A CLASS OF HYPOTHESIS TESTS FOR ONE AND TWO SAMPLE CENSORED SURVIVAL DATA

by

THOMAS R. FLEMING and DAVID P. HARRINGTON

Technical Report Series, No. 9
August 1980

A CLASS OF HYPOTHESIS TESTS FOR ONE AND TWO SAMPLE CENSORED SURVIVAL DATA

Thomas R. Fleming

Department of Medical Statistics and Epidemiology
Mayo Clinic
Rochester, MN 55901

David P. Harrington

Department of Applied Mathematics and Computer Science
University of Virginia
Charlottesville, VA 22901

ABSTRACT

This paper proposes a class of new non-parametric test statistics useful for goodness-of-fit or two-sample hypothesis testing problems when dealing with randomly right censored survival data. The procedures are especially useful when one desires sensitivity to differences in survival distributions that are particularly evident at at least one point in time. This class is also sufficiently rich to allow certain statistics to be chosen which are very sensitive to survival differences occurring over a specified period of interest. The asymptotic distribution of each test statistic is obtained and then employed in the formulation of the corresponding test procedure. Size and power of the new procedures are evaluated for small

and moderate sample sizes using Monte Carlo simulations. The simulations, generated in the two sample situation, also allow comparisons to be made with the behavior of the Gehan-Wilcoxon and log-rank test procedures.

1. INTRODUCTION AND SUMMARY

Let X be a nonnegative random variable, and suppose X is interpreted as the measured time to some predetermined event. Since one of the most common applications of random variables such as X lies in the area of life-testing and survival theory, X will hereafter be referred to as a time of death or failure. In life-testing applications, problems of statistical inference about failure time random variables most commonly arise in one of the following two situations:

(a) One has a random sample $X_1, X_2, \ldots, X_N$ from a homogeneous population of failure times distributed as X, and wishes to test statistical hypotheses about that distribution.

(b) One has two samples, $X_{11}, X_{12}, \ldots, X_{1N_1}$, and X_{21}, $X_{22}, \ldots, X_{2N_2}$ from two possibly different populations and wishes to test the null hypothesis that the underlying survival distributions are in fact the same.

The primary intent of this paper is to present a class of new nonparametric test statistics, for each of the above two situations, in the setting where the survival time random variables are subject to arbitrary random right censoring.

Motivation for the presentation of this new class of test procedures and necessary notation will be provided in the next section. In section three of this paper, the new class of procedures will be developed for the one sample problem. The large sample distribution of the statistics will be given as well as a qualitative discussion of the power of these procedures to detect certain types of departures from the null hypothesis. A procedure presented by Aalen (1976), similar to a member of the class presented here, will also be examined in section three. Section four contains the development of the procedures applicable in the two sample problem and presents an algorithm for the computation of each statistic. Inspection of the power and size of the two sample procedures through Monte Carlo simulations appears in section five, while the final section of the paper outlines the proof of the asymptotic distributions of both the one- and two-sample test statistics.

2. NOTATION AND MOTIVATION

2.1 Statistical Model

We will first establish some notation. Suppose X_{i1}, X_{i2}, ..., X_{iN_i} is an independent, identically distributed collection of failure time random variables. (When there is only a single sample, all of the following notation will be the same, except that the subscript i will be dropped.) If $F_i(t) = P(X_{ij} < t)$, then $S_i(t) = 1-F_i(t)$ is called the survival function associated with X_{ij}; $\nu_i(t) = -\frac{d}{dt} \ln S_i(t)$ and $\beta_i(t) = \int_0^t \nu_i(s)ds$ are called the hazard and the cumulative hazard functions, respectively. Since the death times are subject to right censorship, there will also be a collection of independent, identically distributed censoring times $Y_{i1}, Y_{i2},\ldots, Y_{iN_i}$ associated with each collection of survival times. Let $G_i(t) = P(Y_{ij} < t) = 1-C_i(t)$, and define $\alpha_i(t) = -\ln C_i(t)$. Observe that we allow C_1 and C_2 to differ.

We assume that the only information obtained about the survival time X_{ij} for each individual is $T_{ij} = \min(X_{ij}, Y_{ij})$ and $\delta_{ij} = I_{[X_{ij} \leq Y_{ij}]}$, where $I_{[E]}$ is 1 if the event E occurs and 0 otherwise. If $\delta_{ij} = 1(0)$ then the j^{th} individual in sample i is said to have an observed death (censorship) at time T_{ij}. The type of statistical dependence between X_{ij} and Y_{ij} always causes difficulties in problems of this sort. We will assume throughout this paper that the following two conditions hold for $i = 1,2$:

$$\frac{-\frac{\partial}{\partial t_1}\pi_i(t_1,t_2)\Big|_{t_1=t_2=t}}{\pi_i(t,t)} = \frac{-\frac{d}{dt}S_i(t)}{S_i(t)} \tag{2.1}$$

whenever $\pi_i(t,t) > 0$, where $\pi_i(t_1,t_2) = P(X_{ij} \geq t_1,\ Y_{ij} \geq t_2)$.

$$\pi_i(t) \equiv \pi_i(t,t) = S_i(t)\ C_i(t). \tag{2.2}$$

Gail (1975) has shown that if a condition analogous to (2.1) also holds for $C_i(t)$, then condition (2.2) automatically holds. We would like to allow discrete as well as continuous censoring distributions, and the methods of Gail's proof do not apply in this case; thus we will assume (2.2) directly. We make the interesting remark here that although condition (2.1) is part of the classical assumptions made to insure identifiability in a competing risks setting, it will also prove to be necessary and sufficient in proving a useful martingale property later on.

2.2 Estimation

To define estimators using this censored survival data, let $N_i(t)$ represent the number of individuals in sample i under observation at time t, prior to deaths or censorships occurring at t. Thus, $N_i(t) \equiv \sum_{j=1}^{N_i} I_{[T_{ij} \geq t]}$. Further $D_i(t)$ and $L_i(t)$ represent the number of deaths and censorships, respectively, in sample i at time t. Finally let $\{\tau_j : j=1,\ldots, c\}$ be the set of c distinct censorship times and $\{T_j : j=1,\ldots, d\}$ be the set of d distinct death times in the pooled sample. Estimation will be based upon the right continuous cumulative hazard function estimators

$$\hat{\beta}_i(t) = \sum_{T_j \leq t} \sum_{k=0}^{D_i(T_j)-1} \{N_i(T_j)-k\}^{-1}$$

and

$$\hat{\alpha}_i(t) = \sum_{\tau_j \leq t} \sum_{k=0}^{L_i(\tau_j)-1} \{N_i(\tau_j) - D_i(\tau_j)-k\}^{-1}$$

where we have adopted the convention of breaking ties between deaths and censorships in sample i by assuming the deaths occurred infinitesimally earlier, and where we define $\sum_{k=0}^{-1} f(k) \equiv 0$ for any f. These estimators were originally studied by Nelson (1969) and later by Aalen (1976) who extensively investigated their distributional properties. $S_i(t)$ and $C_i(t)$ in turn are estimated by $\hat{S}_i(t) = \exp\{-\hat{\beta}_i(t)\}$ and $\hat{C}_i(t) = \exp\{-\hat{\alpha}_i(t)\}$, whose properties have been presented recently by Fleming and Harrington (1979).

Observe that we have assumed that the survival distributions S_i are continuous, and hence hereafter we will consider only the situation in which no ties exist between death times. The test procedures to be proposed here, however, can be applied to data having a small or moderate number of tied death times by using these general forms of $\hat{\beta}_i$ and $\hat{S}_i$.

2.3 Motivation for the New Procedures

We will now provide some motivation for this new class of test procedures. The notation used will continue to be that of the two-sample situation. The null hypothesis to be tested then is H_0: $S_1(t) = S_2(t)$ for $0 \leq t \leq \tau$ where τ is a fixed positive number.

An important type of departure from H_0 that may arise is called the "crossing hazards" alternative. When two underlying survival distributions have hazard functions which cross at some point, then the survival curves will exhibit differences over a time interval, but those differences may disappear outside that interval. The crossing hazards phenomenon can often go undetected by test statistics that depend upon cumulative differences in the hazard functions. The Gehan-Wilcoxon (Gehan, 1965) and the log-rank (Peto and Peto, 1972) statistics are of this type (see Prentice and Marek, 1979). It is reasonable to expect, though, that procedures based upon maximum observed differences (perhaps weighted in some fashion) in empirical survival functions or in empirical cumulative hazard functions might be more likely to detect crossing hazards alternatives or more generally any type of departure from H_0 that is particularly evident at one point in time. Further, if the changes in these observed differences in empirical functions could be more heavily weighted over intervals where one might anticipate that the departures from H_0 will occur, one would obtain even greater sensitivity to detect this type of alternative.

The techniques proposed here attempt to deal with these issues. All the hypothesis test statistics are Kolmogorov-Smirnov-type statistics, that is, they are suprema of appropriately scaled empirical processes; they are, therefore, sensitive to differences in underlying survival or cumulative hazard distributions which are large at a particular point in time but may disappear at other time points. The class of proposed procedures is sufficiently rich to allow particular statistics to be chosen which are very sensitive to differences occurring over a specified period of interest. Since we assume that the data are subject to random right censorship, the procedures obtained will in addition have more general applicability than those proposed by Dufour and Maag (1978), Barr and Davidson (1973), Koziol and Byar (1975), and Schey (1977), all of whom considered only certain restrictive forms of censoring.

Throughout this paper, let τ be a fixed constant satisfying $\pi_i(\tau) > 0$, $i = 1,2$, and set $T = [0,\tau]$. Let d be the Skorohod metric on the space D(T) of functions on the interval $[0,\tau]$ which have discontinuities of only the first kind (cf. Billingsley (1968)). Hereafter, the term "weak convergence" will be used with respect to the product metric d_n on the product space $D^n(T)$ for appropriate values of n, and will be denoted by $\Rightarrow$.

3. ONE SAMPLE GOODNESS-OF-FIT TESTS

3.1 Formulation and Large Sample Properties of the Test Statistic

In this section we will treat only the situation of a single homogeneous population of failure times and the subscript i will be dropped. Thus we will be considering testing certain hypotheses about the underlying probability distribution of a random sample of failure times $X_1, X_2, \ldots, X_N$.

Let S_0 be a specified survival function having continuous hazard function $\nu_0(t) = -\frac{d}{dt} \ln S_0(t)$ and cumulative hazard function $\beta_0(t) = -\ln S_0(t)$. A test of whether $S_0(t)$ is equal to the true survival function, $S(t)$, for $t \in T$, is equivalently a test of the hypothesis $H_0: \beta(t) = \beta_0(t)$ for all $t \in T$. Formulation of such a test could be based upon a process which at time t is the statistic

$$\int_0^t \sqrt{N}\, I_{[N(s)>0]}\, d\{\beta_0(s) - \hat{\beta}(s)\}.$$

Observe that when $N(t) > 0$, the statistic represents the difference between the hypothesized and estimated cumulative hazard functions at t weighted by $\sqrt{N}$. In uncensored data, N is a measure of the amount of survival information available throughout the interval T. Integrals, unless otherwise specified, are Lebesque-Stieltjes.

Effectively, the amount of information available in censored data to estimate the change in the survival or cumulative hazard function at time s is only a fraction C(s) of that available in uncensored data. In view of this fact, and in order to define a process which asymptotically, under H_0, will have a covariance function independent of the censoring distribution, we will replace N by $N\hat{C}(s^-)$ and define the process

$$\mathcal{B}_N^0 \equiv \{\int_0^t \{N\hat{C}(s^-)\}^{1/2}\, I_{[N(s)>0]}\, d\{\beta_0(s) - \hat{\beta}(s)\} : 0 \le t \le \tau\}.$$

Note for any function f(t) we define $f(t^-) = \lim_{s \uparrow t} f(s)$. The left hand limit of the estimator of C(s) is used to be consistent with the convention employed in breaking ties between deaths and censorships.

To allow the investigator a degree of flexibility in defining the class of alternatives to the null hypothesis against which greatest sensitivity will be obtained, we will consider test procedures based upon the processes $B_N^{\alpha} = \{B_N^{\alpha}(t): t\varepsilon T\}$, where α is a non-negative fixed real number and $B_N^{\alpha}(t) =$

$$\int_0^t \frac{1}{2}[\{S_0(s)\}^{\alpha} + \{\hat{S}(s^-)\}^{\alpha}]\{\hat{C}(s^-)\}^{1/2} I_{[N(s)>0]}\, d\sqrt{N}\,\{\beta_0(s) - \hat{\beta}(s)\}.$$

The role of the free parameter α or, more specifically, the role of the factor $\frac{1}{2}[\{S_0(s)\}^{\alpha} + \{\hat{S}(s^-)\}^{\alpha}]$ in determining sensitivity to specific classes of alternatives will be discussed after Theorems 3.1 and 3.2 are given; these theorems contain the basic results needed to formulate Kolmogorov-type goodness-of-fit test procedures.

Theorem 3.1. Suppose conditions (2.1) and (2.2) hold, that $\nu(u)$ is a continuous function on T, and that α is a fixed non-negative number. Then, under H_0, $B_N^{\alpha} \Rightarrow B^{\alpha}$, where

$$B^{\alpha} \equiv \{B^{\alpha}(t) = \int_0^t \{S(s)\}^{\alpha - \frac{1}{2}} \{\nu(s)\}^{1/2}\, dW(s): t\varepsilon T\},$$

and where the integrals are stochastic integrals in the quadratic mean with respect to a standard Wiener process $\{W(s): s \geq 0\}$.

The proof of Theorem 3.1 follows from weak convergence results for stochastic integrals of square integrable martingales; an outline for this proof is given in section 6.

It follows from Theorem 3.1 that B^{α} is a zero mean Gaussian process possessing continuous sample paths, independent increments, and variance function

$$\sigma_{\alpha}^2(t) \equiv \text{var } B^{\alpha}(t) = \int_0^t \{S(s)\}^{2\alpha-1} \nu(s)ds = (2\alpha-1)^{-1} [1-\{S(t)\}^{2\alpha-1}]$$

if $\alpha \neq 1/2$, with $\sigma_{\alpha}^2(t) = \beta(t)$ if $\alpha = 1/2$.

Since $\hat{\beta}(\tau)$ is a strongly consistent estimator of $\beta(\tau)$ (Aalen, 1976) and $\hat{S}(\tau) = e^{-\hat{\beta}(\tau)}$ is a strongly consistent estimator of $S(\tau)$ (Fleming and Harrington, 1979), it follows under H_0 that $\hat{\sigma}_{\alpha}^2(\tau)$ is a strongly consistent estimator of $\sigma_{\alpha}^2(\tau)$, where

$$\hat{\sigma}_\alpha^2(\tau) = \int_0^\tau [\tfrac{1}{2}\{\hat{S}(s^-)\}^\alpha + \tfrac{1}{2}\{S_0(s)\}^\alpha]^2 \{\hat{S}(s^-)\}^{-1} d\hat{\beta}(s).$$

Clearly $\hat{\sigma}_\alpha(\tau) \equiv [\hat{\sigma}_\alpha^2(\tau)]^{1/2}$ is also a strongly consistent estimator of $\sigma_\alpha(\tau) \equiv [\sigma_\alpha^2(\tau)]^{1/2}$. In fact, with this choice of $\hat{\sigma}_\alpha(\tau)$, var $\{B_N^\alpha(\tau)/\hat{\sigma}_\alpha(\tau)\} \approx 1$ whether or not the null hypothesis holds.

We now define the Kolmogorov-type one-sided and two-sided goodness-of-fit test statistics respectively by

$$K_N^{\alpha} \equiv \sup_{0\leq t\leq\tau} \{\hat{\sigma}_\alpha(\tau)\}^{-1} B_N^\alpha(t)$$

and

$$\tilde{K}_N^{\alpha} \equiv \sup_{0\leq t\leq\tau} \{\hat{\sigma}_\alpha(\tau)\}^{-1} |B_N^\alpha(t)|.$$

It follows by Theorem 3.1 and by the strong consistency of $\hat{\sigma}_\alpha(\tau)$ that, under H_0, K_N^α and $\tilde{K}_N^\alpha$ converge in distribution to $K^\alpha \equiv \sup_{0\leq t\leq\tau} \{\sigma_\alpha(\tau)\}^{-1} B^\alpha(t)$ and $\tilde{K}^\alpha \equiv \sup_{0\leq t\leq\tau} \{\sigma_\alpha(\tau)\}^{-1} |B^\alpha(t)|$ respectively. Since K^α has the same distribution as $\sup_{0\leq t\leq 1} W(t)$, part a in Theorem 3.2 below follows from the result quoted in Karlin and Taylor (1975, page 346). Part b in Theorem 3.2 follows from Feller (1971, page 343) after observing $\tilde{K}^\alpha$ and $\sup_{0\leq t\leq 1} |W(t)|$ share the same distribution.

Theorem 3.2. Under H_0, as $N \to \infty$,

$$\text{(a) } P(K_N^\alpha \leq y) \to H(y) \equiv 1 - \frac{2}{\sqrt{2\pi}} \int_y^\infty \exp(-x^2/2) \tag{3.1}$$

$$\text{(b) } P(\tilde{K}_N^\alpha \leq y) \to \tilde{H}(y) \equiv \frac{4}{\pi} \sum_{k=0}^\infty \frac{(-1)^k}{2k+1} \exp\{-\pi^2 (2k+1)^2/8y^2\} \tag{3.2}$$

3.2. The Role of α

We begin our discussion of the role the free parameter α plays in determining sensitivity to specific classes of alternatives to H_0 by momentarily considering a property of the survival function. The differential equation satisfied by a survival distribution and its associated cumulative hazard function may be written as $dS = -Sd\beta$. Thus small changes in S at time u may be thought of as $-S(u)$ times the small change in β at u. If instead we look at $-S^{\alpha}d\beta = S^{\alpha-1}dS$, then small changes in S will be weighted by the factor $S^{\alpha-1}$, causing changes in S when S is small to be emphasized when $0 \le \alpha < 1$, and changes in S when S is large to be emphasized when $\alpha > 1$.

To obtain a similar weighting of changes in the difference $S-S_0$, it would be natural to examine the differential

$$\frac{1}{2}\{S^{\alpha-1} + S_0^{\alpha-1}\}\ d(S-S_0) = \frac{1}{2}\{S^{\alpha-1} + S_0^{\alpha-1}\}\ [S_0 d\beta_0 - S d\beta]. \qquad (3.3)$$

To take advantage of the large sample behavior of integrals with respect to the martingale $\{\int_0^t I_{[N(s)>0]}\ d\{\beta_0(s) - \hat{\beta}(s)\}: t \in T\}$, we will approximate (3.3) by

$$\frac{1}{2}\{\hat{S}^{\alpha}(s^-) + S_0^{\alpha}(s)\}\ I_{[N(s)>0]}\ d\{\beta_0(s) - \hat{\beta}(s)\},$$

where unknown functions are now replaced by their estimators. Hence, incorporating the estimate of the censoring distribution in the fashion described earlier, one is naturally led to statistics based on the empirical process B_N^{α}.

In view of these observations, the statistic K_N^{α} has been designed so that if $0 \le \alpha < 1$, K_N^{α} tends to give greater weight to later rather than earlier changes in the survival difference, $S(t) - S_0(t)$. Thus, for small values of α and in particular for the value $\alpha = 0$, the corresponding Kolmogorov-type goodness-of-fit test procedures are especially sensitive to those substantial departures from H_0 which do not occur until late in time; that is at times $t < \tau$ such that $S(t)$ is close to $S(\tau)$. Hence, for example, these procedures could be useful in detecting long-term survival benefits of aggressive coronary heart disease treatments which induce moderate mortality initially.

On the other hand, for $\alpha > 1$, the corresponding K_N^{α} procedures give particularly heavy weight to earlier rather than later changes in the survival difference and in turn are particularly sensitive to substantial departures from H_0 which occur early in time. This sensitivity is not diminished in situations in which the large survival differences disappear later in time. This type of alternative to H_0 in which large early survival differences disappear later in time has been termed the acceleration alternative and has been given a great deal of attention in clinical trials assessing the carcinogenicity of certain drugs, (Transcripts of the Ad Hoc Statistical Meeting, June 11, 1979, of the Working Group to Evaluate the Carcinogenicity of F D and C Red Dye #40. On file, Hearing Clerk, Food and Drug Administration, Washington, D.C.)

A careful analytical study of the sensitivity, i.e. the power, of these newly proposed procedures under various specific alternatives and censoring distributions would obviously be complex and difficult. The proof of Theorem 3.1 given in section 6 shows that we may write our test statistic for a one-sided test as

$$K_N^\alpha = \sup_{0\leq t\leq\tau} \frac{B_N^\alpha(t)}{\hat{\sigma}_\alpha(\tau)} = \sup_{0\leq t\leq\tau} \{\hat{\sigma}_\alpha(\tau)\}^{-1} \{L_N^\alpha(t) + r_N^\alpha(t)\}$$

where

$$r_N^\alpha(t) = \int_0^t \sqrt{N}\, \frac{1}{2} [\{S_0(u)\}^\alpha + \{\hat{S}(u^-)\}^\alpha]\, \hat{C}(u^-)^{1/2} \cdot I_{[N(u)>0]}\, \{\nu_0(u) - \nu(u)\}\, du,$$

and where $L_N^\alpha(t)$ is a stochastic integral with respect to a martingale. When large values of the test statistic are observed one would reject the null hypothesis H_0: $S(t) = S_0(t)$ for $t\varepsilon T$, in favor of the alternative that $S(t) > S_0(t)$ over some interval. The proof of Theorem 3.1 shows that $\{L_N^\alpha(t)/\hat{\sigma}_\alpha(\tau): t\varepsilon T\}$ always converges weakly to a mean zero independent increment Gaussian process with variance function $\Sigma_\alpha^2(t)/\Sigma_\alpha^2(\tau)$, where

$$\Sigma_\alpha^2(s) = \int_0^s [\tfrac{1}{2} [\{S(u)\}^\alpha + \{S_0(u)\}^\alpha]]^2 \{S(u)\}^{-1}\, \nu(u)\, du, \qquad (3.4)$$

and where $\Sigma_\alpha^2(s) = \sigma_\alpha^2(s)$ whenever either H_0 holds or $\alpha = 0$. While

$r_N^\alpha(t) \equiv 0$ under H_0, it is clear that $\lim_{N\to\infty} r_N^\alpha(t) = \infty$ a.s. whenever $\nu(u)$ and $\nu_0(u)$ do not agree on a subset of $[0,t]$ with positive Lebesque measure. Heuristically we can say then that, when a distribution from the alternative class holds, the test statistic in large samples is formulated using a Gaussian process with variance function $\Sigma_\alpha^2(t)/\Sigma_\alpha^2(\tau)$ and with non-constant mean value function $r^\alpha(t)/\Sigma_\alpha(\tau)$, where $\Sigma_\alpha(\tau) = [\Sigma_\alpha^2(\tau)]^{1/2}$, and where

$$r^\alpha(t) = \sqrt{N} \int_0^t \frac{1}{2} [\{S_0(u)\}^\alpha + \{S(u)\}^\alpha] \, C(u)^{1/2} \, \{\nu_0(u) - \nu(u)\} \, du.$$

Clearly, the larger the mean value function, the more likely it is that the test statistic will lie in the rejection region. It is immediately evident that the mean value functions form an increasing sequence in N and are larger for large positive values of $\nu_0(u) - \nu(u)$ on the interval $T = [0,\tau]$.

The role of α in determining a procedure's sensitivity in detecting early or late differences has already been indicated. However, it should be added that the role of the censorship distribution in the power of these test procedures is also very important and is precisely shown in the expression of the mean value function, $r^\alpha(t)/\Sigma_\alpha(\tau)$. Clearly even if one attempts to obtain sensitivity to detect substantial differences occurring later in time, little power will be realized if censorship at that time is heavy. This can be seen mathematically by observing in the formula for $r^\alpha(t)$ that the difference in hazard functions is weighted by the product of $\frac{1}{2} [\{S_0(u)\}^\alpha + \{S(u)\}^\alpha]$ and $C(u)^{1/2}$.

3.3. Another Proposed Kolmogorov-Type Procedure

In comparison to the classical Kolmogorov-Smirnov procedures in uncensored data, it should be noted that the procedure using the test statistic K_N^0 bases rejection of H_0 precisely on whether differences between the cumulative hazard functions, rather than the survival functions, exceed a certain quantity at any point in time. This basic idea was proposed earlier by Aalen (1976) who suggested use of the statistic

$$Q_N \equiv \sup_{0\le t\le\tau} Q_N(t) \equiv \sup_{0\le t<\tau} \{\hat{\sigma}_Q^2(\tau)\}^{-1/2} \sqrt{N} \int_0^t I_{[N(s)>0]}\, d\{\beta_o(s) - \hat{\beta}(s)\}$$

to test H_0, where $\hat{\sigma}_Q^2(\tau)$ is a strongly consistent estimator of

$$\sigma_Q^2(\tau) = \int_0^\tau C(s)^{-1}\, S(s)^{-1}\, d\beta(s).$$

The statistic Q_N is essentially equivalent to K_N^0 is uncensored data. However in censored data, unlike K_N^0, Q_N is formulated by taking the supremum over a process whose large sample covariance function under H_0 is dependent upon the censoring distribution. Specifically, $\{Q_N(t): 0\le t\le\tau\} \Rightarrow \{Q(t): 0\le t\le\tau\}$ where the limiting process is a mean zero Gaussian process with independent increments and variance function $\sigma_Q^2(t)/\sigma_Q^2(\tau)$, clearly dependent upon $C(t)$.

It is of particular interest to compare the Kolmogorov-type goodness-of-fit statistics K_N^0 and Q_N when the alternative hypothesis holds. Using terminology of the previous sub-section, in large samples, K_N^0 is formulated from approximately a Gaussian process with independent increments, having variance function

$$\sigma_0^2(t)/\sigma_0^2(\tau) = \int_0^t S(s)^{-1}\, d\beta(s) \Big/ \int_0^\tau S(s)^{-1}\, d\beta(s)$$

and mean value function

$$r^0(t)/\sigma_0(\tau) = \sqrt{N} \int_0^t C(s)^{1/2}\, d\{\beta_o(s) - \beta(s)\} \Big/ \{\int_0^\tau S(s)^{-1}\, d\beta(s)\}^{1/2}.$$

In comparison Q_N is formulated from a Gaussian process with independent increments, having variance function

$$\sigma_Q^2(t)/\sigma_Q^2(\tau) = \int_0^t C(s)^{-1}\, S(s)^{-1}\, d\beta(s) \Big/ \int_0^\tau C(s)^{-1}\, S(s)^{-1}\, d\beta(s)$$

and mean value function

$$r^Q(t)/\sigma_Q(\tau) = \sqrt{N} \int_0^t d\{\beta_o(s) - \beta(s)\} \Big/ \{\int_0^\tau C(s)^{-1}\, S(s)^{-1}\, d\beta(s)\}^{1/2}.$$

Careful inspection of the variance and mean value functions of both processes reveals that, unlike K_N^0, Q_N has the undesirable property that its probability of rejection of H_0 based upon information up to time t systematically tends to zero when censorship of data after time t is increased.

4. TWO SAMPLE PROCEDURES

4.1. The Test Statistic and its Asymptotic Distribution

Suppose now one has two independent samples of death and censoring time random variables, $(X_{11}, Y_{11}), \ldots, (X_{1N_1}, Y_{1N_1})$ and $(X_{21}, Y_{21}), \ldots, (X_{2N_2}, Y_{2N_2})$. Let $S_i(t) = P(X_{ij} \geq t)$, $i = 1,2$. One can formulate a class of procedures to test $H_0: S_1(t) = S_2(t)$ (= S(t) unspecified) for $t \varepsilon T$, which is similar to that proposed for the one sample goodness-of-fit situation. To define the appropriate test statistics and specify their asymptotic distributions, we need some notation. Let

$$H_{N_1,N_2}(s) = \left\{\frac{N_1\hat{C}_1(s^-)\; N_2\hat{C}_2(s^-)}{N_1\hat{C}_1(s^-) + N_2\hat{C}_2(s^-)}\right\}^{1/2} \frac{1}{2} \cdot [\{\hat{S}_1(s^-)\}^{\alpha} + \{\hat{S}_2(s^-)\}^{\alpha}],$$

$$B^{\alpha}_{N_1,N_2}(t) = \int_0^t H_{N_1,N_2}(u)\; I_{[N_1(u)\; N_2(u) > 0]}\; d\{\hat{\beta}_1(u) - \hat{\beta}_2(u)\},$$

and

$$\mathcal{B}^{\alpha}_{N_1,N_2} = \{B^{\alpha}_{N_1,N_2}(t);\; 0 \leq t \leq \tau\}.$$

$B^{\alpha}_{N_1,N_2}(t)$ is then entirely analogous to $B^{\alpha}_N(t)$ and will be used to test the two sample null hypothesis in exactly the same manner as $B^{\alpha}_N(t)$ was used in the one sample case. The next theorem contains the basic result needed to formulate these Smirnov-type two sample test procedures.

Theorem 4.1. Suppose that conditions (2.1) and (2.2) hold for $i = 1,2$. Then, under H_0, $\mathcal{B}^{\alpha}_{N_1,N_2} \Rightarrow \mathcal{B}^{\alpha}$ as N_1 and $N_2 \to \infty$, so long as $\lim\limits_{N_1 \to \infty} \frac{N_1}{N_2} = \lambda$, $0 < \lambda < \infty$.

The proof of the theorem, outlined in section 6, uses the multivariate version of Rebolledo's theorem, Theorem 3.5 of Rebolledo (1978).

As in the one sample situation, before giving the test statistic, we need to estimate $\sigma_\alpha^2(\tau) \equiv \text{var}\{B^\alpha(\tau)\}$. Again relying essentially upon the strong consistency of the cumulative hazard function estimators, it follows that

$$\{\hat{\hat{\sigma}}_\alpha^2(\tau)\} \equiv \int_0^\tau \{N_1\hat{C}_1(s^-)+N_2\hat{C}_2(s^-)\}^{-1}\,[\tfrac{1}{2}\,[\{\hat{S}_1(s^-)\}^\alpha + \{\hat{S}_2(s^-)\}^\alpha]]^2 .$$

$$I_{[N_1(s)N_2(s)>0]}\cdot[N_2\hat{C}_2(s^-)\{\hat{S}_1(s^-)\}^{-1}d\hat{B}_1(s)+N_1\hat{C}_1(s^-)\{\hat{S}_2(s^-)\}^{-1}d\hat{B}_2(s)]$$

and $\hat{\hat{\sigma}}_\alpha(\tau) \equiv \{\hat{\hat{\sigma}}_\alpha^2(\tau)\}^{1/2}$ are strongly consistent estimators of $\sigma_\alpha^2(\tau)$ and $\sigma_\alpha(\tau)$ respectively. In fact, with this choice of $\hat{\hat{\sigma}}_\alpha(\tau)$, $\text{var}\{B^\alpha_{N_1,N_2}(\tau)\,/\,\hat{\hat{\sigma}}_\alpha(\tau)\} \approx 1$ whether or not the null hypothesis holds. We now define the Smirnov-type one-sided and two-sided two-sample test statistics respectively by

$$K^\alpha_{N_1,N_2} \equiv \{\hat{\hat{\sigma}}_\alpha(\tau)\}^{-1} \sup_{0\le t\le \tau} B^\alpha_{N_1,N_2}(t)$$

and

$$\tilde{K}^\alpha_{N_1,N_2} \equiv \{\hat{\hat{\sigma}}_\alpha(\tau)\}^{-1} \sup_{0\le t\le \tau} \left|B^\alpha_{N_1,N_2}(t)\right|$$

Using the strong consistency property of $\hat{\hat{\sigma}}_\alpha(\tau)$ and the weak convergence result of Theorem 4.1, the large sample distributions of the newly proposed statistics are obtained by employing the identical argument presented in the one-sample situation. This result is given in the lemma below.

<u>Lemma 4.1</u>. Under $H_0:S_1(t) = S_2(t)$ for $t\varepsilon T$, as $N_1,N_2 \to \infty$,

$$P(K^\alpha_{N_1,N_2} \le y) \to H(y) \text{ and } P(\tilde{K}^\alpha_{N_1,N_2} \le y) \to \tilde{H}(y),$$

where $H(y)$ and $\tilde{H}(y)$ were defined in (3.1) and (3.2) respectively.

4.2. Computational Formulas

Although the integral formulas provide the easiest way to formulate the newly proposed test statistics, they do not lend much insight into how the statistic might actually be calculated with a given set of data. We will now give the algorithm that can be used to calculate the one- and two-sided Smirnov-type two-sample tests based upon $\{\hat{\hat{\sigma}}_\alpha(T^*)\}^{-1} \sup_{0\leq t\leq T^*} B^\alpha_{N_1,N_2}(t)$ and $\{\hat{\hat{\sigma}}_\alpha(T^*)\}^{-1} \sup_{0\leq t\leq T^*} |B^\alpha_{N_1,N_2}(t)|$ where $T^* = \max\{T_{ij}: N_1(T_{ij})N_2(T_{ij})>0\}$. (In general, one might set $T' = \min(\tau, T^*)$ for some predetermined τ, and use T' rather than T^* hereafter. This would be necessary for asymptotic results to apply directly.) The algorithm reveals that the computation is more straightforward than one might have anticipated, primarily due to the recursive structure of the calculation.

Recall $N_i(t)$ represents the number of individuals in sample i under observation at time t, prior to any deaths or censorships at time t, and $\{T_j: j=1,\ldots,d\}$ represents the d distinct times of death in the pooled sample. Let $Z_j \varepsilon \{1,2\}$ indicate the sample in which the death at T_j occurred.

(1) Define $J = \max\{j: T_j \leq T^*\}$.

For i=1,2, set $\hat{\beta}_i(T_0) = 0$ and recursively calculate, for all j=1,...,J,

$$\hat{\alpha}_i(T_j^-) = \sum_{k=0}^{N_i-N_i(T_j)-1} (N_i-k)^{-1} - \hat{\beta}_i(T_{j-1})$$

and

$$\hat{\beta}_i(T_j) = \hat{\beta}_i(T_{j-1}) + \{1-|i-Z_j|\}\{N_i(T_j)\}^{-1}. \quad (4.1)$$

(Observe $\hat{\beta}_i(T_j^-) = \hat{\beta}_i(T_{j-1})$ since the estimator $\hat{\beta}_i$ as defined is right continuous.)

In (2)-(4), we calculate $B^\alpha_{N_1,N_2}(T_j)$ for j = 1,...,J.

(2) For all j=1,...,J calculate

$$\eta(T_j) = [[N_1 \exp\{-\hat{\alpha}_1(T_j^-)\}]^{-1} + [N_2 \exp\{-\hat{\alpha}_2(T_j^-)\}]^{-1}]^{-1/2}.$$

(3) For all j=1,...,J calculate

$$V_\alpha(T_j) = 0.5\,[\exp\{-\alpha\hat{\beta}_1(T_j^-)\} + \exp\{-\alpha\hat{\beta}_2(T_j^-)\}].$$

(4) Set $U(T_0) = 0$ and recursively calculate, for all $j=1,\ldots,J$,

$$U(T_j) = U(T_{j-1}) + \eta(T_j)V_\alpha(T_j)[\{1-|1-Z_j|\}\{N_1(T_j)\}^{-1} - \{1-|2-Z_j|\}\{N_2(T_j)\}^{-1}]. \qquad (4.2)$$

In (5) and (6) we calculate $\hat{\hat{\sigma}}_\alpha(T^*)$.

(5) For $j=1,\ldots,J$ and $i=1,2$, calculate

$$\xi_i(T_j) = [\exp\{\hat{\beta}_i(T_j^-)\}]\ \{1-|i-Z_j|\}\cdot [N_i(T_j)\ N_i\ \exp\{-\hat{\alpha}_i(T_j^-)\}]^{-1}. \qquad (4.3)$$

(6) Set $\hat{\hat{\sigma}}_\alpha(T^*) = [\sum_{j=1}^{J} \{\eta(T_j)V_\alpha(T_j)\}^2\{\xi_1(T_j) + \xi_2(T_j)\}]^{1/2}$.

(7) Set $K^\alpha_{N_1,N_2} = \max\{0, \{\hat{\hat{\sigma}}_\alpha(T^*)\}^{-1}U(T_j)\colon j=1,\ldots,J\}$

and set the two-sided statistic

$$\tilde{K}^\alpha_{N_1,N_2} = \max\{\{\hat{\hat{\sigma}}_\alpha(T^*)\}^{-1}|U(T_j)|\colon j=1,\ldots,J\}.$$

(8) Calculate the significance level for the one-sided test from the formula

$$p = p(K^\alpha_{N_1,N_2}) = \frac{2}{\sqrt{2\pi}} \int_{K^\alpha_{N_1,N_2}}^{\infty} \exp[-x^2/2].$$

Calculate the significance level for the two-sided test from the formula

$$\tilde{p} = \tilde{p}(\tilde{K}^\alpha_{N_1,N_2}) = 1 - \frac{4}{\pi}\sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp[-\pi^2(2k+1)^2/\{8(\tilde{K}^\alpha_{N_1,N_2})^2\}].$$

Observe that the algorithm has been presented with the assumption that no ties exist between observed times of death. If some ties in death times do exist, the one and two sample procedures can be applied in that situation by simply replacing the term $\{1-|i-Z_j|\}\{N_i(T_j)\}^{-1}$ in (4.1), (4.2) and (4.3) of the algorithm by the term $\sum_{k=0}^{D_i(T_j)-1} \{N_i(T_j) - k\}^{-1}$ for $i=1,2$. This modification to accomodate tied death times is desirable due to its simplicity. However, the test statistic's distribution was actually derived assuming no ties exist in death times. Thus, if the number of ties is very large, other more complex modifications might yield improved behavior.

5. MONTE CARLO SIMULATIONS

The asymptotic distributions of the newly proposed test procedures have been used in the construction of hypothesis tests of a given size. Therefore, Monte Carlo simulations were used to determine whether the true size of each of these test procedures, in small or moderate sample sizes and under varying amounts of censorship, was accurately approximated by the nominal significance level based on this asymptotic distribution theory. The simulations were also used to confirm the conclusions reached earlier concerning the role of α in determining power. Nearly exact values of the power of $K^{\alpha}_{N_1,N_2}$, for selected values of α, were calculated against some specific alternatives of particular interest. Attention in this section has been restricted to the two sample problem since this is where the greatest interest appears to be.

5.1 Simulation Procedure

In the simulations, eleven distinct configurations of survival and censoring distributions were inspected, with each configuration including two survival distributions used to generate the two samples of failure times, and a single censoring distribution used to generate the two samples of censoring times. Independent censoring and survival random variables were obtained by transforming uniformly distributed random variates produced with a linear congruential random number generator (Knuth, 1969). Each observation time was taken to be the minimum of a survival and a censoring random variable; that is, $T_{ij} = \min(X_{ij}, Y_{ij})$ as indicated earlier. Equal sample sizes of $N_1 = N_2 = 20$ and $N_1 = N_2 = 50$ were used.

To investigate size, the two survival distributions were chosen to be equal. Various survival configurations in which the null hypothesis failed to hold were then simulated with the intent of comparing, in these special cases, the power of the log-rank, the Gehan-Wilcoxon, and the Smirnov-type procedures $K^{\alpha}_{N_1,N_2}$ for $\alpha = 0,1,2,3$ and 4. Most configurations chosen had survival differences which were particularly evident at one point in time since it was this type of departure from H_0 that provided the basic motivation for the formulation of the $K^{\alpha}_{N_1,N_2}$ procedures.

Five hundred pairs of samples (one thousand pairs of samples when evaluating size) were generated for each selected configuration of survival and censoring distributions for the two populations and for each sample size. The proportions of samples in which each one-sided test procedure under consideration rejected H_0 at the $\alpha = 0.05$ significance level were calculated.

5.2 Results

Results pertaining to the evaluation of size of the $K^{\alpha}_{N_1,N_2}$ procedures are presented in Table I. Selected configurations of censoring distributions and equal exponential survival distributions were inspected which yielded lightly, moderately or heavily censored data; specifically, the expected percents censored were 13%, 25%, 37%, 47%, 61% or 68%. All Smirnov-type procedures preserved the nominal size in each situation, with the procedures being somewhat more conservative in the smaller sample size and with the $K^{0}_{N_1,N_2}$ procedure being noticeably more conservative than others.

Table II contains the results obtained from the Monte Carlo simulations employed to evaluate the power of the newly proposed $K^{\alpha}_{N_1,N_2}$ procedures. Figure 1 presents the graphs of the five configurations inspected in the table.

The first configuration (see Figure 1(a)) presents a "proportional hazards" or "Lehmann" alternative. Specifically, two exponential distributions representing a doubling in median survival were generated. This configuration was chosen to compare the behavior of the Smirnov-type procedures to that of the log-rank in the situation in which the latter test procedure would be expected to have its greatest relative sensitivity (see Peto and Peto (1972)). Table II reveals the log-rank procedure is the most sensitive in this situation but its gain in power is not large over that of $K^1_{N_1,N_2}$, the most powerful of the $K^\alpha_{N_1,N_2}$ procedures at this alternative for the values of α examined.

TABLE I SIZE

Monte Carlo Estimates of the Sizes of the Gehan-Wilcoxon, Log-rank and $K^\alpha_{N_1,N_2}$ (α = 0,1,2,3 & 4) One-Sided Test Procedures of $H_0:S_1 = S_2$ vs. $H_1:S_1 < S_2$ (1000 simulations)

$S_1 = S_2$	$C_1 = C_2$	Expected percent censored	$N_1 = N_2$	Smirnov-type, $K^\alpha_{N_1,N_2}$ $\alpha = 0$	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$	Gehan-Wilcoxon	Log-rank
$\lambda = 2^*$	C_A**	13.5%	20	.029	.029	.035	.038	.047	.042	.043
			50	.028	.039	.042	.046	.051	.047	.046
	C_B***	25.4%	20	.020	.033	.037	.042	.042	.056	.045
			50	.036	.048	.051	.047	.040	.054	.050
$\lambda = 1$	C_A	36.8%	20	.033	.033	.033	.033	.038	.058	.060
			50	.035	.035	.034	.035	.033	.033	.040
	C_B	47.4%	20	.023	.034	.035	.030	.033	.048	.049
			50	.035	.043	.045	.045	.036	.050	.054
$\lambda = 0.5$	C_A	60.7%	20	.036	.042	.043	.044	.042	.061	.064
			50	.042	.040	.036	.038	.038	.046	.049
	C_B	67.9%	20	.030	.031	.033	.035	.033	.052	.046
			50	.030	.032	.033	.034	.035	.038	.036

* λ denotes the constant hazard function of S_i, an exponential distribution.

** C_A is defined by: $C_A(t) = 1$ for $0 \le t \le 1$, and $C_A(t) = 0$ for $t > 1$.

*** C_B is defined by: $C_B(t) = 1 - (.4)t$ for $0 \le t \le 1$, and $C_B(t) = 0$ for $t > 1$.

The second and third configurations (see Figures 1(b) and 1(c)) present departures from the null hypothesis in which substantial differences existing between survival distributions later in time fail to exist early in time. As would be anticipated, the log-rank has marginally acceptable power against these alternatives, far better than the unacceptable power of the Gehan-Wilcoxon procedure. In turn, however, the procedure $K^{0}_{N_1,N_2}$ has power clearly better than that of the log-rank test. The power of the $K^{\alpha}_{N_1,N_2}$ procedures to detect these later differences depends dramatically upon the choice of α, clearly confirming previous qualitative conclusions that ability to detect departures from H_0 which occur later in time increases as α decreases.

In the fourth configuration (see Figure 1(d)), large differences exist between survival curves over the middle range of the survival distribution although $S_1 = S_2$ for both small t and large t. The last configuration (see Figure 1(e)) presents the situation in which large early differences between survival curves disappear somewhat later in time. From the formulation of their test statistics, we would anticipate the log-rank procedure to have unacceptable sensitivity to these departures, while the Gehan-Wilcoxon should have marginally acceptable power. This has been confirmed in the simulations. Of all procedures inspected, $K^{1}_{N_1,N_2}$ and $K^{2}_{N_1,N_2}$ were found to have the best power to detect the large "middle" difference in Figure 1(d). Simulations with the last configuration clearly reveal that the $K^{\alpha}_{N_1,N_2}$ procedures for $\alpha \geq 2$ have excellent sensitivity in detecting large early survival differences which disappear somewhat later in time. Further, the simulations confirm earlier qualitative conclusions that this sensitivity to early departures from H_0 is increased by choosing $K^{\alpha}_{N_1,N_2}$ procedures with larger α values.

TABLE II POWER

Monte Carlo Estimates of the Power of the Gehan-Wilcoxon, Log-rank and $K^{\alpha}_{N_1,N_2}$ (α = 0,1,2,3 & 4) One-Sided Test Procedures of $H_0: S_1 = S_2$ vs. $H_1: S_1 < S_2$ (500 simulations)*

S_1	S_2	$N_1=N_2$	Smirnov-type $K^{\alpha}_{N_1,N_2}$ $\alpha=0$	$\alpha=1$	$\alpha=2$	$\alpha=3$	$\alpha=4$	Gehan-Wilcoxon	Log-rank
$\lambda_1=2$,** $\lambda_2=1: t\varepsilon(0,\infty)$		20	.392	.446	.426	.372	.354	.478	.522
Fig. 1a "Proportional Hazards"		50	.752	.832	.784	.700	.618	.820	.872
W(4,1)***	W(2,0.5)	20	.450	.304	.154	.068	.034	.130	.358
Fig. 1b "Late Difference"		50	.932	.710	.280	.090	.014	.188	.692
$\lambda_1=2$, $\lambda_2=2: t\varepsilon(0,.4)$		20	.552	.364	.196	.100	.054	.148	.416
$\lambda_1=4$, $\lambda_2=.4: t\varepsilon(.4,\infty)$		50	.948	.724	.348	.192	.098	.254	.688
Fig. 1c "Late Difference"									
$\lambda_1=2$, $\lambda_2=2: t\varepsilon(0,.1)$									
$\lambda_1=3$, $\lambda_2=.75: t\varepsilon(.1,.4)$		20	.072	.280	.294	.250	.194	.274	.178
$\lambda_1=.75$, $\lambda_2=3: t\varepsilon(.4,.7)$									
$\lambda_1=1$, $\lambda_2=1: t\varepsilon(.7,\infty)$		50	.304	.674	.666	.530	.432	.504	.330
Fig. 1d "Middle Difference"									
$\lambda_1=3$, $\lambda_2=.75: t\varepsilon(0,.2)$									
$\lambda_1=.75$, $\lambda_2=3: t\varepsilon(.2,.4)$		20	.052	.234	.402	.498	.534	.322	.162
$\lambda_1=1$, $\lambda_2=1: t\varepsilon(.4,\infty)$		50	.120	.646	.842	.894	.896	.522	.214
Fig. 1e "Early Difference"									

* Censoring distribution in every case is C_B, defined in Table I.

** λ_i denotes the hazard function of S_i, a piecewise exponential distribution.

*** $W(\lambda,\gamma)$ denotes a Weibull distribution having $S(t) = \exp\{-(\lambda t)^{\gamma}\}$.

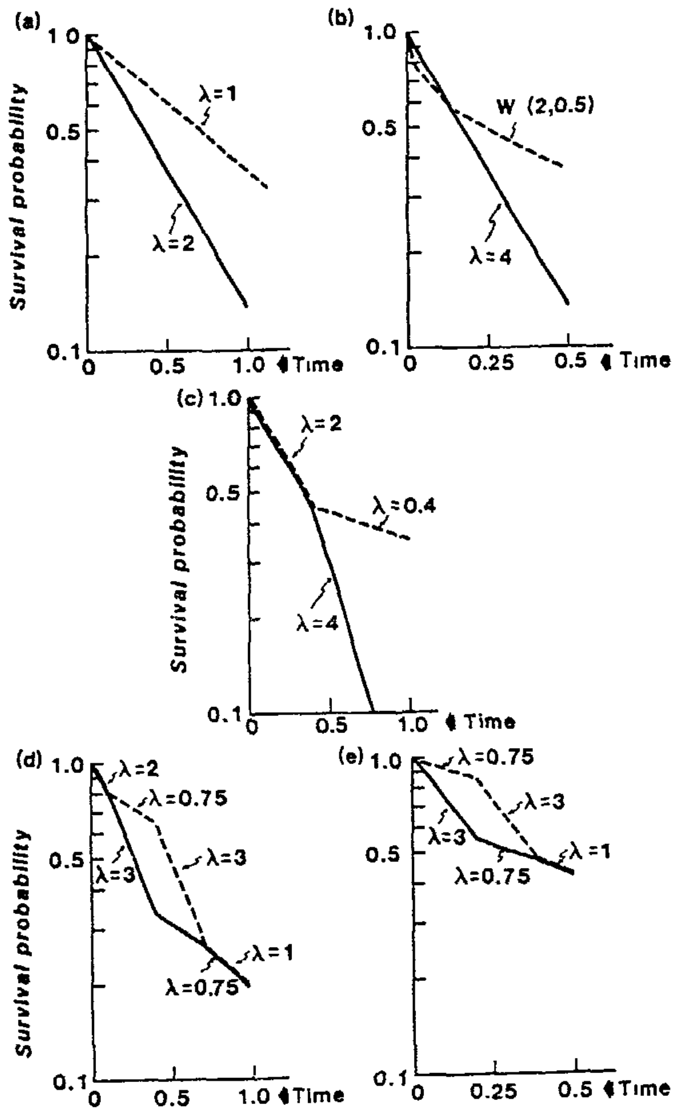

FIG 1. Semi-log plots of survival distributions simulated when evaluating power. Survival distributions are either piecewise exponentials with constant hazard λ over intervals or are Weibull distributions, $W(\lambda,\gamma)$, having survival function $S(t) = \exp\{-(\lambda t)^{\gamma}\}$.

6. ASYMPTOTIC DISTRIBUTION RESULTS

6.1 One Sample Test Statistics

As mentioned in the text, the proof of Theorem 3.1 and of its multivariate analogue, Theorem 4.1, relies on weak convergence results for stochastic integrals of martingales related to counting processes. These results were first published by Aalen (1977); Rebolledo (1978) has shown that some of Aalen's regularity conditions may be simplified considerably. Since the proofs of Theorems 3.1 and 4.1 do not use techniques substantially different from those which have appeared elsewhere in the literature (cf. section 3 in Aalen (1977)), we will only outline the major steps in the proof here. We will first indicate how Theorem 3.1 is proved.

In order to use the results of Aalen and Rebolledo, we need to show that $B_N^{\alpha}(t) = L_N^{\alpha}(t) + r_N^{\alpha}(t)$, where $L_N^{\alpha}(t)$ may be written as a stochastic integral with respect to a counting process based martingale $\overline{M}(t)$, and where,under H_0, $r_N^{\alpha}(t)$ is a remainder term that satisfies $r_N^{\alpha}(t) \Rightarrow 0$ as $N \to \infty$. Lemma 6.1 below is the first step in the construction of $\overline{M}(t)$. Since this lemma provides a connection between the martingale based approach to problems of this type and the classical notion of the identifiability of competing risks, the proof of Lemma 6.1 is given here.

<u>Lemma 6.1</u>. Let X and Y be random variables, defined on an underlying probability space $(\Omega,\mathcal{F},P)$, denoting the death and censoring times for a given individual. Let

$$T_0 = \min (X,Y);$$

$$N^D(t) = I_{[T_0 \le t,\ X \le Y]};$$

$$N^C(t) = I_{[T_0 \le t,\ Y < X]};$$

and

$$\mathcal{F}_t = \sigma\,(N^D(u),\ N^C(u),\ 0 \le u \le t), \quad 0 \le t \le \tau,$$

where $\sigma(\mathcal{A})$ denotes the σ-subalgebra generated by the family of random variables $\mathcal{A}$. Then the stochastic process

$$M(t) = N^D(t) - \int_0^t \nu(u)\, I_{[T_0 \geq u]}\, du, \quad t \varepsilon T,$$

is a square integrable martingale with respect to F_t if and only if condition (2.1) holds.

Proof: $M(t)$ is obviously adapted to F_t, and is clearly square integrable. To see the martingale property, note that

$$E\{M(t+a)|F_t\} = E\{N^D(t+a) - \int_0^{t+a} \nu(u)\, I_{[T_0 \geq u]}\, du|F_t\}$$

$$= N^D(t) + E\,\{N^D(t+a) - N^D(t)|F_t\} - \int_0^t \nu(u)\, I_{[T_0 \geq u]}\, du$$

$$- E\left\{\int_t^{t+a} \nu(u)\, I_{[T_0 \geq u]}\, du|F_t\right\}.$$

To complete the proof, we need only show

$$E\{N^D(t+a) - N^D(t)|F_t\} = \int_t^{t+a} \nu(u)\, P(T_0 \geq u|F_t)\, du \quad \text{a.s.} \,. \qquad (6.1)$$

Let $A = \{\omega\colon N^D(t) + N^C(t) = 1\}$. Then $A \varepsilon F_t$ and both sides of (6.1) are zero on A.

Now examine $A^C \varepsilon F_t$, where A^C is the complement of the event A. On A^C, $N^D(t+a) - N^D(t) = 1$ or 0. Thus

$$\begin{aligned} E\{N^D(t+a) - N^D(t)|A^C\} &= P\{N^D(t+a) - N^D(t) = 1|A^C\} \\ &= P(t < X \leq t+a,\ X \leq Y|A^C) \\ &= P(t < X \leq t+a,\ X \leq Y)/P(X > t,\ Y > t). \end{aligned}$$

On A^C, the right hand side of (6.1) is

$$\int_t^{t+a} \nu(u)\, P(T_0 \geq u|T_0 > t)\, du$$

$$= \{P(T_0 > t)\}^{-1} \int_t^{t+a} \nu(u)\, P(T_0 \geq u)\, du.$$

Since $\{\omega\colon T_0 > t\} = \{\omega\colon X > t, Y > t\}$ we need only show that

$$P(t < X \leq t+a,\ X \leq Y) = \int_t^{t+a} \nu(u)\, P(T_0 \geq u)\, du. \qquad (6.2)$$

Suppose $f(u) = \frac{d}{du} F(u)$. The left hand side of (6.2) may be written as

$$\int_t^{t+a} P(Y \geq u \mid X = u) f(u)\, du,$$

while the right hand side is

$$\int_t^{t+a} [f(u)/\{1-F(u)\}]\; P(X \geq u,\ Y \geq u)\, du$$

$$= \int_t^{t+a} f(u)\; P(Y \geq u \mid X \geq u)\, du.$$

Thus the equality will hold for all a if and only if

$$P(Y \geq u \mid X \geq u) = P(Y \geq u \mid X = u). \tag{6.3}$$

But

$$P(Y \geq u \mid X = u) = -\{f(u)\}^{-1} \frac{\partial}{\partial t_1} \pi(t_1, t_2)\Big|_{t_1 = t_2 = u}$$

and $P(Y \geq u \mid X \geq u) = \pi(u,u)\ \{S(u)\}^{-1}$. Thus equation (6.3) holds if and only if

$$\frac{f(u)}{S(u)} = \frac{\frac{-\partial}{\partial t_1} \pi(t_1, t_2)\Big|_{t_1 = t_2 = u}}{\pi(u,u)},$$

which is precisely condition (2.1).

We remark here that an application of the Doob-Meyer decomposition theorem for submartingales of the class D (see Meyer 1966, chap. VII, sec. 1-4) can be used to show that if the family $\mathcal{F}_t$ is generated by just the process $N^D(t)$, then M(t) is a martingale with respect to $\mathcal{F}_t$ regardless of the dependence relationship between X and Y. Both Aalen (1978) and Lipster and Shiryayev (1978, chap. 18) give excellent summaries of this aspect of the structure of counting processes. Because of some technical requirements that will arise later on, however, we require that the σ-fields $\mathcal{F}_t$ be rich enough to allow the adaptability of $N^C(t)$. Aalen and Johansen (1978) have given a different method for preserving the martingale property of M(t) with enlarged σ-fields which has a more measure theoretic flavor.

Suppose now that the stochastic process $\overline{M}(t)$ is given by

$$\overline{M}(t) = \overline{N}^D(t) - \int_0^t \nu(u)\, N(u)\, du\ ,$$

for $t \in T$, where $\bar{N}^D(t)$ is the number of observed deaths in the sample up to and including time t, and N(u) is the size of the risk set at time u; that is $N(u) = \sum_{j=1}^{N} I_{[T_j \geq u]}$. Using slight variants of the arguments in Aalen (1978), one may show that $\bar{M}(t)$, $t \in T$, is a square integrable martingale with respect to the family of σ-subalgebras $\bar{F}_t = \bigotimes_{j=1}^{N} F_t^{\,j}$, where $F_t^{\,j} = \sigma(I_{[T_j \leq u,\ X_j \leq Y_j]}, I_{[T_j \leq u,\ Y_j < X_j]};\ 0 \leq u \leq t)$. By letting $H_N^\alpha(u) = \frac{1}{2}\,[\{S_0(u)\}^\alpha + \{\hat{S}(u^-)\}^\alpha]\ \{\hat{C}(u^-)N\}^{1/2}$, it is not hard to show, after some algebra, that $B_N^\alpha(t) = L_N^\alpha(t) + r_N^\alpha(t)$, where

$$L_N^\alpha(t) = -\int_0^t H_N^\alpha(u)\ \{N(u)\}^{-1}\ I_{[N(u)>0]}\ d\,\bar{M}(u)$$

and

$$r_N^\alpha(t) = \int_0^t H_N^\alpha(u)\ I_{[N(u)>0]}\ \{\nu_0(u) - \nu(u)\}\ du.$$

(Note that we always take 0/0 to be 0.) Since the conditions of Proposition 3 of Doléans-Dadé and Meyer (1970) are satisfied, the Stieltjes integral $L_N^\alpha(t)$ may be identified with the martingale stochastic integral with respect to $\bar{M}(t)$.

Let $g(u) = \frac{1}{2}\,[\{S(u)\}^\alpha + \{S_0(u)\}^\alpha]\ \{S(u)\}^{-1/2}\ \{\nu(u)\}^{1/2}$ where S(u) is the true underlying survival function for X_j. Theorem 3.4 in Rebolledo may now be applied to the process $L_N^\alpha(t)$ to show that

$$\{L_N^\alpha(t): t \in T\} \Rightarrow \left\{\int_0^t g(u)\ dW(u): t \in T\right\} \quad \text{as } N \to \infty.$$

The details needed to check the sufficient conditions given by Rebelledo in his Theorem 3.4 can be constructed by using modifications of the arguments in Aalen (1977). Furthermore, it is clear that under H_0, $r_N^\alpha(t) \equiv 0$.

6.2 Two-Sample Test Statistics

Many of the details involved in the proof of Theorem 4.1 are nearly identical to those required for the proof of Theorem 3.1. Using Lemma 6.1, it is easy to construct a product probability space $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{P})$ and an increasing family of σ-subalgebras $\overline{\mathcal{F}}_t$ so that the stochastic processes

$$\overline{M}_1(t) = \overline{N}_1^D(t) - \int_0^t N_1(u)\, \nu_1(u)\, du$$

and

$$\overline{M}_2(t) = \overline{N}_2^D(t) - \int_0^t N_2(u)\, \nu_2(u)\, du$$

are orthogonal square integrable martingales with respect to $\overline{\mathcal{F}}_t$. A calculation similar to that required in Theorem 3.1 shows that

$$B^{\alpha}_{N_1,N_2}(t) = L^{\alpha}_{N_1}(t) - L^{\alpha}_{N_2}(t) + r^{\alpha}_{N_1,N_2}(t)$$

where

$$L^{\alpha}_{N_1}(t) = \int_0^t H_{N_1,N_2}(u)\, N_1^{-1}(u)\, I_{[N_1(u)N_2(u)>0]}\, d\,\overline{M}_1(u),$$

$$L^{\alpha}_{N_2}(t) = \int_0^t H_{N_1,N_2}(u)\, N_2^{-1}(u)\, I_{[N_1(u)N_2(u)>0]}\, d\,\overline{M}_2(u),$$

and

$$r^{\alpha}_{N_1,N_2}(t) = \int_0^t H_{N_1,N_2}(u)\;\; I_{[N_1(u)N_2(u)>0]}\, \{\nu_1(u) - \nu_2(u)\}\, du.$$

Under H_0, $r^{\alpha}_{N_1,N_2}(t) = 0$, so it suffices to study the asymptotic behavior of the processes $L^{\alpha}_{N_i}(t)$. Suppose

$$g_i^2(u) = \frac{\lambda^{i-1}\, C_{3-i}(u)}{\lambda\, C_1(u) + C_2(u)}\;\; [\tfrac{1}{2}\, [\{S_1(u)\}^{\alpha} + \{S_2(u)\}^{\alpha}]]^2\, \{S_i(u)\}^{-1}\, \nu_i(s)$$

and that $W_1(t)$ and $W_2(t)$ are independent standard Wiener processes on $[0,\tau]$. Using a method nearly identical to that which yields the weak convergence of $L^{\alpha}_N(t)$ in the proof of Theorem 3.1, and the orthogonality of $\overline{M}_1(t)$ and $\overline{M}_2(t)$ (and hence the orthogonality of $L^{\alpha}_{N_1}(t)$ and $L^{\alpha}_{N_2}(t)$) one can show that

$$\vec{L}^{\alpha}_{N_1,N_2} = \{(L^{\alpha}_{N_1}(t),\ L^{\alpha}_{N_2}(t));\ 0\le t\le\tau\} \Rightarrow \vec{Y}^{\alpha} \text{ in } D^2[0,\tau]$$

where

$$\vec{Y}^{\alpha} = \{(\int_0^t g_1(u)\, dW_1(u),\ \int_0^t g_2(u)\, dW_2(u));\ 0\le u\le\tau\}.$$

The theorem now follows by calculating the asymptotic distribution of $L^{\alpha}_{N_1}(t) - L^{\alpha}_{N_2}(t)$ in the obvious way.

ACKNOWLEDGMENT

The research of the first author was supported in part by NIH Grant CA 24089-01. The second author was partially supported by NSF Grant ENG 78-06733. Support for the simulation results was provided by the F.D.A. Task Order #5 through EBON Research Systems.

BIBLIOGRAPHY

Aalen, O. O. (1976). Nonparametric inference in connection with multiple decrement models. Scand. J. Statist. 3, 15-27.

Aalen, O. O. (1977). Weak convergence of stochastic integrals related to counting processes. Z. Wahrscheinlichkeitstheorie Verw. Gebeite 38, 261-77.

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

Aalen, O. O. & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scand. J. Statist. 5, 141-50.

Barr, D. R. & Davidson, T. G. (1973). A Kolmogorov-Smirnov test for censored samples. Technometrics 15, 739-57.

Billingsley, P. (1968). Convergence of Probability Measures. New York: John Wiley & Sons, Inc.

Doléans-Dadé, C. & Meyer, P. A. (1970). Intégralés stochastiques par rapport aux martingales locales. Seminaire de Probabilities IV. Lecture Notes in Mathematics 124, 77-107. Berlin: Springer-Verlag.

Dufour, R. & Maag, J. R. (1978). Distribution results for modified Kolmogorov-Smirnov statistics for truncated or censored samples. Technometrics 20, 29-32.

Feller, W. (1971). An Introduction to Probability Theory and its Applications. Vol. 2, 2nd ed. New York: John Wiley & Sons, Inc.

Fleming, T. R. & Harrington, D. P. (1979). Nonparametric estimation of the survival distribution in censored data. Technical Report Series No. 8, Section of Medical Research Statistics, Mayo Clinic, 32 pages.

Gail, M. (1975). A review and critique of some models used in competing risk analysis. Biometrics 31, 209-22.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 52, 203-223.

Karlin, S. & Taylor, H. M. (1975). A First Course in Stochastic Processes, 2nd ed. New York: Academic Press.

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

Koziol, J. R. & Byar, D. P. (1975). Percentage points of the asymptotic distributions of one and two sample K-S statistics for truncated or censored data. Technometrics 17, 507-10.

Lipster, R. S. & Shiryayev, A. N. (1978). Statistics of Random Processes. Vol. 2. New York: Springer.

Meyer, P. A. (1966). Probabilities and Potentials. Waltham, Massachusetts: Blaisdell Publishing Company.

Nelson, W. (1969). Hazard plotting for incomplete failure data. J. of Qual. Tech. 1, 27-52.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). J. Royal Statist. Soc., Series A 135, 185-207.

Prentice, R. L. & Marek, P. (1979). A qualitative discrepancy between censored data rank tests. Biometrics 35, 861-869.

Rebolledo, R. (1978). Sur les applications de la théorie des martingales à l'étude statistique d'une famille de processes pontrels. In Journées de Statistique des Processes Stochastiques. Lecture Notes in Mathematics 636, 27-70. Berlin: Springer-Verlag.

Schey, H. M. (1977). The asymptotic distribution of the one sided Kolmogorov-Smirnov statistic for truncated data. Commun. Statist. - Theor. Meth. A6, 1361-65.