the Cauchy df) we restrict the discussion to the case of the logistic df F, i.e. we propose the asymptotic likelihood ratio test for the cone V which corresponds to the logistic df $F(x) = \exp(x)/(1 + \exp(x))$, $x \in \mathbb{R}$.

In the first step let us consider the functions (2.7) with H^{-1} substituted by F^{-1}, i.e. for $\varrho = 1, 2, 3$ we define

$$b_{1\varrho} := \left(-\frac{f'}{f} \circ F^{-1}\right)(D_\varrho \circ F^{-1}) - (d_\varrho \circ F^{-1}),$$

$$b_{2\varrho} := \left(\frac{f}{1 - F} \circ F^{-1}\right)(D_\varrho \circ F^{-1}).$$

By elementary evaluation we get the representations (for $0 < u < 1$)

$$\begin{aligned} b_{11}(u) &= (1 - u)(3u - 1), \\ b_{12}(u) &= 8u(1 - u)(2u - 1), \\ b_{13}(u) &= u(3u - 2) = -b_{11}(1 - u), \end{aligned} \tag{3.2.a}$$

and

$$\begin{aligned} b_{21}(u) &= u(1 - u), \\ b_{22}(u) &= 4u^2(1 - u), \\ b_{23}(u) &= u^2. \end{aligned} \tag{3.2.b}$$

Finally, for the definition of our three representative score functions $\vec{b}_\varrho = (b_\varrho^{(1)}, b_\varrho^{(2)})$, $\varrho = 1, 2, 3$, we need a censoring df G. In order to have some additional flexibility we use the

proportional hazard assumption:

$$(1 - G) = (1 - F)^\lambda \quad \text{for some} \quad \lambda > 0. \tag{3.3}$$

This assumption is equivalent to the assumption that the rv's $(X_1, \dots. X_N)$ and $(\Delta_1, \dots, \Delta_N)$ are independent under $F_1 = F_2 = F$ and G. Obviously assumption (3.3) and $H = 1 - (1 - F)(1 - G)$ imply (for $0 < u < 1$)

$$F_p(u) := F \circ H^{-1}(u) = 1 - (1 - u)^p, \quad \text{with} \quad p = 1/(1 + \lambda),$$

and the $\vec{b}_\varrho = (b_\varrho^{(1)}, b_\varrho^{(2)})$, $\varrho = 1, 2, 3$, according to (2.7) are evaluated as

$$b_\varrho^{(1)} = b_{1\varrho} \circ F_p, \quad b_\varrho^{(2)} = b_{2\varrho} \circ F_p. \tag{3.4}$$

Choosing approximate scores

$$b_{\varrho N i}^{(j)} = b_\varrho^{(j)}\left(\frac{i}{N+1}\right), \quad i = 1, \dots, N, \quad j = 1, 2, \quad \varrho = 1, 2, 3,$$

we get $S_N = (S_{1N}, S_{2N}, S_{3N})^{\mathrm{T}}$ with $(\forall \varrho = 1, 2, 3)$

$$S_{\varrho N} = \sum_{i=1}^{N} c_{ND_i}\left(\Delta^{(i)} b_\varrho^{(1)}\left(\frac{i}{N+1}\right) + (1 - \Delta^{(i)})\, b_\varrho^{(2)}\left(\frac{i}{N+1}\right)\right). \tag{3.5}$$

The corresponding 3×3-matrix $\Gamma = (\langle \vec{b}_i, \vec{b}_j \rangle_0) =: (\gamma_{ij})$ has the following explicit evaluation:

$$\begin{aligned}
\gamma_{11}(\lambda) &= \frac{4}{5+\lambda} - \frac{4}{4+\lambda} + \frac{1}{3+\lambda} \\
\gamma_{22}(\lambda) &= \frac{144}{7+\lambda} - \frac{384}{6+\lambda} + \frac{352}{5+\lambda} - \frac{128}{4+\lambda} + \frac{16}{3+\lambda} \\
\gamma_{33}(\lambda) &= \frac{4}{5+\lambda} - \frac{8}{4+\lambda} + \frac{4}{3+\lambda} \\
\gamma_{12}(\lambda) &= -\frac{24}{6+\lambda} + \frac{44}{5+\lambda} - \frac{24}{4+\lambda} + \frac{4}{3+\lambda} \\
\gamma_{13}(\lambda) &= -\frac{4}{5+\lambda} + \frac{6}{4+\lambda} - \frac{2}{3+\lambda} \\
\gamma_{23}(\lambda) &= \frac{24}{6+\lambda} - \frac{56}{5+\lambda} + \frac{40}{4+\lambda} - \frac{8}{3+\lambda}.
\end{aligned} \tag{3.6}$$

The determinant of $\Gamma = \Gamma(\lambda)$ is strictly positive for all $\lambda > 0$ which can be seen by drawing the graph of the function $\lambda \to \det \Gamma(\lambda)$. An exact proof seems to be very tedious.

Under the assumption (3.3) the regression function p_1, see (2.8), equals the constant p, i.e., $p_1 = p = 1/(1 + \lambda)$. Since

$$\bar{\Delta}_N = \frac{1}{N}\sum_{i=1}^{N} \Delta_N^{(i)} = \frac{1}{N}\sum_{i=1}^{N} \Delta_i$$

is a consistent estimator of p, we replace the parameter p in the definition (3.4) of the score functions by $\bar{\Delta}_N$. Similarly we replace $\lambda = (1 - p)/p$ in (3.6) by $\hat{\lambda}_N = (1 - \bar{\Delta}_N)/\bar{\Delta}_N$.

Now we prove that the assertions of Theorem 2.3 and Theorem 2.4 remain true, if we substitute the original score functions (3.4) by the score functions $b_{1\varrho} \circ F_{\bar{\Delta}_N}$ and $b_{2\varrho} \circ F_{\bar{\Delta}_N}$.

According to the proofs of Theorem 2.3 and Theorem 2.4 it's sufficient to prove (2.23) and (2.24) for the modified quantities $\beta^*_{\varrho N}(i, \Delta_N^{()}(\omega))$, and to use the joint continuity of the function of $(\Gamma, x) \to LR(\Gamma, x)$.

According to the special form of the $b_\varrho^{(j)}$'s in (3.4) we get (2.23) if we can prove

$$\frac{1}{N} \sum_{i=1}^{N} \left(\left(1 - \frac{i}{N+1} \right)^{kp} - \left(1 - \frac{i}{N+1} \right)^{k\bar{\Delta}_N} \right) \Delta_N^{(i)} \to 0 \quad \text{in probab.} \tag{3.7}$$

for all $k > 0$, and $p = \int (1 - G)\, dH_1 > 0$.

Since $|(1 - u)^{kp'} - (1 - u)^{kp}| \leqq k|(1 - u)^{k\tilde{p}} \ln (1 - u)|\, |p' - p|$ with $\tilde{p} \in (p'p)$ and $0 < u < 1$, the L.H.S. in (3.7) is of the order $0_P(|\bar{\Delta}_N - p|\, \bar{\Delta}_N) = o_P(1)$, hence (3.7) is proved.

Finally (2.24) holds true since all the $b_\varrho^{(j)}$'s in (3.4) are bounded.

Remark 3.3. In case of right censoring more often a scale model with positive observations is used instead of the translation model of Section 1, i.e. it's assumed that $X_{ji} \geqq 0$, $j = 1, 2$, $i = 1, \ldots, N$ with $F(0) = G(0) = 0$ and that X_{1i} has the df $F(x \exp(-c_{Ni}\vartheta^{\mathrm{T}} D(x)))$. In this case $F_1 \leqq F_2$ is equivalent to $\vartheta^{\mathrm{T}} D \geqq 0$. But, by switching over to the transformed rv's $\log X_{ji}$ the model becomes a translation model of the form of Section 1 with $F(x)$ replaced by $F(e^x)$ and $D(x)$ replaced by $D(e^x)$. Since the hypothesis of randomness and the stochastically larger alternative remain invariant under the log-transformation, we may use the above test also for the scale model with positive observations.

4. APPLICATION AND MONTE CARLO SIMULATION

In this section we discuss the applicability of the proposed likelihood ratio rank test. Especially we discuss the approximation of the conditional distribution of $LR(\Gamma_N(\Delta_N^{()}), S_N)$ given $\Delta_N^{()} = \delta$ under the null hypothesis by $F_0^*(\Gamma_N(\delta), \cdot)$, cf. Theorem 2.4.

The practical application of the approximate likelihood ratio rank test (LRRT) is rather simple:

We use the statistic $LR(\Gamma_N(\Delta_N^{()}), S_N)$, where $S_N = (S_{1N}, S_{2N}, S_{3N})^{\mathrm{T}}$ and $\Gamma_N(\delta)$ are defined in (3.5) and (2.26), respectively, on the basis of the score functions $b_{1\varrho} \circ F_{\bar{\Delta}_N}$ and $b_{2\varrho} \circ F_{\bar{\Delta}_N}$, $\varrho = 1, 2, 3$, cf. formulae (3.2) to (3.4). The numerical evaluation of the statistic is quite simple, cf. (2.15), since $R = \{1, 2, 3\}$.

Additionally, if Γ is any positive definite 3×3-matrix then we have, cf. Section

3.2 of [1],

$$1 - F_0^*(\Gamma, t) = \sum_{k=1}^{3} w_k P\{\chi_k^2 > t\} \quad \forall t > 0\,, \tag{4.1}$$

with

$$\begin{aligned} w_1 &= (\arccos(\varrho_{12}^*) + \arccos(\varrho_{13}^*) + \arccos(\varrho_{23}^*))/(4\pi)\,, \\ w_2 &= (\arccos(\varrho_{12}) + \arccos(\varrho_{13}) + \arccos(\varrho_{23}))/(4\pi)\,, \\ w_3 &= \tfrac{1}{2} - w_1\,, \end{aligned} \tag{4.2}$$

where $(\varrho_{ij})_{i,j=1,2,3}$ is the correlation matrix corresponding to Γ and where $(\varrho_{ij}^*)_{i,j=1,2,3}$ is the correlation matrix corresponding to Γ^{-1}.

Therefore, if we observe $t_0 = LR(\Gamma_N(\Delta_N^{(\,)}), S_N)$, it's easy to evaluate the approximate conditional p-value $1 - F_0^*(\Gamma_N(\Delta_N^{(\,)}), t_0)$ of the likelihood ratio rank statistic from (4.1) and (4.2).

We checked this approximation by Monte Carlo simulation in the following way:

Under the null hypothesis $F_1 = F_2 = F$ and the proportional hazard assumption $1 - G = (1 - F)^{(1-p)/p}$ for the censoring df G the distribution of the random vector $(R_1, \ldots, R_N, \Delta_N^{(1)}, \ldots, \Delta_N^{(N)})$ does not depend on F. With $F = F_0$ according to the uniform distribution on $(0, 1)$ and 3000 Monte Carlo repetitions we have listed the percentages of the events that the above approximate conditional p-value of the LRRT is smaller than 0·01, 0·05, 0·10, respectively. This has been done for $p = 0{\cdot}2$ to $p = 0{\cdot}9$ and some (n_1, n_2) between (10, 10) and (50, 50). Notice, the value of p is not known in the LRRT-procedure but estimated by $\bar{\Delta}_N$. (Under the proportional hazard assumption $1 - p$ is the probability of censoring under the null hypothesis, i.e. $p = \mathsf{P}_0\{\Delta_1 = 1\}$.)

Additionally we have checked the approximation under the null hypothesis $F_1 = F_2 = F_0$ with (non-proportional hazard) censoring df G according to

$$\begin{aligned} &[\mathrm{A}] \quad G(x) = x\,1(0 < x \leqq 0{\cdot}25) + (0{\cdot}25)\,1(0{\cdot}25 < x \leqq 1) \\ &\qquad\qquad + (x - 0{\cdot}75)\,1(1 < x \leqq 1{\cdot}75) + 1(1{\cdot}75 < x)\,, \\ &[\mathrm{B}] \quad G(x) = (x - 0{\cdot}25)\,1(0{\cdot}25 < x \leqq 0{\cdot}75) + (0{\cdot}5)\,1(0{\cdot}75 < x \leqq 1) \\ &\qquad\qquad + (x - 0{\cdot}5)\,1(1 < x \leqq 1{\cdot}5) + 1(1{\cdot}5 < x)\,, \\ &[\mathrm{C}] \quad G(x) = (2x - 1)\,1(0{\cdot}5 < x \leqq 1) + 1(1 < x)\,. \end{aligned} \tag{4.3}$$

The three LRRT-columns of Table 1 show that the approximation works well with some conservative tendency. Only for small sample sizes and very small p (80% censoring) the aproximation seems to be too conservative.

For comparison we show the corresponding simulation for the linear rank statistic $S_{1N} + S_{3N}$ with S_{1N} and S_{3N} as defined in (3.5) but with p substituted by $\bar{\Delta}_N$. Here the conditional null distribution is approximated by a $\mathcal{N}(0, \hat{\sigma}_N^2)$-distribution, where $\hat{\sigma}_N^2$ is the conditional variance of $S_{1N} + S_{3N}$ given $\Delta_N^{(\,)}$, cf. formulae (2.25) and (2.26). This statistic can be viewed as the correct adaptation of the usual Wilcoxon statistic to random censoring (Cens. Wilcoxon), cf. [5].

Table 1. Monte Carlo simulation under the null hypothesis $F_1 = F_2 = F$ with proportional hazard assumption $1 - G = (1 - F)^{(1-p)/p}$ for different values of p, and also under different types of non-proportional hazard assumptions. The Monte Carlo sample size is 3000.

(n_1, n_2)	Cens Df	LRRT			Cens. Wilcoxon			Wilcoxon		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
(10, 10)	$p = 0{\cdot}9$	0·5	4·8	9·9	0·8	5·1	10·3	1·0	5·2	10·8
	$p = 0{\cdot}8$	0·5	4·6	9·9	0·7	5·1	10·3	1·1	5·6	10·4
	$p = 0{\cdot}7$	0·5	4·7	10·1	0·7	5·1	10·5	1·1	5·4	11·2
	$p = 0{\cdot}6$	0·5	4·8	10·9	0·6	5·5	10·5	1·1	5·9	11·2
	$p = 0{\cdot}5$	0·6	4·7	9·8	0·7	5·3	10·5	0·9	5·6	11·0
	$p = 0{\cdot}4$	0·7	4·4	9·9	0·7	4·9	10·3	0·8	5·6	11·0
	$p = 0{\cdot}2$	0·1	3·8	9·7	0·3	4·9	11·0	0·8	5·2	11·7
	$G = [A]$	0·3	4·6	10·8	1·0	5·1	10·5	0·8	4·9	10·7
	$G = [B]$	0·8	4·4	9·7	1·0	4·8	9·4	0·6	4·7	10·1
	$G = [C]$	0·7	5·5	10·7	1·2	5·5	10·8	1·3	6·2	11·9
(10, 30)	$p = 0{\cdot}9$	0·8	4·9	10·5	0·9	5·1	10·0	0·8	4·8	10·2
	$p = 0{\cdot}8$	0·8	4·9	10·1	0·8	4·7	10·4	1·0	4·9	10·0
	$p = 0{\cdot}7$	0·8	4·7	9·9	0·8	4·7	10·1	0·9	5·1	10·2
	$p = 0{\cdot}6$	0·6	4·6	9·6	0·8	4·9	10·3	1·1	5·4	10·4
	$p = 0{\cdot}5$	0·7	5·0	10·1	0·8	5·4	10·8	1·5	5·4	10·2
	$p = 0{\cdot}4$	0·6	4·2	10·0	0·8	5·3	11·2	1·4	5·3	10·3
	$p = 0{\cdot}2$	0·2	2·6	7·7	0·2	4·2	9·8	1·2	5·4	10·5
	$G = [A]$	0·6	4·2	9·2	0·6	5·0	10·1	1·2	4·5	9·5
	$G = [B]$	0·5	4·2	9·1	0·7	4·1	9·5	0·8	4·6	8·9
	$G = [C]$	0·8	5·2	10·6	0·8	5·1	11·1	0·8	4·6	10·0
(30, 10)	$p = 0{\cdot}9$	1·3	5·0	9·8	1·0	5·5	10·1	1·1	5·2	10·6
	$p = 0{\cdot}8$	1·1	5·1	10·4	1·1	5·5	10·7	0·9	5·2	10·5
	$p = 0{\cdot}7$	1·2	5·1	10·1	1·0	5·6	10·2	1·0	5·6	10·3
	$p = 0{\cdot}6$	1·2	5·4	10·6	1·1	5·8	11·0	0·9	5·1	10·3
	$p = 0{\cdot}5$	1·2	5·6	10·8	1·2	5·6	10·8	1·1	4·9	9·6
	$p = 0{\cdot}4$	1·3	5·9	10·5	1·2	5·8	10·5	1·0	5·0	9·8
	$p = 0{\cdot}2$	1·4	6·4	11·0	1·5	5·8	10·3	0·8	5·1	9·8
	$G = [A]$	0·9	4·9	10·1	0·9	5·0	10·7	1·1	4·5	8·5
	$G = [B]$	1·0	4·9	10·5	0·9	4·9	10·3	0·8	4·8	9·7
	$G = [C]$	1·3	5·6	9·9	1·3	5·2	10·9	1·2	5·6	10·1
(30, 30)	$p = 0{\cdot}9$	1·0	4·9	10·8	1·3	5·3	10·7	1·3	5·8	11·0
	$p = 0{\cdot}8$	1·0	4·8	10·6	1·3	5·3	10·3	1·0	5·7	10·6
	$p = 0{\cdot}7$	0·8	5·1	10·5	1·1	5·3	11·1	1·1	5·4	10·4
	$p = 0{\cdot}6$	1·1	5·1	11·0	1·2	5·6	10·8	1·0	5·3	10·9
	$p = 0{\cdot}5$	0·9	4·9	10·7	1·1	5·4	11·1	1·0	5·0	10·6
	$p = 0{\cdot}4$	0·8	4·8	10·8	0·7	5·4	11·3	1·0	5·0	10·0
	$p = 0{\cdot}2$	0·8	5·3	10·3	0·9	5·1	11·0	1·2	5·4	10·8
	$G = [A]$	0·7	4·6	10·0	0·8	4·9	9·7	0·8	4·7	10·5
	$G = [B]$	0·7	4·9	10·2	0·8	4·8	10·1	0·7	4·7	9·9
	$G = [C]$	0·9	4·8	9·6	1·0	4·9	9·8	0·9	4·5	9·4

Table 1. (Continuation)

(n_1, n_2)	Cens Df	LRRT			Cens. Wilcoxon			Wilcoxon		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
(30, 50)	$p = 0\cdot9$	0·7	4·8	10·4	1·0	5·2	10·6	0·8	4·8	10·0
	$p = 0\cdot8$	0·7	4·8	10·1	0·8	5·1	10·4	0·5	4·7	9·7
	$p = 0\cdot7$	1·0	5·1	10·3	0·8	5·2	10·2	0·6	4·2	9·9
	$p = 0\cdot6$	0·9	5·2	10·0	0·8	5·0	10·1	0·5	4·6	9·7
	$p = 0\cdot5$	0·7	5·2	10·0	0·7	4·5	9·8	0·4	4·4	9·8
	$p = 0\cdot4$	0·5	4·9	9·5	0·6	4·5	9·8	0·5	4·3	0·6
	$p = 0\cdot2$	0·7	4·0	9·6	0·9	4·6	10·1	0·7	4·0	9·6
	$G = [A]$	1·1	4·9	10·3	1·1	5·4	10·2	1·1	5·2	10·0
	$G = [B]$	0·9	5·2	10·3	0·9	4·9	10·4	0·8	5·1	10·4
	$G = [C]$	1·0	5·2	10·8	0·8	5·4	10·2	0·8	5·1	10·6
(50, 30)	$p = 0\cdot9$	0·6	5·5	11·2	0·7	5·3	10·6	0·8	4·8	10·2
	$p = 0\cdot8$	0·6	5·2	11·0	0·7	5·2	10·2	0·6	4·9	9·9
	$p = 0\cdot7$	0·8	5·0	10·8	0·7	5·0	10·5	0·8	4·7	9·8
	$p = 0\cdot6$	0·8	5·1	10·3	0·8	5·1	10·2	0·8	4·6	10·0
	$p = 0\cdot5$	1·1	5·7	10·3	1·2	5·5	10·0	0·8	4·8	9·8
	$p = 0\cdot4$	1·0	5·1	9·9	1·0	5·4	10·6	1·3	5·0	9·7
	$p = 0\cdot2$	1·0	5·9	11·6	1·1	5·6	10·5	1·0	4·7	9·7
	$G = [A]$	0·8	4·5	10·2	1·0	5·1	10·4	0·9	5·0	10·5
	$G = [B]$	1·0	4·9	9·8	1·0	4·9	9·7	1·1	4·8	9·6
	$G = [C]$	0·6	4·4	9·3	0·8	4·5	9·1	0·9	4·6	9·5
(50, 50)	$p = 0\cdot9$	0·5	4·5	10·3	0·7	4·4	9·3	0·6	4·2	9·2
	$p = 0\cdot8$	0·6	4·3	9·7	0·7	4·7	9·5	0·8	4·5	9·5
	$p = 0\cdot7$	0·6	4·3	9·3	0·8	4·6	9·9	0·7	4·4	9·7
	$p = 0\cdot6$	0·7	4·4	9·0	0·8	5·0	9·3	0·8	4·6	9·6
	$p = 0\cdot5$	1·1	5·4	10·8	0·9	5·3	10·1	0·8	5·1	9·5
	$p = 0\cdot4$	0·8	4·5	9·7	0·8	4·8	9·8	0·9	4·8	10·1
	$p = 0\cdot2$	1·0	4·4	9·7	1·2	4·6	9·5	0·6	4·8	10·3
	$G = [A]$	0·5	4·9	9·9	0·5	4·4	10·0	0·7	4·3	9·6
	$G = [B]$	1·0	5·2	10·4	1·1	5·3	10·1	1·3	5·1	10·1
	$G = [C]$	1·2	5·9	10·6	1·2	5·7	10·1	1·1	5·4	10·0

Finally we report the simulation results for the usual Wilcoxon test based on the ranks $R_1, \ldots, R_N$ of $X_1, \ldots, X_N$ only. Here we used the well-known (unconditional) normal approximation (Wilcoxon).

In Table 2 and Table 3 we report the results of Monte Carlo power simulations of the three approximate tests of Table 1. The results are given in the percent scale.

In Table 2 we considered generalized shift alternatives of the form

$$\mathscr{L}(X_{1i}) \sim F(x - c_{Ni} s\, D_\varrho(x))$$

with logistic df F and D_ϱ, $\varrho = 1, 2, 3$, according to (3.1) $[D_4 = 1]$ under the proportional hazard assumption $1 - G = (1 - F)^{(1-p)/p}$. Additionally we considered the

Table 2. Monte Carlo power simulation (in percent scale) at the approximate 5% level under the proportional hazard assumption $1 - G = (1 - F)^{(1-p)/p}$ for different values of p. The Monte Carlo sample size is 3000.
[1] Lower-, [2] Central-, [3] Upper-Logistic Shift with $s = 5{\cdot}0$, [4] Exact Logistic Shift with $s = 5{\cdot}0$, [5] Generalized Cauchy Shift with $s = 5{\cdot}0$

(n_1, n_2)	Cens Df	LRRT [1]	[2]	[3]	[4]	[5]	Cens. Wilcoxon [1]	[2]	[3]	[4]	[5]	Wilcoxon [1]	[2]	[3]	[4]	[5]
(10, 10)	$p = 0{\cdot}8$	37·5	59·2	35·1	76·3	51·2	38·3	57·9	33·4	81·2	43·3	35·7	50·4	28·6	72·8	37·2
	$p = 0{\cdot}6$	35·8	55·8	30·1	71·9	49·5	37·9	56·0	28·8	76·7	43·2	30·7	38·0	19·3	57·5	29·0
	$p = 0{\cdot}4$	33·4	48·9	22·3	63·0	42·4	39·0	48·2	21·4	68·9	37·2	25·1	24·4	12·7	39·1	17·5
	$p = 0{\cdot}2$	26·9	31·6	10·8	44·9	23·1	33·9	32·4	12·5	52·1	22·0	15·9	12·1	7·2	19·4	8·0
(10, 30)	$p = 0{\cdot}8$	43·4	63·5	34·3	78·8	58·0	45·3	60·7	31·4	82·9	48·6	40·9	50·6	35·2	72·0	40·6
	$p = 0{\cdot}6$	40·2	54·9	26·7	70·3	51·2	44·7	52·8	25·2	77·0	43·1	34·4	35·4	16·6	52·2	28·4
	$p = 0{\cdot}4$	34·4	42·3	17·2	57·3	39·2	42·4	40·7	17·3	64·4	33·3	24·5	20·8	9·2	31·3	16·7
	$p = 0{\cdot}2$	19·3	17·7	6·3	29·1	13·5	28·4	21·3	7·9	40·0	14·8	13·8	9·7	6·1	14·6	8·1
(30, 10)	$p = 0{\cdot}8$	40·6	67·3	41·8	82·0	59·7	38·8	66·0	39·1	85·8	49·9	35·5	58·9	32·1	79·9	43·8
	$p = 0{\cdot}6$	40·9	64·9	36·8	78·9	57·6	39·6	64·4	32·7	82·5	50·1	32·2	50·2	19·8	68·4	36·2
	$p = 0{\cdot}4$	40·3	61·0	29·9	74·4	52·0	42·4	61·0	26·3	77·6	46·3	27·5	35·6	11·8	50·1	22·2
	$p = 0{\cdot}2$	40·7	51·5	19·8	66·2	38·2	43·3	49·3	16·5	67·7	30·5	18·6	15·6	6·4	26·8	9·0
(30, 30)	$p = 0{\cdot}8$	42·3	70·7	39·8	82·0	68·7	41·3	67·1	35·7	86·6	56·0	37·7	58·9	29·6	78·5	47·0
	$p = 0{\cdot}6$	41·3	65·3	33·0	77·4	64·6	41·4	62·9	29·9	82·5	52·7	31·9	44·8	19·5	62·7	35·6
	$p = 0{\cdot}4$	38·4	56·8	24·2	71·2	54·5	41·0	54·9	22·3	75·5	44·0	25·1	26·8	11·9	40·0	20·0
	$p = 0{\cdot}2$	34·3	36·3	12·8	55·1	30·4	38·5	34·6	12·7	59·2	22·8	15·5	12·2	7·0	18·5	7·8
(30, 50)	$p = 0{\cdot}8$	43·6	71·8	38·8	80·5	71·3	42·0	67·5	33·8	84·4	56·8	38·1	57·3	26·8	75·7	47·0
	$p = 0{\cdot}6$	43·0	65·0	30·3	76·4	65·3	43·2	61·7	26·3	81·5	51·9	33·4	41·4	17·0	58·6	34·1
	$p = 0{\cdot}4$	39·1	53·6	22·4	67·2	53·6	42·7	50·9	19·4	71·3	41·0	26·3	24·8	10·2	37·4	18·7
	$p = 0{\cdot}2$	32·2	21·1	12·0	49·0	26·2	37·9	28·0	11·6	54·1	21·2	15·4	11·1	6·0	17·5	7·5
(50, 30)	$p = 0{\cdot}8$	44·4	72·6	40·5	83·1	73·3	41·3	68·7	38·0	86·5	57·3	37·8	60·5	30·6	79·4	49·6
	$p = 0{\cdot}6$	43·0	67·7	33·9	79·0	69·8	42·7	66·0	31·4	82·3	56·5	34·0	47·0	20·1	64·7	38·1
	$p = 0{\cdot}4$	42·7	59·6	25·6	73·9	59·8	43·5	58·1	23·3	77·7	48·2	26·7	27·9	11·8	44·7	21·5
	$p = 0{\cdot}2$	38·1	41·5	15·6	60·1	37·1	41·2	38·1	14·2	63·8	27·5	16·5	11·6	7·4	21·8	7·9
(50, 50)	$p = 0{\cdot}8$	44·4	71·3	40·6	82·1	76·0	42·6	68·2	36·3	85·7	59·0	38·7	59·6	29·0	76·9	50·7
	$p = 0{\cdot}6$	43·3	66·9	33·2	77·5	69·7	44·0	63·9	28·7	81·5	56·0	33·5	45·3	17·7	61·0	37·2
	$p = 0{\cdot}4$	40·7	57·9	24·3	70·4	59·6	43·0	54·6	20·9	73·6	47·0	26·0	28·0	10·9	40·7	21·7
	$p = 0{\cdot}2$	34·5	38·0	12·4	53·5	33·0	38·3	34·2	11·6	57·0	23·2	15·7	11·9	6·8	19·4	8·8

Table 3. Monte Carlo power simulation (in percent scale) at the approximate 1%, 5%, and 10% levels under the different types of non-proportional hazard assumptions (4.3) and $\mathscr{L}(X_{1i}) = \mathscr{L}(U + 500_{c_{Ni}}U^4(1-U)^4)$, $\mathscr{L}(U)$ is the uniform distribution on (0, 1). The Monte Carlo sample size is 3000.

(n_1, n_2)	Cens Df	LRRT			Cens. Wilcoxon			Wilcoxon		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
(10, 10)	$G = [A]$	28·3	64·5	78·3	32·8	64·6	78·3	23·2	51·9	68·6
	$G = [B]$	32·4	64·3	78·1	36·6	67·7	79·2	30·8	61·9	76·0
	$G = [C]$	34·3	64·6	78·7	39·9	68·4	80·9	35·3	64·1	78·1
(10, 30)	$G = [A]$	46·4	71·9	82·5	32·1	56·8	68·7	18·4	40·5	51·8
	$G = [B]$	44·5	71·8	82·3	34·5	62·3	75·2	25·3	51·7	64·7
	$G = [C]$	48·9	75·2	84·9	43·2	69·8	80·9	31·4	58·2	71·1
(30, 10)	$G = [A]$	64·7	84·2	90·6	57·1	78·5	86·1	41·6	67·2	77·6
	$G = [B]$	61·2	82·0	89·2	55·6	78·8	87·1	52·0	77·1	85·6
	$G = [C]$	59·0	81·4	88·8	48·5	73·7	84·0	53·7	78·4	86·3
(30, 30)	$G = [A]$	47·5	74·1	84·3	30·9	58·7	72·2	18·7	42·2	57·5
	$G = [B]$	63·0	84·9	91·7	54·9	79·1	86·9	43·0	71·6	82·1
	$G = [C]$	64·0	86·1	92·9	56·7	82·2	90·4	49·5	76·4	85·9
(30, 50)	$G = [A]$	46·4	70·9	82·0	28·9	55·0	68·3	18·6	41·3	55·4
	$G = [B]$	65·3	86·0	92·6	52·5	77·8	87·0	40·6	66·9	78·8
	$G = [C]$	69·7	89·1	94·7	60·3	82·7	91·0	50·2	74·9	84·2
(50, 30)	$G = [A]$	59·1	82·3	89·7	38·9	65·5	78·1	22·7	51·4	65·8
	$G = [B]$	70·7	88·9	94·5	62·7	83·8	91·4	53·6	78·1	86·7
	$G = [C]$	65·1	86·7	93·0	56·6	81·0	89·6	54·9	79·0	88·3
(50, 50)	$G = [A]$	49·8	75·0	84·5	30·7	57·3	70·5	20·1	43·8	57·2
	$G = [B]$	74·7	90·9	94·9	61·3	83·9	91·4	49·1	75·3	85·8
	$G = [C]$	74·1	9†·4	95·6	63·5	85·2	91·5	56·2	79·9	88·9

corresponding generalized Cauchy shift (F = Cauchy df) with $D_\varrho = 16F^2(1-F)^2$.

In Table 3 we considered generalized shift alternatives of the form

$$\mathscr{L}(X_{1i}) = \mathscr{L}(U + 500c_{Ni}U^4(1-U)^4),$$

where $\mathscr{L}(U)$ is the uniform distribution on (0,1). The (non-proportional hazard) censoring dfs G have been selected according to formula (4.3).

For small sample sizes the Cens. Wilcoxon seems to be a good choice, whereas for sample sizes $(n_1, n_2) = (30, 30)$ and higher the LRRT seems to be comparable to the Cens. Wilcoxon for the generalized logistic shift alternatives and substantially better than the Cens. Wilcoxon in the Cauchy situation and in the model situation of Table 3. Therefore we propose the LRRT for sample sizes $n_1 \geqq 30$ and $n_2 \geqq 30$. Obviously the usual Wilcoxon (based on the ranks $R_1, \ldots, R_N$ only) is a bad choice.

(Received December 8, 1989.)

REFERENCES

[1] K. Behnen and G. Neuhaus: Rank Tests with Estimated Scores and Their Application. Teubner, Stuttgart 1989.

[2] J. Hájek and Z. Šidák: Theory of Rank Tests. Academic Press, New York 1967.

[3] A. Janssen: Local asymptotic normality for randomly censored models with applications to rank tests. Preprint 199, Univ. Siegen 1987.

[4] A. Kudo: A multivariate analog of the one-sided test. Biometrika *50* (1963), 403—418.

[5] G. Neuhaus: Asymptotically optimal rank tests for the two-sample problem with randomly censored data. Commun. Statist. — Theory Meth. *17* (1988), 6, 2037—2058.

[6] P. E. Nüesch: On the problem of testing location in multivariate populations for restricted alternatives. Ann. Math. Statist. *37* (1966), 113—119.

[7] P. E. Nüesch: Estimation of monotone parameters and the Kuhn-Tucker conditions. Ann. Math. Statist. *41* (1970), 1800.

[8] H. Strasser: Mathematical Theory of Statistics. de Gruyter, Berlin 1985.

Prof. Dr. Konrad Behnen, Prof. Dr. Georg Neuhaus, Institut für Mathematische Stochastik, Bundesstrasse 55, D-2000 Hamburg. Federal Republic of Germany.