Biometrika (1982), **69**, 3, *pp*. 553–66
Printed in Great Britain

A class of rank test procedures for censored survival data

BY DAVID P. HARRINGTON

Department of Applied Mathematics and Computer Science, University of Virginia, Charlottesville, Virginia, U.S.A.

AND THOMAS R. FLEMING

Department of Medical Statistics and Epidemiology, Mayo Clinic, Rochester, Minnesota, U.S.A.

SUMMARY

A class of linear rank statistics is proposed for the k-sample problem with right-censored survival data. The class contains as special cases the log rank test (Mantel, 1966; Cox, 1972) and a test essentially equivalent to Peto & Peto's (1972) generalization of the Wilcoxon test. Martingale theory is used to establish asymptotic normality of test statistics under the null hypotheses considered, and to derive expressions for asymptotic relative efficiencies under contiguous sequences of alternative hypotheses. A class of distributions is presented which corresponds to the class of rank statistics in the sense that for each distribution there is a statistic with some optimal properties for detecting location alternatives from that distribution. Some Monte Carlo results are displayed which present small sample behaviour.

Some key words: Censored data; Contiguous alternative; Linear rank statistic; Log rank test; Pitman asymptotic relative efficiency; Wilcoxon test.

1. INTRODUCTION

Some attention has been given in the recent statistical literature to various formulations for nonparametric statistics used in survival theory. Prentice (1978) has shown that statistics such as the log rank (Mantel, 1966; Cox, 1972) and Peto & Peto's (1972) generalization of the Wilcoxon statistic can be obtained as linear rank statistics for censored data assumed to come from a log linear model. Prentice & Marek (1979) have shown that in many cases such linear rank statistics may be expressed as vectors whose components consist of sums of weighted differences between the observed deaths in a given sample and the conditionally expected deaths, given the censoring and survival pattern up to the time of an observed death. Aalen (1978) has shown that most of the two-sample tests can be formulated as stochastic integrals with respect to counting processes, and has used martingale theory (Aalen, 1977) to obtain weak convergence results for some estimators and test statistics. Gill (1980) has used this martingale formulation to obtain Pitman asymptotic relative efficiencies for statistics such as the log rank and both Gehan's (1965) and Peto & Peto's (1972) generalizations of Wilcoxon's statistic under contiguous sequences of two-sample location alternatives. In this paper, we propose a class of k-sample nonparametric procedures which includes the log rank and Peto & Peto's statistics as special cases, and we show how the formulations mentioned above can be used to study the properties of these procedures. Other nonparametric k-sample procedures have been proposed by Breslow (1970) and Koziol & Reid (1977).

The most intuitive formulation for these new statistics is in the context outlined by Prentice & Marek (1979). The necessary notation is as follows. Let $\{X_{ij};\ 1 \leqslant j \leqslant N_i, 1 \leqslant i \leqslant r+1\}$ denote the survival variables, i.e. death times, from $r+1$ independent random samples with survival functions $S_i(t) = \mathrm{pr}(X_{ij} \geqslant t)$ $(i = 1, \ldots, r+1)$. We assume throughout that the S_i are absolutely continuous with cumulative hazard and hazard functions defined, respectively, as $\beta_i(t) = -\log S_i(t)$ and $\lambda_i(t) = d\beta_i(t)/dt$. Let $\{Y_{ij};\ 1 \leqslant j \leqslant N_i, 1 \leqslant i \leqslant r+1\}$ denote censoring variables with censoring distributions $C_i(t) = \mathrm{pr}(Y_{ij} \geqslant t)$. We assume that the survival and censoring times are independent and thus

$$\pi_i(t) \equiv \mathrm{pr}(X_{ij} \geqslant t, Y_{ij} \geqslant t) = S_i(t)\, C_i(t).$$

We assume that the data consist of

$$\{X_{ij}^0 = \min(X_{ij}, Y_{ij}), \delta_{ij} = I(X_{ij} \leqslant Y_{ij});\ 1 \leqslant j \leqslant N_i, 1 \leqslant i \leqslant r+1\},$$

where $I(A) = 1$ if the event A has occurred and 0 otherwise, but that the important inference problems are those regarding the S_i. Let $T_1 < \ldots < T_d$ denote the d distinct ordered observed death times in the pooled sample. For $i = 1, \ldots, r$, define $z_{ik} = 1$ if and only if the observed death at T_k is from sample i, and let $\tilde{z}'_k = (z_{1k}, \ldots, z_{rk})$, where $\tilde{x}'$ denotes the transpose of any column vector $\tilde{x}$; $\tilde{z}_k = \tilde{0}$ will mean that the death at T_k is in sample $r+1$. Let n_{ik} be the number of subjects under observation in sample i just prior to T_k; $n_{1k} + \ldots + n_{r+1,k} = n_k$ will be the total size of the risk set in the pooled sample at T_k. If we let $\{w_k;\ 1 \leqslant k \leqslant d\}$ be a set of weights, then Prentice & Marek (1979) and Tarone & Ware (1977) have shown that many nonparametric tests for H_0: $S_1 = \ldots = S_{r+1}$ are based on statistics of the form $\Sigma\, w_k(\tilde{z}_k - \tilde{q}_k)$, where $\tilde{q}'_k = (n_k)^{-1}(n_{1k}, \ldots, n_{rk})$. When $w_k \equiv 1$, the resulting test is the log rank while, when

$$w_k = S^*(T_k) \equiv \prod_{j=1}^{k} n_j/(n_j + 1),$$

the test is Prentice's (1978) generalization of the Wilcoxon test and is similar to the generalization of Peto & Peto (1972).

Clearly, the way to choose proper weights is to pick a set yielding a procedure as sensitive as possible to the types of departures from equality of the S_i that are anticipated in a given experiment. Gill (1980) has shown that in two samples of censored data, the log rank test has unit Pitman asymptotic relative efficiency against a time-transformed sequence of contiguous location alternatives when the S_i are type I extreme value distributions. The log rank test is thus fully efficient for proportional hazards alternatives. The approach illustrated by Prentice (1978) proves the log rank test is the locally most powerful rank test and is fully efficient against time-transformed location alternatives for the extreme value distribution when there are $r+1$ samples of uncensored data. Of course the log rank test reduces in this setting to Savage's exponential scores test, and its properties have been known for some time. Gill (1980) has also shown that Peto & Peto's Wilcoxon test is fully efficient against time-transformed location alternatives for the logistic distribution in two samples of censored data, while Prentice (1978) observed that it is a locally most powerful rank test in $r+1$ samples of uncensored data.

A natural class of weights which generalizes the log rank and Peto & Peto Wilcoxon weights is $w_k(\rho) = \{\hat{S}(T_k)\}^\rho$, for a fixed $\rho \geqslant 0$ where $\hat{S}(T_k)$ is a pooled survival function estimator at T_k. In this paper we study the properties of tests based on the statistics

$$G^\rho = \sum_{k=1}^{d} \{\hat{S}(T_k)\}^\rho (\tilde{z}_k - \tilde{q}_k),$$

where

$$\hat{S}(u) = \prod_{T_k, T_k < u} (n_k - 1)/n_k$$

differs only slightly from $S^*(u)$. Section 2 gives results for this statistic when there are two samples of arbitrarily right censored data. Those results will include asymptotic distribution theory under null and contiguous sequences of alternative hypotheses and a characterization of the parametric location alternatives for which these procedures are fully efficient or locally most powerful rank tests. Section 3 contains results that can be proved when the number of samples differs from 2. In §4 results of Monte Carlo simulations are displayed which support, for two-sample tests, the claim that the asymptotic theory provides reasonably accurate approximations for the properties of the statistics in small and moderate sample sizes.

2. *The two-sample statistics*

The most direct way to obtain the asymptotic distribution theory for the two-sample G^ρ statistics relies on martingale weak convergence theorems. To apply these results we need to recast our two-sample statistic in slightly different notation. Assume $r+1 = 2$; for $i = 1, 2$, let

$$N_i(t) = \sum_{j=1}^{N_i} I(X_{ij}^0 \leqslant t, \delta_{ij} = 1), \quad Y_i(t) = \sum_{j=1}^{N_i} I(X_{ij}^0 \geqslant t).$$

In the earlier notation, $Y_i(T_k) = n_{ik}$.

In this setting G^ρ may be expressed as

$$G^\rho_{N_1,N_2} = \int_0^\infty \{\hat{S}(u)\}^\rho \left\{\frac{Y_1(u)\,Y_2(u)}{Y_1(u)+Y_2(u)}\right\} \left\{\frac{dN_1(u)}{Y_1(u)} - \frac{dN_2(u)}{Y_2(u)}\right\},$$

with the convention that $0/0 = 0$. Note that the class of statistics $G^\rho_{N_1,N_2}$ $(\rho \geqslant 0)$ is a subset of the more general class K^+ of Gill (1980). The asymptotic distribution theory derived by Gill for that class applies here and is summarized in Theorems 2·1 and 2·2 below. Note that Theorem 2·1 may be proved more directly following Fleming & Harrington (1981, §6).

THEOREM 2·1 *Let $\Delta N_i(u) = N_i(u) - \lim N_i(t)$ as $t \uparrow u$, and let the variance estimator be given by*

$$V = \sum_{i=1}^{2} \int_0^\infty \frac{\{\hat{S}(u)\}^{2\rho}}{Y_i(u)} \left\{\frac{Y_1(u)\,Y_2(u)}{Y_1(u)+Y_2(u)}\right\}^2 \left\{1 - \frac{\Delta N_1(u) + \Delta N_2(u) - 1}{Y_1(u) + Y_2(u) - 1}\right\} \frac{d\{N_1(u)+N_2(u)\}}{Y_1(u)+Y_2(u)}.$$

Let $N = N_1 + N_2$ and assume $\lim N_i/N \equiv a_i$, as $N \to \infty$, exists and satisfies $0 < a_i < 1$. Then under the null hypothesis H_0: $S_1 = S_2 = S$:

$$\text{(a)} \quad \lim_{N\to\infty} \left(\frac{N_1+N_2}{N_1 N_2}\right) V = \int_0^\infty \frac{\pi_1(u)\,\pi_2(u)}{a_1\pi_1(u) + a_2\pi_2(u)} \{S(u)\}^{2\rho}\, d\beta(u)$$

in probability, where $\beta(u) = -\log S(u)$.

$$\text{(b)} \quad \lim_{N\to\infty} \{V\}^{-\frac{1}{2}} G^{\rho}_{N_1,N_2} = Z \sim N(0,1)$$

in distribution.

Proof. For $i = 1, 2$, $Y_i(t)/N_i$ is the empirical distribution function estimator of $\pi_i(t)$. Thus by the Glivenko–Cantelli theorem,

$$\sup_{0\leqslant t<\infty} |\, Y_i(t)/N_i - \pi_i(t)\,| \to 0$$

in probability as $N \to \infty$. The proof then follows directly from Proposition 4.3.3 of Gill (1980).

The variance estimator V is identical to the hypergeometric variance estimator from a series of independent 2×2 contingency tables, one at each death time (Mantel, 1966). In the original notation for the test statistic,

$$V = \sum_{k=1}^{d} w_k^2 (n_{1k}/n_k)\,\{(n_k - n_{1k})/n_k\}\,\{(n_k - d_k)/(n_k - 1)\}\, d_k,$$

where d_k is the number of observed deaths or failures at time T_k. Under the model assumed in this paper, ties at observed death times can only be caused by a grouping of the data. In the case of ties,

$$G^{\rho} = \sum_{k=1}^{d} \{\hat{S}(T_k)\}^{\rho} (d_{1k} - d_k n_{1k}/n_k),$$

where d_{1k} is the number of observed deaths from sample 1 at T_k.

Theorem 2·1 provides the means for approximating the significance level of observed values of $G^{\rho}_{N_1,N_2}$. Some information is also available about the asymptotic behaviour of these statistics under alternative hypotheses. Specifically, $G^{\rho}_{N_1,N_2}$ is consistent against the alternatives H_1: $\lambda_1(t) \geqslant \lambda_2(t)$ for $t \in \{u\colon S_1(u)\, S_2(u) > 0\}$, and H_2: $\beta_1(t) \geqslant \beta_2(t)$ $(0 \leqslant t < \infty)$ with each of the inequalities being strict inequalities on some interval. Details are given by Gill (1980, §4.1). The following theorem provides a basis for computing Pitman asymptotic relative efficiencies under a contiguous sequence of alternative hypotheses, and follows directly from Theorem 4.2.1 of Gill (1980).

THEOREM 2·2. *Let* $S_i^N(t)$ $(i = 1, 2)$ *be a sequence of survival functions which satisfies* $\lim S_i^N(t) = S(t)$ *uniformly in* $t \in [0, \infty)$. *Let* β_i^N *be the associated cumulative hazard functions. Define*

$$\gamma_i(t) \equiv \lim_{N\to\infty} \left(\frac{N_1 N_2}{N_1 + N_2}\right)^{\frac{1}{2}} \left\{\frac{d\beta^N(t)}{d\beta} - 1\right\} \quad (i = 1, 2),$$

and assume that the convergence is uniform on each closed subinterval of $\{t\colon S(t) > 0\}$. *Let* $\gamma(t) = \gamma_1(t) - \gamma_2(t)$ *and* $\pi_i(t) = S(t)\, C_i(t)$. *Then for* $0 < t \leqslant \infty$

$$\lim_{N\to\infty} \left(\frac{N_1 + N_2}{N_1 N_2}\right)^{\frac{1}{2}} G^{\rho}_{N_1,N_2} = W \sim N(\mu_\rho, \sigma_\rho^2)$$

in distribution, where

$$\mu_\rho = \int_0^\infty \frac{\pi_1(u)\,\pi_2(u)}{a_1\,\pi_1(u)+a_2\,\pi_2(u)}\gamma(u)\,\{S(u)\}^\rho\,d\beta(u),$$

$$\sigma_\rho^2 = \int_0^\infty \frac{\pi_1(u)\,\pi_2(u)}{a_1\,\pi_1(u)+a_2\,\pi_2(u)}\{S(u)\}^{2\rho}\,d\beta(u). \tag{2·1}$$

The asymptotic efficacy of the statistic $G^\rho_{N_1,N_2}$ under such a sequence of contiguous alternatives is defined in the usual way to be $e(\rho) = (\mu_\rho/\sigma_\rho)^2$. We now describe briefly how this expression for efficacy can be used to find a class of parametric location alternatives against which the statistic $G^\rho_{N_1,N_2}$ is asymptotically fully efficient.

Suppose the contiguous alternatives are indexed by location parameters θ_i^N $(i = 1,2)$ and let $1-S_i^N(t) = \Psi\{g(t)+\theta_i^N\}$, where Ψ is a fixed cumulative distribution function and $g(t)$ is an arbitrary monotonically increasing time transformation. Let $\psi(t) = d\Psi(t)/dt, \gamma(t) = \psi(t)\{1-\Psi(t)\}^{-1}$, $l(t) = \log\gamma(t), l'(t) = d\log\gamma(t)/dt$ and

$$\theta_i^N = \theta_0+(-1)^{i+1}\left\{\frac{N_{3-i}}{N_i(N_1+N_2)}\right\}^{\frac{1}{2}},$$

where θ_0 is unspecified. Assume that we are interested in testing the sequence of null hypotheses H_0^N: $\theta_1^N = \theta_2^N$ against the alternatives H_1^N: $\theta_1^N < \theta_2^N$ or $\bar{H}_1^N$: $\theta_1^N \neq \theta_2^N$, and that we wish to restrict ourselves to statistics of the form

$$\int_0^\infty K(u)\left\{\frac{dN_1(u)}{Y_1(u)}-\frac{dN_2(u)}{Y_2(u)}\right\},$$

where $\{K(u), 0 \leqslant u < \infty\}$ is a stochastic process satisfying the regularity conditions outlined in § 3·3 of Gill (1980). Then Gill has shown that the resulting asymptotic efficacy will be maximized for the contiguous location alternatives above if K is chosen as $K(u) = l'[\Psi^{-1}\{1-\hat{S}(u)\}]\,Y_1(u)\,Y_2(u)/\{Y_1(u)+Y_2(u)\}$, where $\hat{S}$ is the previously defined left-continuous version of the product-limit survival function estimator in the pooled sample. Thus the time-transformed location alternatives against which tests based on $G^\rho_{N_1,N_2}$ should have good sensitivity will include distributions Ψ which satisfy $l'[\Psi^{-1}\{1-\hat{S}(u)\}] = \{\hat{S}(u)\}^\rho$. A sufficient condition ensuring this is $l' = (1-\Psi)^\rho$. If we let $H(t) = 1-\Psi(t)$, a short calculation shows that the underlying survival functions $H(t)$ against which $G^\rho_{N_1,N_2}$ should have good power for time-transformed location alternatives include, but are not necessarily limited to, survival functions $H(t)$ which satisfy the differential equation:

$$\frac{H''(t)}{H'(t)}-\frac{H'(t)}{H(t)} = \{H(t)\}^\rho, \quad t \in \{u\colon H'(u)\,H(u) > 0\}.$$

Theorem 2·3 is a precise statement of the results that are now possible in this setting.

THEOREM 2·3. *Let* $-\infty < t < \infty$ *and let* $H_\rho(t)$ *be the family of survival functions given by*

$$H_0(t) = \exp(-e^t) \quad (\rho = 0), \quad H_\rho(t) = (1+\rho e^t)^{-1/\rho} \quad (\rho > 0).$$

Let $S^\rho(t,\theta) = H_\rho\{g(t)+\theta\}$ *and let* $S^\rho(t,\theta_i^N)$ $(i = 1,2)$ *be a sequence of location alternatives, with* θ_i^N *defined as above. Let* $\rho \geqslant 0$ *be fixed and known; let* z_α *be the* α *quantile of a standard normal distribution.*

(a) *The level α test which rejects H_0^N: $\theta_1^N = \theta_2^N$ in favour of $\bar{H}_1^N$: $\theta_1^N \neq \theta_2^N$ whenever*

$$(V)^{-\frac{1}{2}} |G^{\rho}_{N_1,N_2}| > z_{1-\frac{1}{2}\alpha}$$

has maximum efficacy against the contiguous alternatives $S^{\rho}(t, \theta_i^N)$ $(i = 1, 2)$ among all tests based on statistics of the form

$$\int_0^{\infty} K(u) \left\{ \frac{dN_1(u)}{Y_1(u)} - \frac{dN_2(u)}{Y_2(u)} \right\}.$$

(b) *A level α test which rejects H_0^N according to the criterion given in part* (a) *is a fully efficient test against time-transformed location alternatives to $H_\rho(t)$ if and only if $\pi_1 = \pi_2$ almost surely with respect to the probability measure specified by H_ρ.*

Proof. Part (a) follows from Lemma 5·2·1 of Gill (1980), while part (b) follows from Corollary 5·3·1 in the same paper.

Since tests based on $G^{\rho}_{N_1,N_2}$ reduce to the log rank when $\rho = 0$ and are nearly equal to the Peto & Peto test when $\rho = 1$, it is not surprising that $H_1(t)$ is the usual logistic survival function, nor that as $\rho \to 0$

$$\lim H_\rho(t) = \exp(-e^t).$$

The fully efficient nature of the Peto & Peto test against logistic shift alternatives and of the log rank test against type I extreme-value shift alternatives is already well known. By using the full family of H_ρ distributions, however, it is now possible to study the behaviour of rank tests which are optimal against models exhibiting a specific degree of departure, as determined by ρ, from the popular proportional hazards model in the direction of the often-used logistic location shift model. Interestingly, this family of H_ρ distributions is an important subset of the generalized-F family of distributions discussed by Prentice (1975); in the notation of that paper, H_ρ is obtained by taking $m_1 = 1$ and $m_2 = \rho^{-1}$.

We examine the behaviour of the H_ρ survival functions under two-sample time-transformed location alternatives. Suppose that ρ is fixed and that we wish to consider modelling two samples of survival data with the distributions $S_i(t) = H_\rho\{g(t) + \theta_i\}$ $(i = 1, 2)$. If one takes $\Delta = \theta_1 - \theta_2$ then

$$S_2 = S_1[(S_1)^\rho + \{1 - (S_1)^\rho\} e^\Delta]^{-1/\rho}. \tag{2·2}$$

In fact, since the efficiencies of rank tests are invariant under monotone transformations of the data, tests based on $G^{\rho}_{N_1,N_2}$ will be fully efficient against alternatives in which S_1 is arbitrary, and S_2 is given by (2·2). The right-hand side of (2·2) is a specific instance of the conversion function discussed by Peto & Peto (1972). If we use (2·2), and if $\lambda_i(t)$ is the hazard function corresponding to $S_i(t)$, then

$$\lambda_2 = \lambda_1 e^\Delta [(S_1)^\rho + \{1 - (S_1)^\rho\} e^\Delta]^{-1}.$$

The relative behaviour of the two distributions S_1 and S_2 is now most clearly understood by taking $S_1(t) = e^{-t}$, a unit exponential, and studying the ratio $\lambda_2(t)/\lambda_1(t) = \lambda_2(t)$. In this case

$$\lambda_2(t) = e^\Delta \{e^{-\rho t} + (1 - e^{-\rho t}) e^\Delta\}^{-1}$$

and we call this term $R(\Delta, \rho t)$. At $t = 0$, $\lambda_2(t) = e^{\Delta}\lambda_1(t)$, and hence e^{Δ} represents the initial ratio of the hazard functions. Figure 1 illustrates the behaviour of $R(\Delta, \rho t)$ for some representative values of Δ and t. We have chosen Δ so that $e^{\Delta} = 2^b$ for various values of b. The plots have been made on a semi-$\log_2$ scale, with the horizontal axis corresponding to ρt, since in this setting ρ acts simply as a scale factor.

Notice that $R(\Delta, \frac{1}{2}t)$ is much closer to $R(\Delta, t)$ than to $R(\Delta, 0)$. Thus one must be careful not to assume that choosing $\rho = \frac{1}{2}$ in a modelling situation leads to a set of location alternatives which are 'midway' between extreme value and logistic shift alternatives. For exactly the same reason one should therefore not consider tests based on $G^{\frac{1}{2}}_{N_1,N_2}$ as

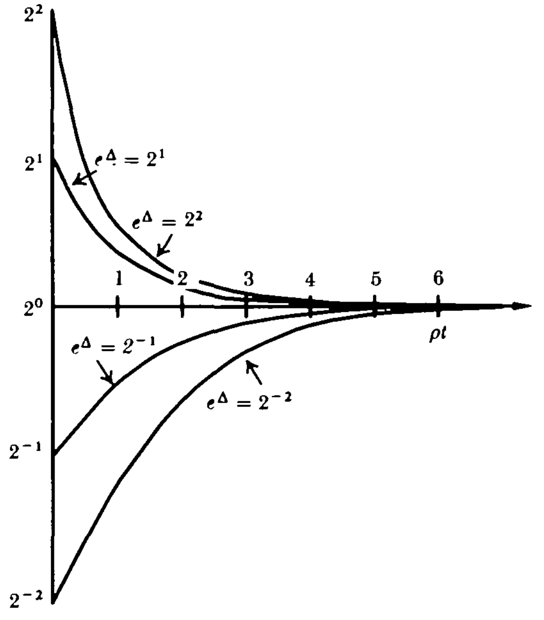

Fig. 1. The hazard ratio $R(\Delta, \rho t)$

tests which provide a balanced compromise between the log rank and the Peto & Peto tests.

It is possible to compute Pitman asymptotic relative efficiencies for the family of test statistics $G^{\rho}_{N_1,N_2}$. For simplicity, we assume that $C_1(t) = C_2(t) = C(t)$, and hence that $\pi_1(t) = \pi_2(t) = \pi(t)$ under H_0.

In our setting

$$\theta_i^N - \theta_0 = (-1)^{i+1}\left\{\frac{N_{3-i}}{N_i(N_1+N_2)}\right\}^{\frac{1}{2}}.$$

Since

$$\frac{d\beta_i^N(t)}{d\beta} - 1 \equiv \frac{\lambda(t, \theta_i^N) - \lambda(t, \theta_0)}{\lambda(t, \theta_0)},$$

we have that

$$\gamma_i(t) = \lim_{N\to\infty}\left(\frac{N_{3-i}}{N_1+N_2}\right)\left\{\frac{N_i(N_1+N_2)}{N_{3-i}}\right\}^{\frac{1}{2}}\left\{\frac{\lambda(t, \theta_i^N) - \lambda(t, \theta_0)}{\lambda(t, \theta_0)}\right\}$$

$$= \left[(-1)^{i+1} a_{3-i}\frac{\partial}{\partial\theta}\log\lambda(t, \theta)\right]_{\theta=\theta_0}$$

Thus we have

$$\gamma(t) = \gamma_1(t) - \gamma_2(t) = \left[\frac{\partial}{\partial\theta}\log\lambda(t,\theta)\right]_{\theta=\theta_0}.$$

We examine the behaviour of the statistic $G^{\rho}_{N_1,N_2}$ under contiguous alternatives for a distribution H_{ρ^*}, where ρ^* and ρ may or may not be equal. For $S(t,\theta) = H_{\rho^*}\{g(t)+\theta\}$,

$$\log\lambda(t,\theta) = g(t)+\theta+\log g'(t) - \log\{1+\rho^* e^{g(t)+\theta}\}.$$

Hence $\gamma(t) = \{1+\rho^* e^{g(t)+\theta_0}\}^{-1} = \{S^{\rho^*}(t,\theta_0)\}^{\rho^*}$. Denote $S^{\rho^*}(t,\theta_0)$ by $S(t)$. The asymptotic efficacy of $G^{\rho}_{N_1,N_2}$ for the alternatives $S^{\rho^*}(t,\theta_i^N) = H_{\rho^*}\{g(t)+\theta_i^N\}$ is given by

$$\begin{aligned} e_{\rho^*}(\rho) &= (\mu_\rho/\sigma_\rho)^2 \\ &= \left[\int_0^\infty \pi(u)\{S(u)\}^{\rho^*+\rho}\,d\beta(u)\right]^2 \Big/ \int_0^\infty \pi(u)\{S(u)\}^{2\rho}\,d\beta(u). \end{aligned}$$

The expression $e_{\rho^*}(\rho)/e_{\rho^*}(\rho^*)$ is the Pitman asymptotic relative efficiency of tests based on $G^{\rho}_{N_1,N_2}$ for location alternatives under H_{ρ^*} with respect to the fully efficient test $G^{\rho^*}_{N_1,N_2}$. The ratio may be evaluated in closed form when, for instance, $C(u) \equiv \{S(u)\}^{\alpha}$, and equals $(2\rho+\alpha+1)(2\rho^*+\alpha+1)/(\rho^*+\rho+\alpha+1)^2$. Although one would not anticipate that $C(u) = \{S(u)\}^{\alpha}$ in many cases, values of α can be used in the above expression to infer qualitative information about the effect that the severity of censorship has on the asymptotic relative efficiencies of these procedures. Of course, numerical integration can be used to evaluate $e_{\rho^*}(\rho)$ for any pair of survival and censoring distributions.

3. Results when the number of samples differs from two

3·1. *More than two samples*

Properties of the k sample, $k > 2$, versions of the G^{ρ} statistic can be established with the theory of linear rank statistics. When the data are uncensored, with $r+1 > 2$, the following result holds.

THEOREM 3·1. *Let* $S_i(t) = H_\rho\{g(t)+\theta_i\}$ $(i = 1,\ldots,r+1)$. *Suppose we wish to test* H_0: $\theta_1 = \theta_2 = \ldots = \theta_{r+1}$ *against the global alternative* H_1: $\theta_i \neq \theta_j$ *for some pair* (i,j) *with* $i \neq j$. *Then tests based on* G^{ρ} *are asymptotically equivalent to the locally most powerful rank test for testing* H_0.

Proof. Let $N = N_1 + \ldots + N_{r+1}$. Then one can show that G^{ρ} may be written as a linear rank statistic $\Sigma_k c_N^*(k)\tilde{z}_k$ with

$$c_N^*(k) = \{\hat{S}(T_k)\}^{\rho} - \sum_{l=1}^{k}\{\hat{S}(T_l)\}^{\rho}/n_l.$$

To construct the locally most powerful rank test, let

$$F_\rho(t) = 1 - H_\rho(t),\quad f_\rho(t) = \frac{dF_\rho(t)}{dt},\quad f'_\rho(t) = \frac{df_\rho(t)}{dt}\quad (-\infty < t < \infty),$$

$$\phi_\rho(u) = -f'_\rho\{F_\rho^{-1}(u)\}/f_\rho\{F_\rho^{-1}(u)\}\quad (0 < u < 1).$$

Then, if we let scores $\{c_N(k);\ k = 1,\ldots,N\}$ be given by $c_N(k) = -\phi_\rho\{k/(N+1)\}$, the locally most powerful rank test is asymptotically equivalent to that based on $\Sigma_k c_N(k)\tilde{z}_k$.

The asymptotic equivalence of a test based on the scores $c_N^*(k)$ to that based on scores $c_N(k)$ then follows from the fact that $\lim N^{-1} \Sigma_k \{c_N(k) - c_N^*(k)\}^2 = 0$, which is not difficult to establish in this situation. See Randles & Wolfe (1979, pp. 287, 319) for the sufficiency of this condition.

Since $G^\rho = \Sigma_k c_N^*(k) \tilde{z}_k$, with scores $c_N^*(k)$ asymptotically equivalent to those given by $c_N(k) = -\phi_\rho\{k/(N+1)\}$, we may rely on the large-sample distribution theory for $\tilde{R} = \Sigma_k c_N(k) \tilde{z}_k$ to find approximate critical regions for tests of H_0: $\theta_1 = \ldots = \theta_{r+1}$. Theorem V.2.2 of Hájek & Sĭdák (1967, p. 170) establishes that $\tilde{R}$, and hence G^ρ, may be used to construct a hypothesis test based upon the χ^2 distribution. Specifically, if we define

$$\bar{c}_N = N^{-1} \sum_{k=1}^{N} c_N(k),$$

$$\tilde{R}_c' = \{N_1^{-\frac{1}{2}}(R_1 - N_1 \bar{c}_N), \ldots, N_r^{-\frac{1}{2}}(R_r - N_r \bar{c}_N)\},$$

where R_i is the ith component of $\tilde{R}$, then one may show that under H_0

$$Q = (N-1)\left[\sum_{k=1}^{N} \{c_N(k) - \bar{c}_N\}^2\right]^{-1} \tilde{R}_c' \tilde{R}_c$$

is asymptotically distributed as χ_r^2. Of course, all of the above results hold with $c_N(k)$ replaced by $c_N^*(k)$.

In censored data, very little seems to be known currently about the properties of rank statistics when there are more than two samples. We conjecture, however, that the methods used by Gill (1980) in computing efficiencies for two sample tests can be extended to situations in which there are more than two samples. Prentice (1978) has proposed a method for modifying the usual score function tests to censored data, and it is possible to compute his modified scores for the H_ρ distributions. Suppose m_k is the number of censored observations in the interval $[T_k, T_{k+1})$, and let $\tilde{z}_{kl}$ $(l = 1, \ldots, m_k)$ be the 0–1 regression vectors indicating sample membership of each of the censored observations. Then in our situation and in the notation of Prentice, censored data rank tests may be based on

$$\tilde{v} = \sum_{k=1}^{d} \{\tilde{z}_k \hat{c}_N(k) + \tilde{s}_{(k)} \hat{C}_N(k)\},$$

where $\hat{c}_N(k)$ is a score for an uncensored observation, $\hat{C}_N(k)$ is a score for a censored observation, and $\tilde{s}_{(k)} = \Sigma_l \tilde{z}_{kl}$. The scores $\hat{c}_N(k)$ and $\hat{C}_N(k)$ are given by

$$\hat{c}_N(k) = \frac{1}{\rho} - \left(\frac{\rho+1}{\rho}\right) \prod_{j=1}^{k} \left(\frac{n_j}{n_j+\rho}\right), \quad \hat{C}_N(k) = \frac{1}{\rho}\left\{1 - \prod_{j=1}^{k} \left(\frac{n_j}{n_j+\rho}\right)\right\}.$$

Also $\hat{c}_N(k)$ are the exact scores in uncensored data.

Hypothesis tests may be based on the censored data rank statistics in much the same way as for the uncensored data case. The form of χ^2 tests, which are based on the assumption of asymptotic multivariate normality for the vector $\tilde{v}$, is clearly specified by Prentice (1978, pp. 170–5). We will not examine formal proofs here establishing this asymptotic normality. It is expected, however, that such proofs can be constructed, under appropriate assumptions regarding the censoring distributions, by appealing to the results of Hájek & Sĭdák (1967, p. 152).

3·2. *One-sample test statistic*

As mentioned earlier, one-sample goodness-of-fit G^ρ test procedures can be formulated. In this situation the statistics can easily be approached from the stochastic integral point of view. The argument here is similar to that used by Woolson (1981), and is heuristic in nature.

Recall that for two samples

$$G^\rho_{N_1,N_2} = \int_0^\infty \{\hat{S}(u)\}^\rho \left\{\frac{Y_1(u)\,Y_2(u)}{Y_1(u)+Y_2(u)}\right\}\left\{\frac{dN_1(u)}{Y_1(u)} - \frac{dN_2(u)}{Y_2(u)}\right\}.$$

Suppose now we think of N_1 as being arbitrarily large, giving us complete information about a survival distribution S. In this case

$$\{Y_1(u)\,Y_2(u)\}/\{Y_1(u)+Y_2(u)\} \simeq Y_2(u), \quad \hat{S}(u) \simeq S(u), \quad dN_1(u)/\{Y_1(u)\} \simeq d\beta(u).$$

Let S_0 be a hypothesized distribution, and $\beta_0 = -\log S_0$. Set $N = N_2$, $Y(u) = Y_2(u)$, $X^0_j = X^0_{2j}$ and $\delta_j = \delta_{2j}$; we may write G^ρ in this situation as

$$G^\rho_N = \int_0^\infty \{S_0(u)\}^\rho \{Y(u)\,d\beta_0(u) - dN(u)\}.$$

If $\rho > 0$, this becomes

$$\sum_{j=1}^{N} \rho^{-1}[1-\{S_0(X^0_j)\}^\rho] - [\delta_j\{S_0(X^0_j)\}^\rho],$$

while, for $\rho = 0$, G^ρ_N is

$$\sum_{j=1}^{N} \{\beta_0(X^0_j) - \delta_j\} = \sum_{j=1}^{N} [\log\{S_0(X^0_j)\}^{-1} - \delta_j].$$

The statistics G^ρ_N can be used to formulate a family of one-sample test statistics which include a one-sample version of the log rank statistic as a special case. Approximate critical values for these tests can be found by appealing to the following theorem, which follows from the results of Gill (1980).

THEOREM 3·2. *Assume that the mild regularity conditions hold which are outlined in* § 4.2 *of Gill* (1980).

(a) *Under* H_0: $S = S_0$, *the statistic* $N^{-\frac{1}{2}} G^\rho_N$ *is asymptotically normally distributed with mean* 0 *and variance*

$$\int_0^\infty \{S_0(u)\}^{2\rho}\,\pi(u)\,d\beta_0(u) = \int_0^\infty \{S_0(u)\}^{2\rho+1}\,C(u)\,d\beta_0(u).$$

(b) *The asymptotic variance in part* (a) *may be consistently estimated by*

$$\int_0^\infty \{S_0(u)\}^{2\rho}\,Y(u)\,N^{-1}\,d\beta_0(u) = \begin{cases} N^{-1}\sum_{j=1}^{N}(2\rho)^{-1}[1-\{S_0(X^0_j)\}^{2\rho}] & (\rho > 0), \\ N^{-1}\sum_{j=1}^{N}\log S_0(X^0_j), & (\rho = 0). \end{cases}$$

It is interesting to note that in uncensored data, Theorem 3·2(a) implies that the statistic

$$\{(2\rho+1)/N\}^{\frac{1}{2}} \sum_{j=1}^{N} \int_0^{X_j^0} \{S_0(u)\}^\rho \, d\beta_0(u) - [\delta_j\{S_0(X_j^0)\}^\rho]$$

has asymptotically a standard normal distribution. Observe also from Theorem 3·2 that the one-sample censored data log rank statistic

$$\left[\sum_{j=1}^{N} \delta_j - \sum_{j=1}^{N} \log\{S_0(X_j^0)\}^{-1}\right]^2 \Big/ \sum_{j=1}^{N} \log\{S_0(X_j^0)\}^{-1}$$

has asymptotically a χ^2 distribution with one degree of freedom. In this setting, $\Sigma_j \delta_j$ is the observed number of deaths and $\Sigma_j \delta_j - \Sigma_j \log\{S_0(X_j^0)\}^{-1}$ has mean zero. We close this section with a reference to a recent unpublished independent manuscript by P. K. Andersen, Ø. Borgan, R. Gill and N. Keiding brought to our attention during the refereeing process on our work. They use martingale theory to establish asymptotic distribution theory under the null hypothesis for general classes of k-sample test statistics.

4. Results of Monte Carlo simulations

4·1. *General remarks*

We have indicated how the asymptotic distributions of the G^ρ statistics ($\rho \geqslant 0$) can be used to construct hypothesis tests of a given size. Simulation has confirmed that the true size of each of these test procedures, in small or moderate equal sample sizes and under varying amounts of censorship, is indeed accurately approximated by the nominal significance level based upon this asymptotic distribution theory. The simulations also have been used to confirm, in uncensored data, the analytical conclusions reached earlier concerning the role of ρ in determining power. The results from the evaluation of power were particularly interesting and will be given in the remainder of this section.

4·2. *Simulation procedure*

In the simulations to evaluate power, four distinct configurations of survival distributions were inspected, with each configuration including two survival distributions used to generate two samples of failure times. Attention in this simulation study was restricted to the two-sample problem. If, as earlier, we let $S^{\rho^*}(t,\theta) = H_{\rho^*}\{g(t)+\theta\}$ denote a time-transformed location shift of $H_{\rho^*}(t)$, then the four configurations considered were

$$S^{\rho^*_i}(\theta_{1i},\theta_{2i}) \equiv \{S^{\rho^*_i}(t,\theta_{1i}), S^{\rho^*_i}(t,\theta_{2i})\} \quad (i = 1,2,3,4).$$

In turn, the four test statistics evaluated were G^{ρ_i} ($i = 1,2,3,4$). Since it was of particular interest to obtain in the class of configurations $\{S^{\rho^*}(\theta_1,\theta_2);\ \rho \geqslant 0\}$ a small sample comparison of the behaviour of the log rank and Wilcoxon test statistics, that is G^0 and G^1 respectively, with that of other G^ρ statistics, the values of ρ_i^* and ρ_i chosen were 0, $\frac{1}{2}$, 1 and 2.

Let, for example, the time transformation $g(t) = \log t$. Then the resulting survival distributions are $S^{\rho^*}(t,\theta) = (1+\rho^* e^\theta t)^{-1/\rho^*}$ if $\rho^* > 0$ while $S^0(t,\theta) = \exp(-e^\theta t)$. Thus, using the transformation $(U^{-\rho^*}-1)/(\rho^* e^\theta)$ when $\rho^* > 0$, and $-(\log U)/e^\theta$ when $\rho^* = 0$, the appropriate independent survival random variables were obtained by transforming independent uniformly distributed variates, U, produced with a linear congruential random number generator as described by Knuth (1981, §3.2).

Since the main purpose was to investigate in small and moderate samples the performance of G^ρ procedures derived using asymptotic properties, sample sizes $N_1 = N_2 = 20$ and $N_1 = N_2 = 50$ were considered.

Five hundred pairs of samples were generated for each selected configuration of survival distributions and for each sample size. The proportions of samples in which each one-sided test procedure under consideration rejected H_0 at the $\alpha = 0{\cdot}01$ and $\alpha = 0{\cdot}05$ significance levels were calculated.

4·3. *Power results*

Results of the Monte Carlo study pertaining to the evaluation of power of the set of procedures $\{G^\rho\colon \rho = 0, \frac{1}{2}, 1, 2\}$ are presented in Table 1. Figure 2 presents the plots of the hazard functions corresponding to the four survival configurations inspected in

Table 1. *Monte Carlo estimates of the power of the G^ρ ($\rho = 0, \frac{1}{2}, 1, 2$) one-sided test procedures of H_0: $S_1 = S_2$ against H_A: $S_2 > S_1$; 500 simulations; $N_1 = N_2 = N^*$*

	S_1† (ρ^*, e^{θ_1})	S_2 (ρ^*, e^{θ_2})	N^*	Level of test	G^ρ $\rho = 0$	$\rho = \frac{1}{2}$	$\rho = 1$	$\rho = 2$
I	(0, 2)	(0, 1)	20	0·01	0·386	0·338	0·292	0·204
				0·05	0·668	0·620	0·578	0·456
			50	0·01	0·858	0·800	0·734	0·610
				0·05	0·954	0·938	0·894	0·812
					(1·000)	(0·889)	(0·750)	(0·556)
II	($\frac{1}{2}$, 2·25)	($\frac{1}{2}$, 1)	20	0·01	0·308	0·320	0·290	0·258
				0·05	0·548	0·576	0·564	0·516
			50	0·01	0·646	0·694	0·682	0·604
				0·05	0·844	0·878	0·868	0·830
					(0·889)	(1·000)	(0·960)	(0·816)
III	(1, 2·5)	(1, 1)	20	0·01	0·206	0·222	0·234	0·204
				0·05	0·444	0·470	0·488	0·470
			50	0·01	0·534	0·616	0·624	0·598
				0·05	0·754	0·834	0·864	0·828
					(0·750)	(0·960)	(1·000)	(0·938)
IV	(2,3)	(2, 1)	20	0·01	0·148	0·186	0·202	0·206
				0·05	0·336	0·402	0·416	0·426
			50	0·01	0·294	0·406	0·470	0·516
				0·05	0·534	0·662	0·722	0·742
					(0·556)	(0·816)	(0·938)	(1·000)

† Each survival distribution is of the form $S^{\rho^*}(t, \theta_i) = (1 + \rho^* e^{\theta_i} t)^{-1/\rho^*}$. Tabled numbers in parentheses give Pitman asymptotic relative efficiencies, $e_{\rho^*}(\rho)/e_{\rho^*}(\rho^*) = (2\rho + 1)(2\rho^* + 1)/(\rho + \rho^* + 1)^2$

Table 1. The small and moderate sample size relative power of these four tests is entirely consistent with their large-sample asymptotic relative efficiencies. In the time-transformed extreme value location alternative, configuration I, and in the time-

transformed logistic location alternative, configuration III, G^0, that is the log rank, has asymptotic relative efficiency 1 and 0·75 respectively while G^1, that is the Wilcoxon, has asymptotic relative efficiency 0·75 and 1 respectively. This clear superiority of G^0 over G^1 in I and of G^1 over G^0 in III is equally apparent in small samples. In addition, Table 1 reveals in smaller samples that the loss in power obtained by using $G^{\frac{1}{2}}$ rather than G^1 in

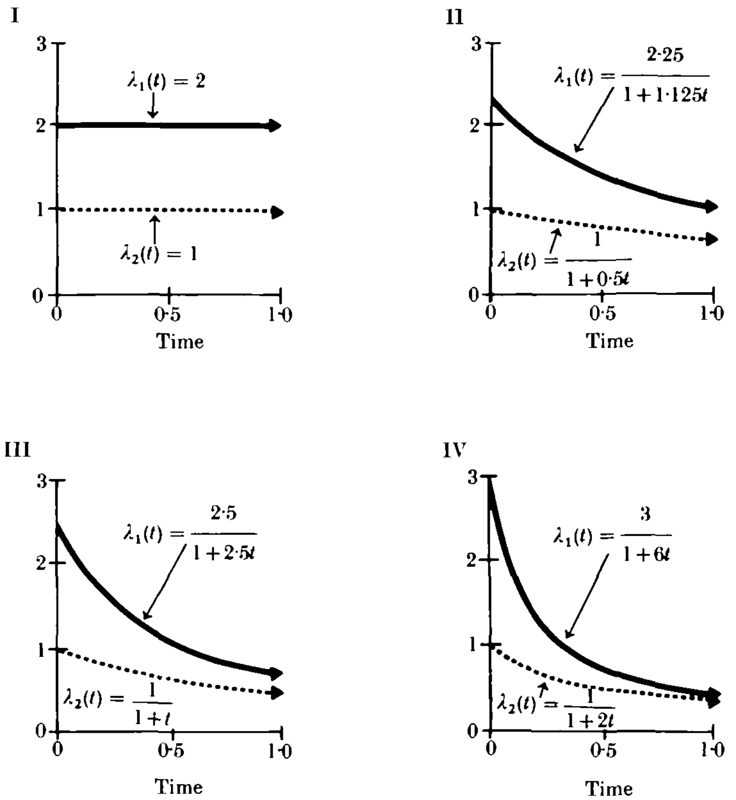

Fig. 2. Hazard function plots for alternative hypothesis simulation configurations; see Table 1 for survival configurations I, II, III and IV.

III is less than that obtained by using $G^{\frac{1}{2}}$ rather than G^0 in I, and secondly that $G^{\frac{1}{2}}$ is more powerful than G^2 in III. This confirms earlier efficiency results. Again, in configuration II we find agreement between our small sample results and asymptotic relative efficiency calculations. Specifically $G^{\frac{1}{2}}$ is more powerful than G^1 which in turn is more powerful than G^0. Finally, small sample confirmation of the facts that G^0 and G^2 have relatively low power in configurations IV and I respectively.

We thank Dr Stephanie Green for many helpful comments during revision of this manuscript. The research was supported by grants from the National Science Foundation and the National Cancer Institute.

References

AALEN, O. O. (1977). Weak convergence of stochastic integrals related to counting processes. *Z. Wahr. verw. Geb.* **38**, 251–77.

AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701–26.

BRESLOW, N. (1970). A generalized Kruskal–Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* **57**, 579–94.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc.* B **34**, 187–220.

FLEMING, T. R. & HARRINGTON, D. P. (1981). A class of hypothesis tests for one and two samples of censored survival data. *Comm. Statist.* A **10**, 763–94.

GEHAN. E. (1965). A generalized Wilcoxon test for comparing arbitrarily singly censored samples. *Biometrika* **52**, 203–23.

GILL, R. D. (1980). *Censoring and Stochastic Integrals.* Mathematical Centre Tracts 124. Amsterdam: Mathematische Centre.

HÁJEK, J. & SIDÁK, Z. (1967). *Theory of Rank Tests.* New York: Academic Press.

KNUTH, D. E. (1981). *The Art of Computer Programming.* **2**, *Seminumerical Algorithms*, 2nd edition. Reading, Mass: Addison-Wesley.

KOZIOL, J. A. & REID, N. (1977). On multiple comparisons among K samples subject to unequal patterns of censorship. *Comm. Statist.* A **6**, 1149–64.

MANTEL, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemo. Rep.* **50**, 163–70.

PETO, R. & PETO, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). *J. R. Statist. Soc.* A **135**, 185–206.

PRENTICE, R. L. (1975). Discrimination among some parameteric models. *Biometrika* **62**, 607–14.

PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167–79.

PRENTICE, R. L. & MAREK, P. (1979). A qualitative discrepancy between censored data rank tests. *Biometrics* **35**, 861–7.

RANDLES, R. H. & WOLFE, D. A. (1979). *Introducion to the Theory of Nonparametric Statistics.* New York: Wiley.

TARONE, R. E. & WARE, J. (1977). On distribution free tests for equality of survival distributions. *Biometrika* **64**, 156–60.

WOOLSON, R. F. (1981). Rank tests and a one-sample log rank test for comparing observed survival data to a standard population. *Biometrics* **37**, 687–96.

[*Received February* 1981. *Revised August* 1981]