A CLASS OF RANK TEST PROCEDURES FOR CENSORED SURVIVAL DATA

by

DAVID P. HARRINGTON and THOMAS R. FLEMING

Technical Report Series, No. 12
February 1981

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 22901

Thomas R. Fleming
Department of Medical Research Statistics and Epidemiology
Mayo Clinic
Rochester, Minnesota 55901

Summary

A class of linear rank statistics are proposed for the k-sample problem with right censored survival data. The class contains as special cases the logrank test (Mantel, 1966; Cox, 1972) and a test essentially equivalent to the Peto and Peto (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 suggest that the asymptotic properties of these statistics in the two-sample case hold fairly well in small and moderate size samples.

Some key words: Censored data, contiguous alternatives, linear rank statistic, logrank test, Pitman asymptotic relative efficiency, Wilcoxon test.

I. Introduction

Some attention has been given in the recent survival theory literature to alternative formulations for the popular nonparametric statistics used in testing the null hypothesis of equality of several underlying survival distributions. Prentice (1978) has shown that statistics such as the logrank (L-R) (Mantel, 1966) and Peto and Peto's generalization of the Wilcoxon (1972) (PPW) can be obtained as linear rank statistics for censored data assumed to come from a log linear model. Prentice and 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 popular two sample tests can be formulated as stochastic integrals with respect to appropriately defined counting processes, and has used martingale theory [Aalen (1977)] in this context to obtain weak convergence results for some estimators and test statistics. Gill (1979) has used this martingale formulation to obtain Pitman asymptotic relative efficiencies (a.r.e.) for statistics such as logrank and both the Gehan (1965) and Peto and Peto (1972) generalizations of Wilcoxon's statistic under contiguous sequences of two-sample location alternatives. In this paper, we will propose a class of nonparametric procedures which includes the L-R and the PPW as special cases, and we will show how the formulations mentioned above can be used to study the properties of these procedures.

The most intuitive formulation for these new statistics is in the context outlined by Prentice and Marek (1979). The notation necessary for this formulation is as follows. Let $\{X_{ij};\ 1 \leq j \leq N_i,\ 1 \leq i \leq r+1\}$ denote the survival variables (i.e., death times) from r+1 independent random samples with underlying survival functions $S_i(t) = P(X_{ij} \geq t)$, $i=1,2,\ldots,r+1$. We will assume throughout that the S_i are absolutely continuous with cumulative hazard and hazard functions defined, respectively, as $\beta_i(t) = -\ln S_i(t)$ and $\lambda_i(t) = \frac{d}{dt}\beta_i(t)$. Let $\{Y_{ij};\ 1 \leq j \leq N_i,\ 1 \leq i \leq r+1\}$ denote censoring variables with censoring distributions $C_i(t) = P(Y_{ij} \geq t)$. We will assume that the observed data consist of $\{X^0_{ij} = \min(X_{ij}, Y_{ij}),\ \delta_{ij} = I[X_{ij} \leq Y_{ij}];\ 1 \leq j \leq N_i,\ 1 \leq i \leq r+1\}$, 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. We will assume that the survival and censoring times are independent and thus $\pi_i(t) \equiv P(X^0_{ij} \geq t) = S_i(t)\, C_i(t)$. Let $T_1 < T_2 < \ldots < T_d$ denote the d distinct ordered observed death times in the pooled sample, and let n_{ik} be the number of subjects under observation in sample i just prior to T_k, i.e., n_{ik} is the size of the risk set in sample i at T_k. $\sum_{i=1}^{r+1} n_{ik} = n_k$ will be the size of the total risk set in the pooled sample at T_k. We will let $\vec{z}_k{}' = (z_{1k},\ldots,z_{rk})$ denote the transpose of a column vector of 0-1 regression coefficients denoting sample membership: $z_{ik}=1$ if and only if the observed death at T_k is from sample i. $\vec{z}_k = \vec{0}$ will mean that the death at T_k is in sample r+1. If we let $\{w_j;\ 1 \leq j \leq d\}$ be a set of weights, then Prentice and Marek (1979) have shown that several of the popular nonparametric tests for H_0: $S_1 = S_2 = \ldots = S_{r+1}$ can be based on statistics of the form $\sum_{k=1}^{d} w_k(\vec{z}_k - \vec{q}_k)$, where $\vec{q}_k{}' = (n_k)^{-1}\,(n_{1k}, n_{2k}, \ldots, n_{rk})$. When $w_k \equiv 1$, the resulting test is the L-R, while when $w_k = \tilde{S}(T_k)$, with $\tilde{S}$ being the pooled sample survival function estimator $\tilde{S}(T_k) = \prod_{j=1}^{k} n_j/(n_j+1)$, the test is Prentice's (1978) generalization of the Wilcoxon test for censored data and is similar to the generalization of Peto and Peto (1972).

It is clear that a researcher should have some flexibility in choosing the weight function w_k, but the only proper way to choose weights is to pick a set yielding a test procedure as sensitive as possible to the types of departures from equality of the S_i that are anticipated in a given experiment. Gill (1979) has shown that in two samples of censored data, the L-R test has Pitman a.r.e. 1 (i.e., is fully efficient) against a time-transformed sequence of contiguous location alternatives when S_i is a type 1 extreme value survival function. The L-R test is thus fully efficient for a contiguous sequence of proportional hazards alternatives. The approach illustrated by Prentice (1978) proves the L-R 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 L-R test reduces in this setting to the Savage exponential scores test, and its properties have been known for some time. Gill (1979) has also shown that the PPW is fully efficient against time-transformed location alternatives for the logistic distribution in two samples of censored data, while again the approach in Prentice (1978) yields the result that it is a locally most powerful rank test in r+1 samples of uncensored data.

The most natural class of weights to use which generalizes the L-R and PPW weights seems to be of the form $w_k(\rho) = \{S^*(T_k)\}^\rho$, for a fixed $\rho \geq 0$ and for $S^*(T_k)$ a pooled survival function estimator at T_k. When $\rho=0$, the L-R weights are obtained, while for $\rho=1$ and $S^*=\tilde{S}$, the generalized Wilcoxon weights are produced. In this paper we will study the properties of tests based on the statistics

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

where $\hat{S}(u)$ is the left continuous version of the Kaplan-Meier product limit estimator of the survival function for the pooled r+1 samples, i.e., $\hat{S}(u) = \prod_{T_k : T_k < u} (n_k - 1)/n_k$. Section II will give 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, a characterization of the parametric location alternatives for which these procedures are fully efficient or locally most powerful rank tests, and tables of the Pitman a.r.e. for pairs of members of this class. Section III will examine those results that can be proved when the number of samples is different from two. In Section IV results of Monte Carlo simulations are displayed which support, in the two-sample situation, the claim that the asymptotic theory provides reasonably accurate approximations for the actual properties of the statistics in small and moderate sample sizes.

II. The Two-Sample Statistics

For two samples of censored survival data the most direct way to obtain the asymptotic distribution theory for the G^ρ statistics is through the use of weak convergence theorems for martingales associated with stochastic integrals of counting processes. To apply these results we need to recast our two-sample statistic in slightly different notation. Assume r+1=2; let

$$N_i(t) = \sum_{k=1}^{N_i} I[X_{ik}^0 \le t, \quad \delta_{ik} = 1], \quad i=1,2,$$

and

$$Y_i(t) = \sum_{k=1}^{N_i} I[X_{ik}^0 \ge t].$$

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

It is not hard to verify that with this notation G^ρ may be expressed as

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

with the convention that 0/0 = 0. With this formulation it is easy to see the class of statistics $G^{\rho}_{N_1,N_2}$, $\rho \geq 0$, is a subset of the more general class $^+$ considered in Gill (1979). The asymptotic distribution theory derived by Gill for that class applies here and may be summarized in Theorems 2.1 and 2.2 below. (We note in passing that Theorem 2.1 may be proved more directly using the method of proof outlined for a similar statistic in Section 6 of Fleming and Harrington (1981).)

Theorem 2.1. Let $\Delta N_i(u) = N_i(u) - \lim_{t \uparrow u} N_i(u)$ and let variance estimators $V_\ell(t)$, $\ell=1,2$, be given by the formulas:

$$V_1(t) = \sum_{i=1}^{2} \int_0^t \frac{\{\hat{S}(u)\}^{2\rho}}{Y_i(u)} \cdot \left\{\frac{Y_1(u)\, Y_2(u)}{Y_1(u)+Y_2(u)}\right\}^2 \left\{1 - \frac{\Delta N_i(u)-1}{Y_i(u)-1}\right\} \frac{dN_i(u)}{Y_i(u)}$$

$$V_2(t) = \sum_{i=1}^{2} \int_0^t \frac{\{\hat{S}(u)\}^{2\rho}}{Y_i(u)} \cdot \left\{\frac{Y_1(u)\, Y_2(u)}{Y_1(u)+Y_2(u)}\right\}^2 \left\{1 - \frac{\Delta N_1(u)+\Delta N_2(u)-1}{Y_1(u)+Y_2(u)-1}\right\} \cdot \frac{d\{N_1(u)+N_2(u)\}}{Y_1(u)+Y_2(u)} .$$

Let $N = N_1+N_2$ and assume $\lim_{N\to\infty} N_i/N \equiv a_i$ exists and satisfies $0<a_i<1$. Then under the null hypothesis H_0: $S_1=S_2$ (= S unspecified):

$$\text{(a)} \lim_{N\to\infty} \left(\frac{N_1+N_2}{N_1\, N_2}\right) V_\ell(t) = \int_0^t \frac{\pi_1(u)\pi_2(u)}{a_1\pi_1(u)+a_2\pi_2(u)} \{S(u)\}^{2\rho}\, d\beta(u), \quad \ell=1,2,$$

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

(b) $\lim_{N\to\infty} \{V_\ell(\infty)\}^{-\frac{1}{2}} G^\rho_{N_1,N_2} = Z \sim N(0,1)$ in distribution.

Proof: For i=1,2 observe that $Y_i(t)/N_i$ is simply the empirical distribution function estimator of $\pi_i(t)$. Thus $\{N^{\frac{1}{2}}\{Y_i(t)/N_i - \pi_i(t)\} : 0 \le t < \infty\}$ converges weakly to a time-transformed Brownian bridge, implying

$$\sup_{0 \le t < \infty} \left| Y_i(t)/N_i - \pi_i(t) \right| \longrightarrow 0$$

in probability as $N\to\infty$. The proof then follows directly from Proposition 4.3.3 in Gill (1979).

There are several pertinent remarks that can be made here. First, Gill's complete asymptotic results can yield more powerful results about the statistic $G^\rho_{N_1,N_2}$ than those we have stated here. It is possible to show that under certain conditions the empirical processes

$$\left\{ \left(\frac{N_1+N_2}{N_1 N_2}\right)^{\frac{1}{2}} G^\rho_{N_1,N_2}(t) : 0 \le t < \infty \right\}$$

converge weakly to a mean zero independent increment Gaussian process under H_0, even for some survival distributions that have discrete probability mass at certain time points. We will not need these general results here, however.

Second, the variance estimator $V_2(\infty)$ is identical to the hypergeometric variance estimator that arises naturally when one views the survival data as producing a series of independent 2 X 2 contingency tables, one at each death time, as was originally done by Mantel (1966) and as illustrated by Prentice and Marek (1979). In the original notation for the test statistic, $V_2(\infty)$ takes the more familiar form

$$V_2(\infty) = \sum_{k=1}^{d} w_k^2 \left(\frac{n_{1k}}{n_k}\right)\left(1 - \frac{n_{1k}}{n_k}\right)\left(\frac{n_k - d_k}{n_k - 1}\right) d_k,$$

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

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

where z_{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 $G^\rho_{N_1,N_2}$. In order to understand better the asymptotic power of these procedures, however, some information is needed about the behavior of these statistics under alternative hypotheses. It is not hard to show that $G^\rho_{N_1,N_2}$ is consistent against the alternatives H_1: $\lambda_1(t) \geq \lambda_2(t)$, $t \in \{u: S_1(u)S_2(u) > 0\}$; and H_2: $\beta_1(t) \geq \beta_2(t)$, $0 \leq t < \infty$, with each of the inequalities being strict inequalities on some interval. (See Gill (1979), Section 4.1 for details.) Asymptotic power functions must be compared under a sequence of alternative hypotheses that approaches the null or with a sequence of significance levels that approaches zero. We will calculate Pitman a.r.e. under a contiguous sequence of alternative hypotheses. The following theorem follows directly from Theorem 4.2.1 in Gill (1979).

<u>Theorem 2.2</u>. Let $S_i^N(t)$, i=1,2, be a sequence (in N) of survival functions which satisfies $\lim_{N\to\infty} S_i^N(t) = S(t)$ uniformly in $t \in [0,\infty)$. Let β_i^N be the associated cumulative hazard functions. Suppose we define

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

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

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

in distribution, where

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

and (2.1)

$$\{\sigma_\rho(t)\}^2 = \int_0^t \frac{\pi_1(u)\pi_2(u)}{a_1\pi_1(u) + a_2\pi_2(u)} \{S(u)\}^{2\rho} d\beta(u) .$$

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 $e(\rho,t) = \{\mu_\rho(t)/\sigma_\rho(t)\}^2$ 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^N_i, $i=1,2$, and let $1 - S^N_i(t) = \Psi(g(t) + \theta^N_i)$, where Ψ is a fixed cumulative distribution function and $g(t)$ is an arbitrary monotonically increasing time transformation. Let $\psi(t) = \frac{d}{dt}\Psi(t)$, $\lambda(t) = \psi(t)\{1-\Psi(t)\}^{-1}$, $\ell(t) = \ell n\,\lambda(t)$, and $\ell'(t) = \frac{d}{dt}\ell n\,\lambda(t)$. To achieve the right rate of convergence for the contiguous alternatives we let

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

where θ_0 is unspecified and c is an arbitrary positive constant. 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 $\tilde{H}_1^N : \theta_1^N \neq \theta_2^N$, and that we wish to restrict ourselves to statistics of the form

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

where K(u), $0 \leq u < \infty$, is a stochastic process satisfying the regularity conditions outlined in Section 3.3 of Gill (1979). Then Gill has shown that the resulting asymptotic efficacy will be maximized for the contiguous location alternatives above if K is chosen as $K(u) = \ell' [\Psi^{-1} \{1-\hat{S}(u)\}]$, where $\hat{S}$ is a 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

$$\ell' [\Psi^{-1} \{1-\hat{S}(u)\}] = \{\hat{S}(u)\}^{\rho} .$$

A sufficient condition to ensure this is to have $\ell' = (1-\Psi)^{\rho}$. Letting $H(t) = 1-\Psi(t)$, 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:

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

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

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

Let $S^\rho(t,\theta) = H_\rho\{g(t)+\theta\}$ be a time-transformed location shift of $H_\rho(t)$, 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 $\rho\geq 0$ 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 favor of $\tilde{H}_1^N$: $\theta_1^N \neq \theta_2^N$ whenever

$$\{V_\ell(\infty)\}^{-\frac{1}{2}} \left| G^\rho_{N_1,N_2} \right| > z_{1-\alpha/2} \qquad (\ell=1 \text{ or } 2)$$

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 in Gill (1979), 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 L-R when $\rho=0$ and asymptotically to the PPW when $\rho=1$, it is not surprising that $H_1(t)$ is the usual logistic survival function, nor that $\lim_{\rho\to 0} H_\rho(t) = \exp(-e^t)$. The fully efficient nature of PPW against logistic shift alternatives and of the L-R test against type 1 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 behavior of rank tests which are optimal against models exhibiting a specific degree (as determined by ρ) of departure 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). Specifically, Prentice considered the log-linear model in which the error distribution was assumed to be that of the logarithm of an F variate on $2m_1^*$ and $2m_2^*$ degrees of freedom. H_ρ is obtained by taking $m_1^*=1$ and $m_2^*=\rho^{-1}$.

Using results quoted earlier, it is possible to tabulate Pitman asymptotic relative efficiencies comparing a test based on $G^{\rho_1}_{N_1,N_2}$ with one based on $G^{\rho_2}_{N_1,N_2}$ when the underlying survival function is H_{ρ^*}, where ρ^* may or may not equal one of the ρ_i, $i=1,2$. Before we do this, however, we feel it is instructive to examine the behavior of the H_ρ survival functions under two-sample time-transformed location alternatives.

Suppose that ρ is fixed and that we wish to consider modeling 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 an easy calculation shows that

$$S_2 = S_1 \, [(S_1)^\rho + \{1-(S_1^\rho)\}e^\Delta]^{-1/\rho} \quad . \qquad (2.2)$$

In fact, since the efficiencies of rank tests are invariant under monotonically increasing 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 equation 2.2. The right hand side of equation 2.2 is a specific instance of the conversion function discussed by Peto and Peto (1972). Using equation 2.2, it is not difficult to show that if $\lambda_1(t)$ is the hazard function corresponding to $S_1(t)$, then

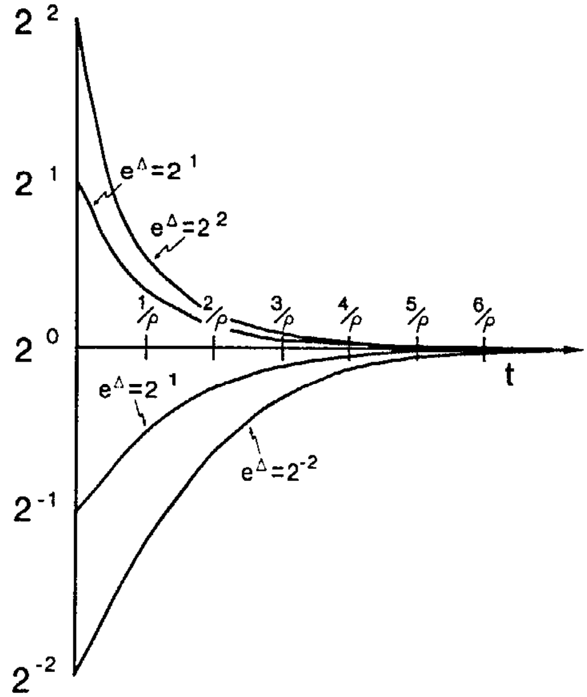
$$\lambda_2 = \lambda_1 \, e^{\Delta} \, [(S_1)^{\rho} + \{1 - (S_1)^{\rho}\}e^{\Delta}]^{-1} \quad .$$

The relative behavior 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 will 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 2.1 illustrates the behavior 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 marked in multiples of ρ^{-1}, since in this setting ρ acts simply as a scale factor.

It is clear from the form of $R(\Delta,\rho t)$ that $\lim_{\rho\to 0} R(\Delta,\rho t) = e^{\Delta}$ for any fixed t, i.e., smaller values of ρ yield alternatives in which the hazards are more nearly proportional. Late in the survival distribution, however, ρ may need to be quite small before $\lambda_2(t)$ becomes "close" to e^{Δ}. Table 2.1 displays values of $R(\Delta,\rho t)$ for selected values of $S_1(t) = e^{-t}$. Notice for instance that for $e^{\Delta}=4$, $R(\Delta, t/2)$ 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 modeling situation



680968 18

Figure 2.1. The Hazard Ratio $R(\Delta,\rho t)$

Table 2.1

Values of $R(\Delta,\rho t) = e^{\Delta}\{e^{-\rho t} + (1-e^{-\rho t})e^{\Delta}\}^{-1}$

(ρ, Δ) \ t	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	3.0
(½, ½)	.500	.525	.550	.574	.599	.622	.646	.668	.690	.711	.731	.818
(½, 2)	2.000	1.826	1.693	1.588	1.504	1.435	1.378	1.330	1.290	1.255	1.225	1.126
(1, ¼)	.250	.289	.332	.378	.426	.475	.525	.579	.623	.668	.711	.870
(1, ½)	.500	.550	.599	.646	.690	.731	.769	.802	.832	.858	.881	.953
(1, 2)	2.000	1.693	1.504	1.378	1.290	1.225	1.177	1.141	1.112	1.090	1.073	1.026
(1, 4)	4.000	2.591	2.011	1.700	1.508	1.381	1.292	1.227	1.178	1.142	1.113	1.039
(2, ½)	.500	.599	.690	.769	.832	.881	.917	.943	.961	.973	.982	.998
(2, 2)	2.000	1.504	1.290	1.177	1.112	1.073	1.048	1.031	1.021	1.014	1.009	1.001
(4, ¼)	.250	.426	.623	.786	.891	.948	.976	.989	.995	.998	.999	1.000
(4, ½)	.500	.690	.832	.917	.961	.872	.992	.996	.998	.999	1.000	1.000
(4, 2)	2.000	1.290	1.112	1.048	1.021	1.009	1.004	1.002	1.001	1.000	1.000	1.000
(4, 4)	4.000	1.508	1.178	1.073	1.032	1.014	1.006	1.003	1.001	1.001	1.000	1.000
(ρ, 1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

leads to a set of location alternatives which are "midway" between proportional hazards 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 tests which provide a balanced compromise between the L-R and the PPW tests.

As mentioned earlier, it is possible to compute Pitman asymptotic relative efficiencies for the family of test statistics $G^{\rho}_{N_1,N_2}$. For simplicity, we will assume that $C_1(t) = C_2(t) = C(t)$, and hence that $\pi_1(t) = \pi_2(t) = \pi(t)$ under H_0. The following argument shows that it is possible to simplify the tabulation of asymptotic relative efficiencies.

In our setting

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

Since

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

we have that

$$\gamma_i(t) = \lim_{N\to\infty}\left(\frac{N_1 N_2}{N_1+N_2}\right)^{\frac{1}{2}}\left\{\frac{d\beta_i^N}{d\beta_1}(t) - 1\right\}$$

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

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

$$= c(-1)^{i+1} a_{3-i}\frac{\partial}{\partial\theta}\ln\lambda(t,\theta)\Big|_{\theta=\theta_0} .$$

Thus we have

$$\gamma(t) = \gamma_1(t) - \gamma_2(t)$$
$$= c \frac{\partial}{\partial\theta} \ell n\, \lambda(t,\theta)\Big|_{\theta=\theta_0} .$$

We will be examining the behavior of the statistic $G^{\rho}_{N_1,N_2}$ under contiguous location alternatives for a distribution H_{ρ^*}, where ρ^* and ρ may or may not be equal. Assume now that c=1. For $S(t,\theta) = H_{\rho^*}\{g(t)+\theta\}$,

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

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 for $G^{\rho}_{N_1,N_2}$ computed at time t for contiguous location alternatives $S^{\rho^*}(t,\theta_i^N) = H_{\rho^*}\{g(t) + \theta_i^N\}$ is given by

$$e_{\rho^*}(\rho,t) = \{\mu_\rho(t) \;/\; \sigma_\rho(t)\}^2$$
$$= \left[\int_0^t \frac{\pi_1(u)\, \pi_2(u)}{a_1\pi_1(u) + a_2\pi_2(u)} \{S(u)\}^{\rho^*} \{S(u)\}^{\rho}\, d\beta(u) \right]^2 \Big/ \int_0^t \frac{\pi_1(u)\, \pi_2(u)}{a_1\pi_1(u) + a_2\pi_2(u)} \{S(u)\}^{2\rho}\, d\beta(u)$$
$$= [\int_0^t \pi(u)\, \{S(u)\}^{\rho^*+\rho}\, d\beta(u)]^2 \;/ \int_0^t \pi(u)\, \{S(u)\}^{2\rho}\, d\beta(u).$$

This expression may be easily evaluated when, for instance, $C(u) = \{S(u)\}^{\alpha}$. (α=0 would imply that the data are uncensored.) In these cases, we have

$$e_{\rho*}(\rho,t) = [-\int_0^t \{S(u)\}^{\rho*+\rho+\alpha} \, dS(u)]^2 \Big/ \int_0^t -\{S(u)\}^{2\rho+\alpha} \, dS(u)$$

$$= \frac{2\rho+\alpha+1}{(\rho*+\rho+\alpha+1)^2} \; \frac{[1 - \{S(t)\}^{\rho*+\rho+\alpha+1}]^2}{[1 - \{S(t)\}^{2\rho+\alpha+1}]} \; .$$

Clearly, when $\rho=\rho*$, we obtain

$$e_{\rho*}(\rho*,t) = \frac{1}{2\rho*+\alpha+1} \; . \; [1 - \{S(t)\}^{2\rho*+\alpha+1}] \; .$$

The ratio $e_{\rho*}(\rho,t)/e_{\rho*}(\rho*,t)$ is the asymptotic relative efficiency of tests based on $G^{\rho}_{N_1,N_2}$ for location alternatives under $H_{\rho*}$ with respect to the fully efficient test $G^{\rho*}_{N_1,N_2}$. Using the results above, we have

$$\frac{e_{\rho*}(\rho,t)}{e_{\rho*}(\rho*,t)} = \frac{(2\rho+\alpha+1)\;(2\rho*+\alpha+1)\;[1-\{S(t)\}^{\rho*+\rho+\alpha+1}]^2}{(\rho*+\rho+\alpha+1)\;[1-\{S(t)\}^{2\rho+\alpha+1}]\;[1-\{S(t)\}^{2\rho*+\alpha+1}]}$$

Obviously $\{e_{\rho*}(\rho,t)/e_{\rho*}(\rho*,t)\} = 1$ when $\rho=\rho*$, as it should. Although one would not anticipate that $C(t) = \{S(t)\}^{\alpha}$ 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. Large values of α could be used as a model for heavily censored data.

The expression given above for $e_{\rho*}(\rho,t)/e_{\rho*}(\rho*,t)$ could be used to plot a.r.e as a function of $S(t)$, the correct null hypothesis survival probability at time t, for selected values of $\rho*$, ρ and α. For the sake of economy, however, we have only provided values of $\lim_{t\to\infty} \{e_{\rho*}(\rho,t)/e_{\rho*}(\rho*,t)\}$ in Table 2.2.

Table 2.2

Asymptotic Relative Efficiencies: $\lim_{t \to \infty} \{e_{\rho^*}(\rho,t)/e_{\rho^*}(\rho^*,t)\}$

	ρ^* / ρ	0.0	1.0	2.0	3.0	4.0
	0.0	1.000	.750	.556	.438	.360
	0.5	.889	.960	.816	.691	.595
α = 0	1.0	.750	1.000	.938	.840	.750
	2.0	.556	.938	1.000	.972	.918
	3.0	.438	.840	.972	1.000	.984
	4.0	.360	.750	.918	.984	1.000
	0.0	1.000	.889	.750	.640	.556
	0.5	.960	.980	.889	.793	.710
α = 1	1.0	.889	1.000	.960	.889	.816
	2.0	.750	.960	1.000	.980	.938
	3.0	.640	.889	.980	1.000	.988
	4.0	.556	.816	.938	.988	1.000
	0.0	1.000	.938	.840	.750	.673
	0.5	.978	.988	.926	.852	.782
α = 2	1.0	.938	1.000	.972	.918	.859
	2.0	.840	.972	1.000	.984	.951
	3 0	.750	.918	.984	1.000	.990
	4.0	.673	.859	.951	.990	1.000

III. Test Statistics for other than Two Samples

Although tests for a significant difference between two homogeneous populations of failure times are commonly used, there are many situations which require a test for a significant difference among r+1 populations, $r \geq 2$, or which call for a goodness-of-fit type test to compare the distribution of a single sample with a hypothesized standard. The G^{ρ} procedures discussed above can easily be generalized to situations involving any number of samples. We will consider first their formulation when there are more than two samples; we will then examine the one sample analogues.

III.1 More Than Two Samples

The easiest way to extend the discussion of the statistics examined here to the setting in which there are more than two samples is in the context of linear rank statistics. Recall that the statistic G^{ρ} was introduced in the form

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

and that such a statistic is asymptotically fully efficient against time-transformed location shifts for survival functions of the form $H_{\rho}(t) = (1+\rho e^{t})^{-1/\rho}$ for two samples of censored data. When the data are uncensored, but the number of samples $r+1 \geq 2$, the following result holds.

<u>Theorem 3.1</u>. Let $S_i(t) = H_{\rho}\{g(t)+\theta_i\}$ $i=1,2,\ldots,r+1$. 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.

The proof of Theorem 3.1 will be provided by the lemmas below.

Lemma 3.2. Let $N = \sum_{i=1}^{r+1} N_i$. Then G^ρ may be written as a linear rank statistic $\sum_{k=1}^{N} c_N^*(k)\vec{z}_k$ with

$$c_N^*(k) = \{\hat{S}(T_k)\}^\rho - \sum_{\ell=1}^{k} \{\hat{S}(T_\ell)\}^\rho / n_\ell$$

Proof: Suppose we have any statistic of the form $\sum_{k=1}^{N} w_k(\vec{z}_k - \vec{q}_k)$ and wish to write it in the form $\sum_{k=1}^{N} c_N^*(k)\vec{z}_k$. Since $(n_{1k}, n_{2k}, \ldots, n_{rk})' = \sum_{\ell=k}^{N} \vec{z}_\ell$ we can write $\vec{q}_k = (n_k)^{-1} \sum_{\ell=k}^{N} \vec{z}_k$. Thus

$$\sum_{k=1}^{N} w_k (\vec{z}_k - \vec{q}_k) = \sum_{k=1}^{N} w_k \vec{z}_k - \sum_{k=1}^{N} \left\{ \frac{w_k}{n_k} \sum_{\ell=k}^{N} \vec{z}_\ell \right\}$$

$$= \sum_{\ell=1}^{N} w_\ell z_\ell - \sum_{\ell=1}^{N} \sum_{k=1}^{\ell} \vec{z}_\ell \frac{w_k}{n_k}$$

$$= \sum_{\ell=1}^{N} \vec{z}_\ell (w_\ell - \sum_{j=1}^{\ell} w_j n_j^{-1}) .$$

For the last expression to equal $\sum_{\ell=1}^{N} c_N^* (\ell) \vec{z}_\ell$ over all sets of regression vectors $\{\vec{z}_\ell : \ell=1,2,\ldots,N\}$ we must have

$$c_N^*(\ell) = w_\ell - \sum_{j=1}^{\ell} w_j n_j^{-1} .$$

The lemma is now proved by taking $w_\ell = \{\hat{S}(T_\ell)\}^\rho$.

Lemma 3.3. Let $\{c_N^*(k);\ k=1,2,\ldots,N\}$ be as above, and let

$$F_\rho(t) = 1 - H_\rho(t) ,$$

$$f_\rho(t) = \frac{d}{dt} F_\rho(t), \qquad -\infty<t<\infty,$$

$$f_\rho'(t) = \frac{d}{dt} f_\rho(t),$$

$$\phi_\rho(u) = - \frac{f_\rho'\{F_\rho^{-1}(u)\}}{f_\rho\{F_\rho^{-1}(u)\}}, \qquad 0<u<1.$$

Let scores $\{c_N(k): k=1,2,\ldots,N\}$ be given by $c_N(k) = -\phi_\rho(\frac{k}{N+1})$. Then tests of H_0 based on the linear rank statistic $\sum_{k=1}^{N} c_N^*(k)\vec{z}_k$ are asymptotically equivalent to the locally most powerful rank tests based on $\sum_{k=1}^{N} c_N(k)\vec{z}_k$.

<u>Proof</u>: In order to prove the asymptotic equivalence of a test based on the scores $c_N(k) = -\phi_\rho(\frac{k}{N+1})$ to that based on another set of scores $c_N^*(k)$ we must show

$$\lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} \{c_N(k) - c_N^*(k)\}^2 = 0.$$

(See, for example, Randles and Wolfe, (1979), pages 287 and 319.)

It is not hard to show that $c_N(k) = c_N^*(k)$ for $\rho=0$, i.e., for the situation in which $H_\rho(t)$ is an extreme value survival function. In what follows, therefore, we will always assume that $\rho>0$. A short calculation shows that $\phi_\rho(u) = -\rho^{-1}\,[(1-u)^\rho(1+\rho) - 1]$. For uncensored data $\hat{S}(T_k) = N^{-1}(N-k+1)$ and $n_k = n-k+1$. If we let

$$a_N(k) = \left(\frac{N+k+1}{N}\right)^\rho - \left(\frac{N-k+1}{N+1}\right)^\rho$$

and

$$b_N(k) = N^{-\rho} \sum_{j=1}^{k} (N-j+1)^{\rho-1} + \rho^{-1} \left\{\left(\frac{N-k+1}{N+1}\right)^{\rho} -1\right\} ,$$

then we may write

$$N^{-1} \sum_{k=1}^{N} \{c_N(k) - c_N^*(k)\}^2 = N^{-1} \sum_{k=1}^{N} \{a_N(k) - b_N(k)\}^2$$

$$= N^{-1} [\sum_{k=1}^{N} \{a_N(k)\}^2 - 2 \sum_{k=1}^{N} a_N(k)b_N(k)$$

$$+ \sum_{k=1}^{N} \{b_N(k)\}^2].$$

It is clear that $|a_N(k)| \leq 1$, and a short calculation shows that $|b_N(k)| \leq 3+\rho^{-1}$. To show that each of the above three terms converges to zero, it is thus sufficient to establish that

$$\lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} a_N(k) = 0$$

and

$$\lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} b_N(k) = 0.$$

First

$$\limsup_{N\to\infty} N^{-1} \sum_{k=1}^{N} a_N(k)$$

$$= \limsup_{N\to\infty} N^{-1} \sum_{k=1}^{N} \left(\frac{N-k+1}{N}\right)^{\rho} \left\{1-\left(\frac{N}{N+1}\right)^{\rho}\right\}$$

$$\leq \limsup_{N\to\infty} \left\{1- \left(\frac{N}{N+1}\right)^{\rho}\right\} = 0.$$

Since $\liminf_{N\to\infty} N^{-1} \sum_{k=1}^{N} a_N(k) \geq 0$, we have that $\lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} a_N(k)$ exists and equals zero.

Second,

$$\lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} b_N(k)$$

$$= \lim_{N\to\infty} [N^{-2} \sum_{k=1}^{N} \sum_{j=1}^{k} \left(\frac{N-j+1}{N}\right)^{\rho-1}$$

$$+ N^{-1} \sum_{k=1}^{N} \rho^{-1} \{ (\frac{N-k+1}{N+1})^{\rho} -1\}]$$

$$= \lim_{N\to\infty} N^{-2} \sum_{k=1}^{N} \sum_{j=1}^{k} (1 - \frac{j-1}{N})^{\rho-1}$$

$$+ \lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} \rho^{-1} \{(1 - \frac{k}{N+1})^{\rho} -1\}$$

$$= \int_{x=0}^{1} \int_{y=0}^{x} (1-y)^{\rho-1}\, dydx + \int_{x=0}^{1} \rho^{-1}\{(1-x)^{\rho} -1\}dx$$

$$= 0.$$

Since Lemma 3.3 shows that tests using the score function $c_N^*(k)$ are asymptotically equivalent to the approximate score function test based on $\phi_\rho(\frac{k}{N+1})$, Theorem 3.1 is established. We note in passing, continuing to consider uncensored data, that it is possible to compute the exact score function for the locally most powerful rank test for location alternatives for the distribution $H_\rho(t)$ when ρ is a positive integer. Such a test is a score function test based on the scores

$$\tilde{c}_N(k) = E\ \{\phi_\rho(U_{(k)})\}\ ,$$

where $U_{(k)}$ is the k^{th} order statistic in a random sample of N variables, each uniformly distributed on [0,1]. In this case, we would have

$$\tilde{c}_N(k) = E\ [-\rho^{-1}[\{1-U_{(k)}\}^\rho\ (1+\rho)\ -\ 1]]$$

$$= \frac{1}{\rho} - (\frac{1+\rho}{\rho})\ E\ \{U_{(N-k+1)}\}^\rho$$

$$= \frac{1}{\rho} - (\frac{1+\rho}{\rho}) \prod_{\ell=0}^{\rho-1} (\frac{N+\rho-k-\ell}{N+\rho-\ell})\ .$$

Since $G^\rho = \sum_{k=1}^{N} c_N^*(k)\vec{z}_k$, with scores $c_N^*(k)$ asymptotically equivalent to those given by $c_N(k) = -\phi_\rho(\frac{k}{N+1})$, we may rely on the large sample distribution theory for $\vec{R} = \sum_{k=1}^{N} c_N(k)\ \vec{z}_k$ to find approximate critical regions for tests of H_0: $\theta_1=\theta_2=\ldots=\theta_{r+1}$. Theorem V.2.2 (p.170) in Hájek and Šidák (1967) establishes that $\vec{R}$, and hence G^ρ, may be used to construct a hypothesis test based upon the χ^2 distribution. Specifically, if we define

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

and

$$\vec{R}_c = \{N_1^{-\frac{1}{2}}\ (R_1-N_1\bar{c}_N),\ N_2^{-\frac{1}{2}}\ (R_2-N_2\bar{c}_N)\ \ldots\ N_r^{-\frac{1}{2}}(R_r\ -\ N_r\bar{c}_N)\}'\ ,$$

where R_i is the i^{th} component of $\vec{R}$, then, one may show that under H_0

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

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. 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 $\vec{z}_{k,j}$, $j=1,2,\ldots m_k$, 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

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

where $\hat{c}_N(k)$ is a score for an uncensored observation, $\hat{C}_N(k)$ is a score for a censored observation, and $\vec{s}_{(k)} = \sum_{j=1}^{m_k} \vec{z}_{kj}$. The scores $\hat{c}_N(k)$ and $\hat{C}_N(k)$ are given by

$$\hat{c}_N(k) = \int_{u_1 < \ldots < u_k} \ldots \int \phi(u_k) \prod_{j=1}^{k} \{n_j (1-u_j)^{m_j} du_j\}$$

and

$$\hat{C}_N(k) = \int_{u_1 < \ldots < u_k} \int \Phi(u_k) \prod_{j=1}^{k} \{n_j (1-u_j)^{m_j} du_j\} ,$$

with

$$\phi(u) = - \frac{f'_\rho\{H_\rho^{-1}(1-u)\}}{f_\rho\{H_\rho^{-1}(1-u)\}}$$

and

$$\Phi(u) = f_\rho\{H_\rho^{-1}(1-u)\} \, (1-u)^{-1} .$$

Straight forward calculations show that

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

and

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

It is not hard to show that $\hat{C}_N(k) = \tilde{C}_N(k)$ for uncensored data.

Hypothesis tests may be based on the censored data rank statstics 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 $\vec{v}$, is clearly specified in Prentice (1978), pp. 170-175. 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 in Hájek and Šidák (1967, p. 152).

III.2. One Sample Test Statistics

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_2 as being arbitrarily large, giving us complete information about a survival distribution S. In this case

$\{Y_1(u)\ Y_2(u)\}/\{Y_1(u) + Y_2(u)\} \approx Y_1(u)$, $\hat{S}(u) \approx S(u)$ and $dN_2(u)/\{Y_2(u)\} \approx d\beta(u)$. Let S_0 be a hypothesized distribution, and $\beta_0 = -\ell n\ S_0$. Notationally, let $N=N_1$, $Y(u) = Y_1(u)$, $X_k^0 = X_{1k}^0$ and $\delta_k = \delta_{1k}$; we may write G^ρ in this situation as

$$G_N^\rho = \int_0^\infty \{S_0(u)\}^\rho \{dN(u) - Y(u)d\beta_0(u)\}$$

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

$$= \sum_{k=1}^N \delta_k \{S_0(X_k^0)\}^\rho - \sum_{k=1}^N \int_0^{X_k^0} \{S_0(u)\}^\rho\ d\beta_0(u).$$

If $\rho>0$, this becomes

$$\sum_{k=1}^N [\delta_k \{S_0(X_k^0)\}^\rho - \frac{1}{\rho} [1-\{S_0(X_k^0)\}^\rho]]$$

while for $\rho=0$, G_N^ρ is

$$\sum_{k=1}^N \{\delta_k - \beta_0(X_k^0)\} = \sum_{k=1}^N [\delta_k - \ell n \{S_0(X_k^0)\}^{-1}].$$

The statistics G_N^ρ can be used to formulate a family of one-sample test statistics which include a one-sample version of the logrank 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 in Gill.

<u>Theorem 3.2.1</u>. Assume that the mild regularity conditions hold which are outlined in Section 4.2 of Gill. Then under H_0: $S=S_0$, the statistic $N^{-\frac{1}{2}}G_N^\rho$ 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

$$\int_0^\infty \{S_0(u)\}^{2\rho}\, Y(u)\, N^{-1}\, d\beta_0(u)$$

$$= N^{-1} \sum_{k=1}^{N} (2\rho)^{-1} [1-\{S_0(X_k^0)\}^{2\rho}] \quad \text{if } \rho>0,$$

$$= -N^{-1} \sum_{k=1}^{N} \ln S_0(X_k^0), \qquad \text{if } \rho=0.$$

It is interesting to note that in uncensored data ($C(u)\equiv 1$), Theorem 3.2.1(a) implies that the statistic

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

has asymptotically a standard normal distribution. Observe also from Theorem 3.2.1 that the one-sample censored data logrank statistic

$$[\sum_{k=1}^{N} \delta_k - \sum_{k=1}^{N} \ln \{S_0(X_k^0)\}^{-1}]^2 \,/\, \sum_{k=1}^{N} \ln \{S_0(X_k^0)\}^{-1}$$

has asymptotically a χ^2 distribution with one degree of freedom. In this setting, $\sum_{i=1}^{N} \delta_i$ is the observed number of deaths and $\sum_{k=1}^{N} \ln \{S_0(X_k^0)\}^{-1}$ is the conditionally expected number of deaths, given the X_k^0.

IV. Monte Carlo Simulations

The asymptotic distributions of the newly proposed test statistics, G^ρ, $\rho\geq 0$, have been used in the construction of hypothesis tests of a given size. Therefore, Monte Carlo simulations were used to confirm that the true size of each of these test procedures, in small or moderate sample sizes and under varying amounts of censorship, was indeed accurately approximated by the nominal significance level based upon this asymptotic distribution theory. The simulations were also used to confirm, in uncensored data, the analytical conclusions reached earlier concerning the role of ρ in determining power. It is those results from the evaluation of power which were particularly interesting and hence will be given in the remainder of this section.

IV.1. 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 since this is where the greatest interest appears to be. 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 $\mathcal{S}^{\rho_i}(\theta_{1i}, \theta_{2i}) \equiv \{S^{\rho_i}(t,\theta_{1i}), S^{\rho_i}(t,\theta_{2i})\}$ for 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 $\{\mathcal{S}^{\rho}(\theta_1, \theta_2); \rho\geq 0\}$ a small sample comparison of the behavior of the logrank and Wilcoxon test statistics (i.e., G^0 and G^1 respectively) with that of other G^{ρ} statistics, the values of ρ_1 chosen were 0, ½, 1 and 2.

Let, for example, the time transformation $g(t) = \ell n t$. 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 $-[\ell n U]/e^{\theta}$ 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 (Knuth, 1969).

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 $\alpha=0.01$ and $\alpha=0.05$ significance levels were calculated.

IV.2 Power Results

Results of the Monte Carlo study pertaining to the evaluation of power of the set of procedures $\{G^\rho: \rho = 0, \frac{1}{2}, 1, 2\}$ are presented in Table 4.1. Figure 4.1 presents the plots of the hazard functions corresponding to the four survival configurations inspected in the tables. The table reveals that the small and moderate sample relative power of these four test procedures is entirely consistent with their large sample a.r.e. given in Table 2.1. In the time-transformed extreme value location alternative (configuration I), and in the time-transformed logistic location alternative (configuration III), G^0(i.e., the logrank) has a.r.e. 1 and 0.75 respectively while G^1(i.e., the Wilcoxon) has a.r.e. 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 4.1 reveals in smaller samples that the loss in power obtained by using $G^{\frac{1}{2}}$ rather than G^1 in 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. Both of these observations also conform to what earlier a.r.e. results indicated. Again, in configuration II we find agreement between our small sample results and a.r.e. calculations. Specifically $G^{\frac{1}{2}}$ is more powerful than G^1

which in turn is more powerful than G^0. Finally, we observe small sample confirmation of the facts that G^0 and G^2 have relatively low power in configurations IV and I respectively.

The research of the first author was supported by a grant from the National Science Foundation; the second author received support from the National Institutes of Health for his work.

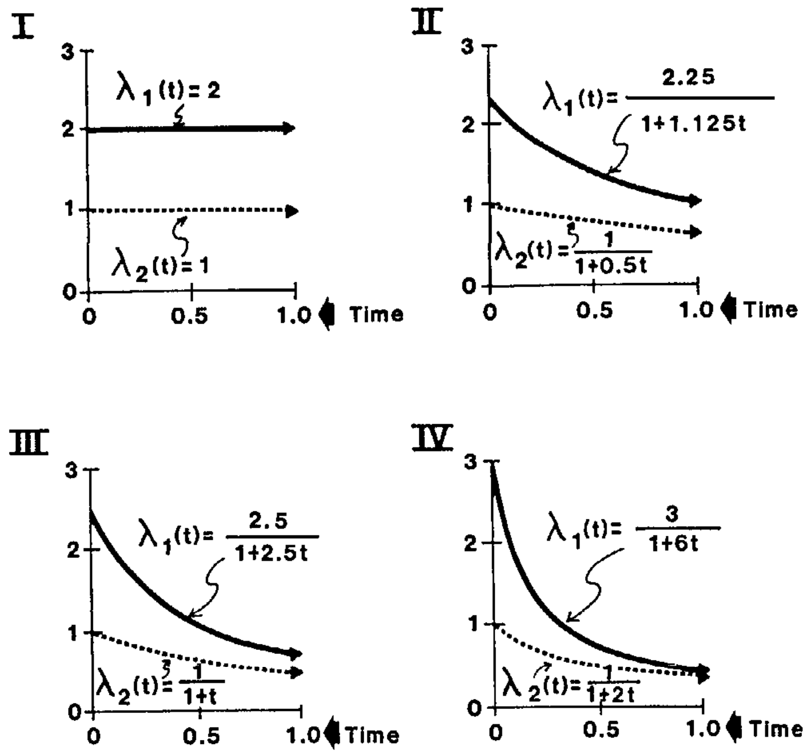


Figure 4.1. Hazard Function Plots for the Alternative Hypothesis Simulation Configurations

Table 4.1

Monte Carlo Estimates of the Power of the G^{ρ}(ρ = 0, ½, 1, 2) One Sided Test Procedures of H_0: $S_1 = S_2$ vs. H_A: $S_1 < S_2$ (500 Simulations)

	S_1* (ρ, e^{θ_1})	S_2 (ρ, e^{θ_2})	$N_1=N_2$	Level of Test	G^0	$G^{1/2}$	G^1	G^2
I.	(0,2)	(0,1)	20	.01	.386	.338	.292	.204
				.05	.668	.620	.578	.456
			50	.01	.858	.800	.734	.610
				.05	.954	.938	.894	.812
II.	(½,2.25)	(½,1)	20	.01	.308	.320	.290	.258
				.05	.548	.576	564	.516
			50	.01	.646	.694	.682	.604
				.05	.844	.878	.868	.830
III.	(1,2.5)	(1,1)	20	.01	.206	.222	.234	.204
				.05	.444	.470	.488	.470
			50	.01	.534	.616	.624	.598
				.05	.754	.834	.864	.828
IV.	(2,3)	(2,1)	20	.01	.148	.186	.202	.206
				.05	.336	.402	.416	.426
			50	.01	.294	.406	.470	.516
				.05	.534	.662	.722	.742

* Each survival distribution is of the form $S^{\rho}(t,\theta_1) = (1+\rho e^{\theta_1} t)^{-1/\rho}$

REFERENCES

Aalen, O. O. (1977). Weak convergence of stochastic integrals related to counting processes. Z. Wahrscheinlichkeitstheorie. Verw. Gebeite 38, 251-77.

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. Annals of Statist. 6, 701-26.

Fleming, T. R. and Harrington, D. P. (1981). A class of hypothesis tests for one and two samples of censored survival data. Communications in Statist. A 10, Number 8.

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

Gill, R. D. (1979). Censoring and Stochastic Integrals. Ph.D. Dissertation, Mathematische Centre, Amsterdam.

Hájek, J. and Šidak, Z. (1967). Theory of Rank Tests. New York: Academic Press.

Knuth, D. E. (1969). The Art of Computer Programming: Volume 2, Seminumerical Algorithms. Reading, Massachusetts: Addison-Wesley.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemotherapy Rep. 50, 163-70.

Peto, R. and 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 parametric models. Biometrika 62, 607-14.

Prentice, R. L. (1978). Linear rank tests with right censored data. Biometrika 65, 167-79.

Prentice, R. L. and Marek, P. (1979). A qualitative discrepancy between censored data rank tests. Biometrics 35, 861-7.

Randles, R. H. and Wolfe, D. A. (1979) Introduction to the Theory of Nonparametric Statistics. New York: John Wiley and Sons, Inc.

Woolson, R. F. (1981). Rank tests for comparing observed and expected survival data. Biometrics. (In press.)