A Survey of a Class of Nonparametric Two-Sample Tests for Right Censored Failure Time Data

Harold K. Colomay

Department of Mathematics and Statistics
McGill University
Montréal, Québec

July 1992

A thesis submitted to the Faculty of Graduate Studies
in partial fulfillment of the requirements
for the degree of Master of Science

ABSTRACT

This thesis takes an in-depth focus at a specific class of nonparametric two-sample procedures for right censored failure time data–standardized weighted log-rank (SWL) statistics. This family of tests comprises the very famous Gehan, Efron, and log-rank procedures. The first two of these reduce to the Wilcoxon test with censoring absent, while the third one is a censored data generalization of the Savage test. Two particular topics of interest to us are (1) the generation of SWL statistics as score tests within the context of some popular regression models, and (2) asymptotic and small sample behavior.

RÉSUMÉ

Le présent mémoire explore de façon détaillée une catégorie spécifique de procédures pour deux échantillons censurés à droite de temps jusqu'à défaillance, nommément les procédures impliquant les statistiques pondérées et normées de logarithmes de rangs (statistiques SWL). Cette famille de tests comprend les célebres tests de Gehan, Efron et de logarithmes de rangs. Les deux premières procédures se réduisent au test de Wilcoxon en l'absence de censure, alors que la troisième est une généralisation aux observations censurées du test de Savage. Deux sujets plus spécialement approfondis ici sont (1) la production de statistiques SWL dans le but d'effectuer des tests de cote au sein de certains modèles de régression courants, et (2) les propriétés de ces statistiques dans les cas d'échantillons de petite et de grande taille.

PREFACE

None of the contents of this work are original, with the exception of Monte Carlo simulations which assess the goodness of fit of the bootstrap distribution of some SWL statistics in small and moderate size samples.

I would like to take this opportunity to express my deepest appreciation to Professor M.G. Gu, my thesis supervisor, for his invaluable assistance in the preparation of this treatise both in the written and Monte Carlo simulations portions. Moreover, I am grateful to Professor Gu for lending me several of his textbooks, and for giving me a copy of his 1991 paper.

Finally, I wish to thank Mr. Alain Vandal for translating the abstract into French.

NOTATION, SYMBOLS, AND CONVENTIONS

$\mathbf{a}$	bold symbols, such as $\mathbf{a}$, denote column or row vectors
$\mathbf{a}^T$	the transpose of $\mathbf{a}$
$\Pr(A)$	probability of the event A
$\Pr(A \mid \mathbf{X} = \mathbf{x}, \mathbf{B}_1, \ldots, \mathbf{B}_k)$	conditional probability of the event A, given $\mathbf{X} = \mathbf{x}$ and given the events $\mathbf{B}_1, \ldots, \mathbf{B}_k$ have occurred. Sometimes, we shall simply write $\Pr(A \mid \mathbf{X})$. Unless otherwise indicated, $\mathbf{x}$ shall be considered as being random.
$f_{\mathbf{X}}(\mathbf{x})$	joint density or probability function of $\mathbf{X}$. If each element of $\mathbf{X}$ is a discrete variate, the notation "$\Pr(\mathbf{X} = \mathbf{x})$" shall as well be employed. Unless otherwise indicated, $\mathbf{x}$ shall be considered as being random.
$f_{\mathbf{X} \mid \mathbf{Y}}(\mathbf{x} \mid \mathbf{y})$	conditional density or probability function of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$. If each element of $\mathbf{X}$ is a discrete variate, the notation "$\Pr(\mathbf{X} = \mathbf{x} \mid \mathbf{Y} = \mathbf{y})$" shall as well be employed. Unless otherwise indicated, both $\mathbf{x}$ and $\mathbf{y}$ shall be considered as being random.

In the last three cases, the dummy variables "$\mathbf{x}$" and "$\mathbf{y}$" may be replaced by any other dummy variables.

i.i.d.	independently and identically distributed
d.f.	distribution function. We shall always assume a d.f. to be nonrandom.

ARE	asymptotic relative efficiency
$I(A)$	indicator of an event A. It is a 0,1 random variable which is unity when the event A has occurred and zero otherwise.
Asvar(X)	asymptotic variance of X. For the purposes of this treatise, we shall always assume that Asvar$(\cdot)$ exists.
$N(\mu, \sigma^2)$	the normal distribution with mean μ and variance σ^2
$\xrightarrow{P}$	convergence in probability
$\xrightarrow{D}$	convergence in distribution
$\gg$	is much greater than
$\#A$	number of elements in the set A
A^c	the complement of the set A
$\emptyset$	the empty set
$A \subset B$	A is a proper subset of B

Let $X(t)$ be either a random or nonrandom function of time. Then "$X(t^-)$" shall denote "$\lim_{h\to 0^-} X(t+h)$," while "$X(t^+)$" shall denote "$\lim_{h\to 0^+} X(t+h)$."

Throughout, we use the convention $0/0 = 0$.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

Failure time data refer typically to a set of independent, continuous, positive-valued random variables (each variable corresponding to a different item) which represent times to occurrence of some undesired point event. Prime examples of such an event include death, onset of disease, or cessation of function of an electrical component. It is quite permissible, however, for the event of interest to be of a nonfailurelike nature (for example, the learning of a new skill or the changing of residence). In addition, the random variables need neither be a measure of time nor be continuous. For instance, each variable may represent the number of attempts required to successfully perform a certain task. More formally then, failure time data is defined as a collection of independent positive random variables, each one of which is associated with a unique item and indicates, in some sense, the immediacy of occurrence of some point event.

For the purposes of this thesis, when considering failure time data in a generalized setting, the event of interest shall be denoted by the word "failure" with the terms "failure time," "event time," and "lifetime" being used interchangeably. Moreover, from hereon, we shall restrict ourselves to the case of absolutely continuous failure times.

We now introduce the concept of *right censorship*. Right censorship refers to the process by which a lifetime is not observed exactly but is known only to exceed a certain value. This particular value is referred to as a *right censored lifetime*, and the item and lifetime variate in question are said to have been *right censored*. If a failure is observed, we describe the corresponding lifetime value as being an *uncensored lifetime*, and we say that the lifetime variate of interest was uncensored. By convention, an item cannot simultaneously fail and be right censored. To describe in mathematical terms the failure or censoring process associated with a given item, let Y be the variable which denotes either the uncensored or right censored event time (hereafter referred to as the survival time variable), and let T be the event time variate. If failure is observed, then $Y = T$; if T is right censored, then $Y < T$. For convenience, from hereon, the term "censoring" shall be used in place of the term "right censoring."

Survival analysis is the branch of statistics that encompasses a variety of techniques for analyzing failure time data regardless of whether or not censoring is present. A major area of interest among researchers in survival analysis is the following two-sample scenario: Based on two sets of lifetimes subject to right censorship — the lifetimes of each sample being identically distributed — test the null hypothesis that the two respective failure time distribution functions are identical. This statistical problem can present itself in a variety of ways. For example, consider a clinical trial in which two different treatments for a fatal disease are being compared and patients enter consecutively. The variable of interest here is time to death from initiation of treatment. A patient yields a censored lifetime if he withdraws from the study due to intolerable side effects, or moves away and is thus lost to follow-up, or dies of some other cause, or is still alive on the predetermined termination date of the trial.

The two-sample situation, described at the beginning of the previous paragraph, also arises in those carcinogenesis experiments which compare the potency of two types of carcinogenic chemicals. In such experiments, one variety of carcinogen is applied repeatedly to one group of animals, while another variety is applied repeatedly to a second group–the two groups being put on test at the same time. In this scenario, the random variable of interest is the time from starting the applications to finding a tumour. An event time becomes censored if an animal dies without developing a tumour, or if an animal is alive without a tumour when the experiment is terminated.

Our two-sample scenario can as well be seen in an industrial reliability study which attempts to compare the lifetimes of two types of electrical components. As in the previous case, all items are placed on test simultaneously, with a decision made to end the study after a time L has elapsed. The statistic of concern here is time from commencement of operation of the components until failure. A censored lifetime is observed only if failure has not occurred prior to the conclusion of the study.

In the comparison of two samples with uncensored data, rank tests have often been proposed as alternatives to parametric methods. Although the rank tests themselves are derived with certain alternative hypotheses in mind for which optimum parametric procedures exist, the former generally possess greater robustness than the latter against misspecification of the parametric form of the two distribution functions and are generally less sensitive to outliers. In addition, rank tests usually are asymptotically fully efficient when the distribution functions are correctly described parametrically. In small-sample situations, rank procedures generally experience only a small loss in efficiency compared to their parametric analogues when the latter are appropriate.

During the past thirty years, several tests based on the generalized rank vector have been developed for tackling the two-sample problem with censoring. Without restricting ourselves to the two-sample problem, the generalized rank vector for n survival times $Y_1, \ldots, Y_n$ can be represented as follows:

$$\mathbf{R}_G = ((N(Y_1), \Delta_1), \ldots, (N(Y_n), \Delta_n)),$$

where $N(t) = \sum_{j=1}^{n} I(Y_j \leq t, \Delta_j = 1)$, and

$$\Delta_i = \begin{cases} 1 & \text{if item } i \text{ fails,} \\ 0 & \text{if item } i \text{ is censored.} \end{cases}$$

For the two-sample scenario with sample sizes n_1 and n_2, we assume that items $1, \ldots, n_1$ comprise sample 1, while the remainder form sample 2. Remark that when censoring is absent, $\mathbf{R}_G$ reduces to the so called "rank vector."

Among the oldest of the two-sample generalized rank tests are the log-rank (Mantel, 1966; Peto and Peto, 1972; Cox, 1972), Gehan (Gehan, 1965), and Efron (Efron, 1967) test statistics. More recently developed significant procedures include Prentice's test (Prentice, 1978), the Tarone-Ware class of statistics (Tarone and Ware, 1977), and the Harrington-Fleming family of statistics (Harrington and Fleming, 1982). It should be noted that the log-rank test is a censored data generalization of the Savage test (Savage, 1956), while Gehan's, Prentice's, and Efron's procedure each simplify to Wilcoxon's test (Wilcoxon, 1945) with no censoring present.

Each of the tests mentioned in the last paragraph, in addition to being a generalized rank procedure, is a standardized weighted log-rank (SWL) statistic. It is our intention to fully examine the characteristics of the family of such statistics. To describe explicitly the form of this class of two-sample procedures, let $T_1^o < \cdots < T_K^o$ denote the ordered uncensored failure times for the combined sample of size n. For $i = 1, \ldots, K$, let $D_{1i} = 1$ if the failure at T_i^o is from sample 1, and let $D_{1i} = 0$ if otherwise. Let R_{1i} and R_{2i} be, respectively, the number of sample 1 and sample 2 subjects with lifetimes known to be greater than or equal to T_i^o, and set $R_i = R_{1i} + R_{2i}$. Then the statistics of interest have the form

$$U_n = \frac{n^{-1/2} \sum_{i=1}^{K} W_i(D_{1i} - R_{1i}/R_i)}{\sqrt{V}}.$$

Here, W_i is a "weight" associated with T_i^o and dependent on

$$D_{11},\ R_{11},\ R_{21},\ D_{12},\ R_{12},\ R_{22},\ \ldots,\ D_{1,i-1},\ R_{1,i-1},\ R_{2,i-1},\ R_{1i},\ R_{2i},$$

while V is either the exact variance of the numerator under H_0 or the product of n^{-1} and an estimator of the null hypothesis variance of the summation term. The exact variance can only be utilized if it is free of the common failure time d.f. (assumed to be unknown) and of all unspecified d.f.'s linked to the censoring mechanism. If V is a random variable, it usually–but not necessarily–is a function of only the generalized rank vector. Note that, unless V is the asymptotic variance of the numerator, $n^{-1/2}$ cancels out.

When $W_i = 1$, for $i = 1, \ldots, K$, the numerator of U_n (with or without $n^{-1/2}$) is the log-rank statistic, hence the name "standardized weighted log-rank statistic." The name given to U_n depends solely on our choice for the W_i's and is unaffected by the choice of V; moreover, $-U_n$ and $+U_n$ have the same name. The actual weights implemented are chosen so as to maximize the power of U_n against the particular alternative hypothesis in mind. For ease of clarity, a particular U_n, with or without V specified, shall be referenced interchangeably with the words "test," "test statistic," and "procedure," while a particular weighted log-rank statistic (with or without $n^{-1/2}$) shall be referenced via the term "statistic." Hence, the "log-rank test" is U_n with $W_i = 1,\ i = 1, \ldots, K$. More generally, the standardized version of any given statistic shall be denoted interchangeably by the words "test," "test statistic," and "procedure," while the unstandardized version shall be denoted by the term "statistic."

Before outlining the actual contents of this thesis, we make some remarks concerning all censoring mechanisms considered in this treatise. Suppose that associated with item i ($i = 1, \ldots, n$) of a life-testing experiment (not necessarily concerned with the two-sample problem) is the possibly time-dependent, possibly random regressor variable $Z_i(t)$. Let

$$X_i(t) = \{Z_i(u) :\ 0 < u \leq t\}.$$

(If in fact $Z_i(t)$ does not vary with time, then $Z_i(t) = X_i(t) = Z_i$, for some Z_i.) Then, for the purposes of this treatise we make the assumption that, conditional on $X_i(t)$, the censoring and failure mechanisms act independently of one another. Such a censoring scheme is described as being "independent" or "arbitrary," and yields the following two consequences: Firstly, if item j is at risk[1] at time t^-, then, conditional on $X_j(t)$, the failure time hazard function[2] of j is unaffected by the fact that j was uncensored in $(0, t)$. Secondly, if an item with regressor variable path $X(t)$ is

[1] If an item is at risk at time t (t^-), this means that it has neither failed nor been censored in $(0, t]$ ($(0,t)$).

[2] The hazard function corresponding to, say, the random variable T is defined as

$$\lim_{h \to 0^+} \frac{\Pr(T < t + h | T \geq t)}{h}.$$

withdrawn from risk at time t, this is in no way an indicator of its prognosis relative to other items at risk at t with path $X(t)$.

We now give examples of independent censoring processes employing the notation that T_i, Y_i are, respectively, the lifetime and survival time variates associated with item i ($i = 1, \ldots, n$).

Example 1: Simple type I censorship

In this case, all study subjects are placed on test at the same time with the decision made to terminate the experiment after a time L has elapsed. The potential censoring time[3] for all subjects is thus L.

Example 2: Progressive type I censorship (or fixed censorship)

Here, as in example 1, the study is terminated at some prespecified time, say L. In fixed censorship, however, the subjects enter the study at random in the interval $[0, a]$, where $a \leq L$. Letting E_i denote the entry time variate corresponding to item i, the potential censoring time variable for item i is thus $L - E_i$. The E_i's are assumed to be independently distributed of each other and of the event times. For the purposes of this thesis, we shall assume as well that the E_i's are identically distributed. Note that the number of items which enter the study, n, is random. While it is customary to condition on n and on the entry times, we shall, in this treatise, condition on only n. (For the two-sample problem, we shall condition on the size of

---

For the purposes of this thesis, we shall treat any given hazard function as a nonrandom quantity.

[3]By "potential censoring time," we mean the time at which the item in question is destined to be censored should failure not occur; hence, if in fact an item is censored, its censored lifetime coincides with its potential censoring time. On the other hand, if an item fails, its potential censoring time is either exactly observable or right censored at some value which is greater than or equal to its lifetime value. From hereon, we shall restrict our use of the term "potential censoring time" to the censoring models of examples (1), (2), and (3).

each sample.)

Example 3: Random censorship

If associated with each T_i is the potential censoring time variate C_i such that the C_i's are independent of one another and of the T_i's, then we are in the *random censorship* situation. The C_i's need not be absolutely continuous random variables. Remark that $Y_i = \min(T_i, C_i)$, $i = 1, \ldots, n$.

Examples 1 and 2 belong to the subclass of random censorship models where the potential censoring time for each item is observable. A random censorship model where potential censoring times corresponding to uncensored event times cannot be observed is seen in a one-sample clinical trial where one cause of censoring is withdrawal from the study due to severe side effects of the treatment.

Example 4: Simple type II censorship

In this situation, all subjects are placed on test at the same time, and observation ceases after a predetermined number of failures $r \leq n$. Thus, if $T_1^o < \cdots < T_n^o$ are the order statistics of the T_i's, then $Y_i = \min(T_r^0, T_i)$, $i = 1, \ldots, n$.

Example 5: Progressive type II censorship

This censoring model is a generalization of that of example 4. Suppose n individuals enter the study simultaneously. The observation plan now is as follows: At the time of the first observed failure, we remove from the experiment a simple random sample of m_1 subjects from the still unfailed $n-1$. Then at the next observed failure time, a further m_2 individuals are selected at random from those still on test and removed. This procedure is carried on until a total of s failures have been observed, with m_k subjects being withdrawn at the k^{th} stage, $k = 1, \ldots, s$;

$$\sum_{k=1}^{s} (m_k + 1) = n.$$

We comment here that $m_1, \ldots, m_s$ and s are fixed in advance.

We now list in brief form the actual contents of this thesis: In chapter 2, we present elementary concepts of our class of two-sample tests in a somewhat heuristic and informal fashion. In chapters 3 and 4, we generate SWL statistics from, respectively, the proportional hazards model and the accelerated failure time model. Finally, in chapter 5, we examine the large and small-sample behavior of our family of tests.

CHAPTER 2
GENERAL OVERVIEW OF CLASS OF TESTS

2.1 Introduction.

The purpose of this chapter is to present, with little theoretical detail, fundamental concepts of SWL statistics. In section 2.2, we derive heuristically this class of two-sample tests and describe the most common variance estimators. In section 2.3, we express a weighted log-rank statistic (or a WL statistic) as a sum of scores, as a generalized Mann-Whitney statistic, and as a member of Gill's class K (Gill, 1980). Section 2.4 considers the construction of censored data extensions of linear rank statistics. The chapter then concludes with a discussion of the most well known tests.

2.2 Informal Development of SWL Statistics.

Our approach to generating SWL statistics is in the spirit of Mantel (1966).

Let $\{X_{j1}, \ldots, X_{jn_j}\}$ and $\{T_{j1}, \ldots, T_{jn_j}\}$ be, respectively, the set of survival time and failure time variates for group j; moreover, let $\Delta_{j1}, \ldots, \Delta_{jn_j}$ be the censoring indicators for sample j defined by

$$\Delta_{ji} = \begin{cases} 1 & \text{if } X_{ji} = T_{ji}, \\ 0 & \text{if } X_{ji} < T_{ji} \end{cases}$$

$(i = 1, \ldots, n_j;\ j = 1, 2;\ n_1 + n_2 = n)$. The data, therefore, is of the form

$$(X_{11}, \Delta_{11}), \ldots, (X_{1n_1}, \Delta_{1n_1}), (X_{21}, \Delta_{21}), \ldots, (X_{2n_2}, \Delta_{2n_2}).$$

Now, let $F_j(t)$ which is unknown be the d.f. corresponding to $T_{j1}, \ldots, T_{jn_j}$. We wish to test the null hypothesis

$$H_0 : F_1(t) = F_2(t) = F(t) \quad \forall t \geq 0,$$

where $F(t)$ is unspecified, either against the one-sided alternative

$$H_1 : F_1(t) \geq F_2(t) \quad \forall t \geq 0,$$

with strict inequality for at least one t, or against the two-sided alternative

$$H_2 : F_1(t) \neq F_2(t),$$

for at least one t. Since we are assuming lifetime to be an absolutely continuous random variable, no two uncensored lifetimes can coincide with one another. Hence, let $T_1^o < \cdots < T_K^o$ denote the ordered uncensored event times for the sample formed by pooling the two groups of data, and set $D_{1i} = 1$ or 0 depending on whether the failure at T_i^o is from sample 1 or 2. In accordance with convention, an item which fails at T_i^o does so without simultaneously being censored. Finally, suppose R_i study subjects are at risk at T_i^{o-} and that R_{ji} is the corresponding number in sample j $(j = 1, 2)$. By their very nature, the variables $D_{1i}, R_{1i}, R_{2i}, R_i$ utilize the information that an i^{th} smallest uncensored failure time exists, yet they ignore the value of T_i^o.

Because independent censoring is in effect, we would expect that little information about differences between $F_1(t)$ and $F_2(t)$ would be obtained from the ordering of censorings between successive failures. Bearing in mind both this point and the fact that we wish to develop a *nonparametric procedure* for testing H_0, it follows that we need only concern ourselves with that portion of the data displayed in the following sequence of contingency tables, the i^{th} of which is associated with T_i^o:

	number of failures	number at risk
Group I	D_{1i}	R_{1i}
Group II	$1 - D_{1i}$	R_{2i}
Total	1	R_i

$i = 1, \ldots, K$. Conditional on R_{1i}, R_{2i}, and on independent censoring being in effect (the regressor variable being group membership), D_{1i} has a Bernoulli distribution under H_0 with mean $R_{1i}R_i^{-1}$ and variance $V_i = R_{1i}R_{2i}R_i^{-2}$. Now, for $i = 1, \ldots, K$, let $\mathbf{X}_i = (R_{1i}, R_{2i})$, and assume

$$\Pr(D_{1i} = d_{1i}|\mathcal{H}_i) = \Pr(D_{1i} = d_{1i}|\mathbf{X}_i),$$

where

$$\mathcal{H}_i = \mathcal{H}(T_i^o),$$

$$\mathcal{H}(t) = \left(R_1(t), R_2(t), N(t^-), \{R_{1i}, R_{2i}, D_{1i} : \; i \leq N(t^-)\}\right),$$

$$R_j(t) = \sum_{i=1}^{n_j} I(X_{ji} \geq t),$$

and

$$N(t) = \sum_{j=1}^{2}\sum_{i=1}^{n_j} I(X_{ji} \leq t, \Delta_{ji} = 1).$$

Setting $L_i = W_i(D_{1i} - R_{1i}R_i^{-1})$, we thus have

$$E_{H_0}(L_i|\mathcal{H}_i) = 0$$

and

$$\mathrm{Var}_{H_0}(L_i|\mathcal{H}_i) = W_i^2 V_i,$$

where $W_i = W(T_i^o)$, and where $W(t)$ is either a random weight function completely determined by $\mathcal{H}(t)$ or a nonrandom quantity independent of time. Now, for $i < j$ and under H_0,

$$\mathrm{Cov}(L_i, L_j) = E\left[L_i E(L_j|\mathcal{H}_j)\right] - E\left[E(L_i|\mathcal{H}_i)\right] E\left[E(L_j|\mathcal{H}_j)\right] = 0.$$

Hence, letting $L_{1n} = \sum_{i=1}^{K} L_i$ and treating K as a fixed quantity if it is in fact random, we have under H_0 that $E(L_{1n}) = 0$ and

$$\mathrm{Var}(L_{1n}) = \sum_{i=1}^{K} \mathrm{Var}(L_i) = \sum_{i=1}^{K} E\left[\mathrm{Var}(L_i|\mathcal{H}_i)\right] + \sum_{i=1}^{K} \mathrm{Var}\left[E(L_i|\mathcal{H}_i)\right] = E(V_{cp}),$$

where $V_{cp} = \sum_{i=1}^{K}(V_i W_i^2)$. Thus, V_{cp} is unbiased for $\mathrm{Var}_{H_0}(L_{1n})$. If K is a random variable and is treated as such, the above expressions for $E_{H_0}(L_{1n})$ and $\mathrm{Var}_{H_0}(L_{1n})$ can be obtained using martingale theory.

For simple and progressive type II censorship, there is only one d.f. to be concerned with under H_0, that being $F(t)$, which is of course unspecified. Evaluation of $\mathrm{Var}_{H_0}(L_{1n})$ in both of these instances, however, does not require knowledge of $F(t)$, regardless of the choice for $W(t)$. The random censorship model, on the other hand, incorporates under H_0, $F(t)$, as well as censoring d.f.'s[4] which are assumed to be distinct from $F(t)$. If the censoring d.f.'s are discrete and specified (for example, simple type I censorship), then $\mathrm{Var}_{H_0}(L_{1n})$ is free of $F(t)$ and thus calculable, regardless of the weight function employed. For the case, though, where the censoring d.f.'s be they discrete or continuous–are unspecified, this variance will be

[4] From hereon, the term "censoring d.f." shall mean the d.f. corresponding to the potential censoring time variate in question.

distribution-free only for a specific type of weight function (more on the above points in subsection 5.3.1). Whatever the censoring scheme, though, it can be shown that, under H_0, $n^{-1}V_{cp}$ is consistent for $\mathrm{Var}_{H_0}(n^{-1/2}L_{1n})$ if, as $n \to \infty$,

$$(1)\ K \xrightarrow{P} \infty \quad (\text{or } K \to \infty)$$

and

$$(2)\ \min\{n_1, n_2\} \to \infty;$$

thus, the use of V_{cp} rather than $\mathrm{Var}_{H_0}(L_{1n})$ is appropriate if K, n_1, and n_2 are reasonably large. Latta (1981) referred to V_{cp} as the *conditional permutation variance estimator*. For brevity, we, as well, shall use the term *conditional permutation variance*.

Before presenting another frequently-used variance estimator, we introduce the concept of *equal censoring patterns*. By definition, two samples have equal censoring patterns if

$$p_{11}(t) = p_{12}(t) = \cdots = p_{1n_1}(t) = p_{21}(t) = p_{22}(t) = \cdots = p_{2n_2}(t) \quad \forall t,$$

where

$$p_{ji}(t) = \Pr(X_{ji} \in [t, t + dt), \Delta_{ji} = 0 | T_{ji} \geq t)$$

(Lagakos, 1979). Examples of censoring scenarios, where the above condition holds, include: simple type I censorship, fixed censorship, random censorship with potential censoring times identically distributed, and simple and progressive type II censorship with H_0 true. In the following two paragraphs, all expectations, variances, and statements concerning consistency are under the assumption that H_0 is true and equal censoring patterns are present.

As will be shown in subsection 2.3.1, L_{1n} can be written as

$$\sum_{i=1}^{K} (D_{1i}Q_i + M_{1i}Q_i^*),$$

where

$$Q_i = W_i - \sum_{j=1}^{i} W_j/R_j$$

is a score associated with T_i^o, where

$$Q_i^* = -\sum_{j=1}^{i} W_j/R_j$$

is a score corresponding to all censored lifetimes in $[T_i^o, T_{i+1}^o)$, and where M_{1i} is the number of group 1 censorings in this interval. By convention, we set $Q_0^* = 0$, $T_0^o = 0$, and $T_{K+1}^o = +\infty$. Now, let $\mathbf{S}$ be the collection

$$(Q_{01}^*, \ldots, Q_{0M_0}^*, Q_1, Q_{11}^*, \ldots, Q_{1M_1}^*, \ldots, Q_K, Q_{K1}^*, \ldots, Q_{KM_K}),$$

where M_i (for $i = 0, \ldots, K$) is the total number of censored failure times in $[T_i^o, T_{i+1}^o)$, and where $Q_{i1}^*, \ldots, Q_{iM_i}^*$ have the value Q_i^*, yet are considered as distinct elements. (We comment here that $\mathbf{S}$ provides no direct information concerning whether a given score corresponds to a censored or uncensored failure time.) In addition, let $\mathbf{P} = (K, \mathbf{S})$. Therefore,

$$\begin{aligned} H_1(a) &= \Pr(L_{1n} \le a|\mathbf{P}) \\ &= \Pr\left(\sum_{i=1}^{n_1} C_i \le a\right) \quad \forall a \in (-\infty, +\infty), \end{aligned}$$

where $C_1, \ldots, C_{n_1}$ is a sample of size n_1 obtained by withdrawal without replacement from $\mathbf{P}$. $H_1(a)$ is often referred to as the "permutation d.f. of L_{1n}." Using results on sampling from finite populations, we thus have that

$$E(L_{1n}|\mathbf{P}) = E\left(\sum_{i=1}^{n_1} C_i\right) = \frac{n_1 \sum_{i=1}^{K}(Q_i + M_i Q_i^*)}{n} = 0,$$

while

$$\begin{aligned} V_p &= \operatorname{Var}(L_{1n}|\mathbf{P}) \\ &= \operatorname{Var}\left(\sum_{i=1}^{n_1} C_i\right) \\ &= n_1 n_2 (n(n-1))^{-1} \sum_{i=1}^{K} \left(Q_i^2 + M_i Q_i^{*2}\right). \end{aligned}$$

The conditional variance V_p is referred to as either the *permutation variance estimator* or the *permutation variance*. In the evaluation of the above conditional expectation and variance, we used the result

$$\sum_{i=1}^{K} (Q_i + M_i Q_i^*) = 0.$$

This equality follows directly from the fact that the scores satisfy the equations

$$W_i = Q_i - Q_i^* = R_i(Q_{i-1}^* - Q_i^*),\ i = 1, \ldots, K.$$

One further result which is obtainable from this set of equations is

$$\sum_{i=1}^{K} (Q_i^2 + M_i Q_i^{*2}) = \sum_{i=1}^{K} W_i^2 (1 - R_i^{-1})$$

(Cox and Oakes, 1984, p. 141). Hence, the permutation variance can be expressed in two distinct forms.

Now,

$$\operatorname{Var}(L_{1n}) = E\left[\operatorname{Var}(L_{1n}|\mathbf{P})\right] + \operatorname{Var}\left[E(L_{1n}|\mathbf{P})\right] = E(V_p),$$

and so V_p is unbiased for $\operatorname{Var}(L_{1n})$. V_p/n, moreover, consistently estimates

$$\operatorname{Var}(n^{-1/2} L_{1n})$$

under conditions (1) and (2) of p. 14 and thus can adequately replace it when K, n_1, n_2 are large. V_p is the exact variance if the scores, the M_i's, and K are nonrandom (for example, simple type II censorship with log-rank scores).

Note that V_{cp} and V_p are functions of strictly the generalized rank vector

$$\mathbf{R}_G = ((N(X_{11}), \Delta_{11}), \ldots, (N(X_{1n_1}), \Delta_{1n_1}),$$
$$(N(X_{21}), \Delta_{21}), \ldots, (N(X_{2n_2}), \Delta_{2n_2})).$$

In chapter 4, we present another variance estimator which is a function of only $\mathbf{R}_G$, while in chapter 5 we put forth variance estimators which are not solely dependent on this vector. The suitability under H_0 and H_1 of all proposed estimators of $\mathrm{Var}_{H_0}(L_{1n})$, for small n, is an issue to be dealt with in chapter 5.

The general class of procedures we propose, therefore, for testing H_0 is given by

$$U_n = \frac{n^{-1/2} L_{1n}}{\sqrt{V}} = \frac{n^{-1/2} \sum_{i=1}^{K} W_i(D_{1i} - R_{1i} R_i^{-1})}{\sqrt{V}},$$

where V is $\mathrm{Var}_{H_0}(n^{-1/2} L_{1n})$ or an unbiased estimator thereof under H_0, or where V is

$$\mathrm{Asvar}_{H_0}(n^{-1/2} L_{1n})$$

or a consistent estimator thereof under H_0. With respect to H_1, one would reject H_0 if U_n is sufficiently large, while, for H_2, one would reject H_0 if $|U_n|$ is sufficiently large. Note that the numerator of U_n is solely dependent upon $(\mathcal{H}(T_K^o), D_{1K})$, which in turn is determined by $\mathbf{R}_G$ only. As with any statistic based strictly on $\mathbf{R}_G$, then, L_{1n} discards the exact failure and censoring times as well as the ordering of censored lifetimes between adjacent failures.

If the L_i's were independently and identically distributed under H_0, then

$$n^{-1/2}L_{1n}/\sqrt{\mathrm{Asvar}_{H_0}(n^{-1/2}L_{1n})}$$

would be asymptotically a standard normal random variable by the central limit theorem. Of course, the L_i's are highly dependent and are not identically distributed; nevertheless, asymptotic normality still holds as we shall demonstrate in chapter 5.

For the random censorship model, the small-sample null distribution of L_{1n} is dependent on all the d.f.'s in question, even if $\mathrm{Var}_{H_0}(L_{1n})$ is distribution-free. On the other hand, with simple or progressive type II censorship in effect, the small-sample null distribution of L_{1n} is free of $F(t)$.

2.3 Representations of a WL Statistic.

2.3.1 The Sum of Scores Form.

Another popular family of statistics for testing H_0, based on the generalized rank vector $\mathbf{R}_G$ as defined in section 2.2, is of the form

$$L_{2n} = \sum_{i=1}^{K}(D_{1i}Q_i + M_{1i}Q_i^*),$$

where Q_i is a score corresponding to T_i^o, Q_i^* is a score associated with all censored lifetimes in $[T_i^o, T_{i+1}^o)$, and D_{1i}, M_{1i} are as previously defined. Q_i and Q_i^* are such that $Q_i = Q(T_i^o)$, $Q_i^* = Q^*(T_i^o)$ with both $Q(t)$ and $Q_i^*(t)$ dependent on $\mathcal{H}(t)$; moreover, we set $Q_0^* = 0$, $T_0^o = 0$ and $T_{K+1}^o = +\infty$. An obvious question which arises is, under what conditions can L_{2n}

be expressed as a WL statistic? Prentice and Marek (1979) provided the answer to this. Firstly, set $W_i = R_i(Q^*_{i-1} - Q^*_i)$, $i = 1, \ldots, K$. (Remark that W_i is determined by $\mathcal{H}_i$.) Then

$$\begin{aligned}
L_{2n} &= \sum_{i=1}^{K}(D_{1i}Q_i + M_{1i}Q^*_i) \\
&= \sum_{i=1}^{K} D_{1i}(Q_i - Q^*_i) - \sum_{i=1}^{K}\left[(D_{1i} + M_{1i})\sum_{j=1}^{i} W_j R_j^{-1}\right] \\
&= \sum_{i=1}^{K} D_{1i}(Q_i - Q^*_i) - \sum_{i=1}^{K}\left[W_i R_i^{-1}\sum_{j=i}^{K}(D_{1j} + M_{1j})\right] \\
&= \sum_{i=1}^{K} D_{1i}(Q_i - Q^*_i - W_i) + \sum_{i=1}^{K} W_i(D_{1i} - R_{1i}R_i^{-1}) \\
&= \sum_{i=1}^{K} W_i(D_{1i} - R_{1i}R_i^{-1}),
\end{aligned} \tag{2.1}$$

if

$$W_i = Q_i - Q^*_i = R_i(Q^*_{i-1} - Q^*_i), \quad i = 1, \ldots, K. \tag{2.2}$$

On the other hand, statistics of the form (2.1) can always be expressed in terms of a sum of scores by letting

$$Q_i = W_i - \sum_{j=1}^{i}\frac{W_j}{R_j}, \qquad Q^*_i = -\sum_{j=1}^{i}\frac{W_j}{R_j}, \quad i = 1, \ldots, K.$$

These scores are obtained by first setting

$$W_i = R_i(Q^*_{i-1} - Q^*_i), \quad i = 1, \ldots, K$$

(thus yielding the Q^*_i's as defined above), and then by setting

$$W_i = Q_i - Q^*_i, \quad i = 1, \ldots, K$$

(hence yielding the Q_i's as given above).

The classical linear rank statistic for testing H_0 with censoring absent has the form

$$L_{2n}^* = \sum_{j=1}^{n} D_{1i} Q_i^u,$$

where $Q_i^u = Q^u(T_i^o)$ is a nonrandom score associated with T_i^o, and where $Q^u(t)$ either is a random function determined by $N(t^-)$ or is a nonrandom, time-independent quantity. Upon substituting $Q_1^u, Q_2^u, \ldots, Q_n^u$ for $Q_1, Q_2, \ldots, Q_n$ in (2.2), we obtain

$$Q_1^{*u}, Q_2^{*u}, \ldots, Q_n^{*u},$$

where

$$Q_i^{*u} = Q^{*u}(T_i^o) = \begin{cases} \sum_{j=1}^{i} Q_j^u/(i-n) & \text{if } i < n, \\ 0 & \text{if } i = n, \end{cases}$$

and where

$$Q^{*u}(t) = \begin{cases} \sum_{j=1}^{N(t^-)+1} Q_j^u/(N(t^-)+1-n) & \text{if } t \le T_{n-1}^o, \\ 0 & \text{if } t > T_{n-1}^o. \end{cases}$$

Hence, L_{2n}^* can be written as

$$\sum_{i=1}^{n} W_i^u (D_{1i} - R_{1i} R_i^{-1}),$$

where $W_i^u = W(T_i^o) = Q_i^u - Q_i^{*u}$, and where

$$W^u(t) = Q^u(t) - Q^{u*}(t).$$

L_{2n}^*, therefore, belongs to the class of WL statistics.

2.3.2 Generalized Mann-Whitney Statistic Form.

The following discussion is due to Gu, Lai, and Lan (1991). For any pair (X_{1i}, X_{2j}), define

$$U(X_{1i}, X_{2j}) = U_{ij} = \begin{cases} -W(X_{2j})/R(X_{2j}), & \text{if } \Delta_{2j} = 1 \text{ and } X_{2j} \leq X_{1i}, \\ W(X_{1i})/R(X_{1i}), & \text{if } \Delta_{1i} = 1 \text{ and } X_{1i} \leq X_{2j}, \\ 0, & \text{in all other cases,} \end{cases}$$

where $R(u) = R_1(u) + R_2(u)$. Then

$$\sum_{i=1}^{K} W_i(D_{1i} - R_{1i}R_i^{-1})$$

can be written as

$$\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} U_{ij}. \tag{2.3}$$

To prove this, note that if $D_{1i} = 1$, then $T_i^o = X_{1r}$ (uncensored) for some r, and so

$$\begin{aligned} W(T_i^o)(D_{1i} - R_{1i}R_i^{-1}) &= [W(T_i^o)/R_i]\,(R_i - R_{1i}) \\ &= [W(X_{1r})/R(X_{1r})] \times [\#\{X_{2j} :\ X_{2j} \geq X_{1r}\}] \\ &= \sum_{X_{2j}:\ X_{1r} \leq X_{2j}} (W(X_{1r})/R(X_{1r}))\,. \end{aligned}$$

Likewise, if $D_{1i} = 0$, then $T_i^o = X_{2t}$ (uncensored) for some t and

$$\begin{aligned} W(T_i^o)\,(D_{1i} - R_{1i}/R_i) &= -\,[W(X_{2t})/R(X_{2t})] \times [\#\{X_{1i} :\ X_{1i} \geq X_{2t}\}] \\ &= -\sum_{X_{1i}:\ X_{1i} \geq X_{2t}} [W(X_{2t})/R(X_{2t})]\,. \end{aligned}$$

The desired result follows accordingly.

If no censoring is present and $W(t) = R(t)$, then (2.3) is in fact the Mann-Whitney statistic, thus justifying the name "generalized Mann-Whitney statistic."

2.3.3 Expression as a Statistic Belonging to Gill's Class K.

For $j = 1, 2$, let

$$N_j(t) = \sum_{i=1}^{n_j} I(X_{ji} \leq t, \Delta_{ji} = 1).$$

Then a WL statistic can be written as

$$\int_0^\infty W(t) \left\{ dN_1(t) - \frac{R_1(t)}{R(t)} dN(t) \right\}. \tag{2.4}$$

Remark, however, that for $R_1(t)R_2(t) > 0$,

$$\begin{aligned} dN_1(t) - dN(t)\frac{R_1(t)}{R(t)} &= \frac{R_2(t)dN_1(t) - dN_2(t)R_1(t)}{R(t)} \\ &= \frac{R_1(t)R_2(t)}{R(t)} \left[\frac{dN_1(t)}{R_1(t)} - \frac{dN_2(t)}{R_2(t)} \right]. \end{aligned}$$

Hence, (2.4) is equal to

$$\int_0^\infty K(t) \left[\frac{dN_1(t)}{R_1(t)} - \frac{dN_2(t)}{R_2(t)} \right] = \int_0^\infty K(t) d\left\{ \hat{\Lambda}_1(t) - \hat{\Lambda}_2(t) \right\}, \tag{2.5}$$

where

$$K(t) = \frac{R_1(t)R_2(t)W(t)}{R(t)}$$

and

$$\hat{\Lambda}_i(t) = \int_0^t \frac{dN_i(s)}{R_i(s)}, \quad i = 1, 2.$$

Of course,

$$\hat{\Lambda}_i(t) \xrightarrow{P} \Lambda_i(t) = \int_0^t \lambda_i(u)\, du,$$

as $n \to \infty$, where $\lambda_i(u)$ is the hazard function for sample i.

Gill (1980) investigated the properties of two-sample tests based on statistics of the form (2.5), in which $K(t)$ is a possibly random function determined by

$$\{N_j(s^-), R_j(s): \ s \leq t, \ j = 1, 2\}$$

and required to be zero whenever $R_1(t)R_2(t) = 0$.

2.4 Generation of Censored Data Counterparts of a Linear Rank Statistic.

One particular subset of the family of classical linear rank statistics for uncensored data comprises statistics of the form

$$S_n = \sum_{i=1}^{n} D_{1i}\phi(i/n),$$

where $\phi(i/n)$ is a nonrandom score attached to T_i^o, and where ϕ is a nonrandom function defined on $[0,1]$ and assumed to satisfy

$$\int_0^1 \phi(u)\, du = 0.$$

Now, under H_0,

$$i/n \xrightarrow{P} F(t_i^o)$$

as $n \to \infty$, where t_i^o is the realized value of T_i^o. For censored samples, therefore, it seems only natural to prescore an uncensored failure time T_i^o as $\tilde{F}_{KM1}(t_i^o)$, where

$$\tilde{F}_{KM1}(t) = 1 - \prod_{j:\ j \le N(t^-)+1} (1 - R_j^{-1})$$

is the Kaplan-Meier (1958) estimator of $F(t)$ under H_0 based on the data from both samples. We then apply the function ϕ to obtain the score

$$Q_i = \phi(\tilde{F}_{KM1}(t_i^o)) = \phi(\tilde{F}_{KM1}(T_i^o))$$

(c.f. Gu, Lai, and Lan 1991; Prentice and Marek 1979). Once we have generated $Q_1, \ldots, Q_K$, we can calculate $Q_1^*, \ldots, Q_K^*$ via equation (2.2), thus producing a statistic of the form of L_{2n}, which in turn is expressible as a WL statistic. Here,

$$Q_i^* = \sum_{k=1}^{i} \left[\left(\prod_{j=k+1}^{i} \frac{-R_j}{1-R_j} \right) \frac{Q_k}{1-R_k} \right],$$

with

$$\prod_{j=i+1}^{i} \left(\frac{-R_j}{1-R_j} \right)$$

set equal to one. If $R_K = 1$, we set $Q_K^* = 0$.

In actual practice, the Kaplan-Meier estimator is written as

$$\tilde{F}_{KM2}(t) = 1 - \prod_{j\ \ j \leq N(t)} (1 - R_j^{-1}); \tag{2.6}$$

however, the above censored data statistic, with $\tilde{F}_{KM2}(t)$ taking the place of $\tilde{F}_{KM1}(t)$, has a weight function which is dependent on $N(t)$ and so does not belong to the class of WL statistics. $\tilde{F}_{KM1}(t)$, though, is asymptotically equivalent to $\tilde{F}_{KM2}(t)$. Therefore, from now throughout the remainder of this thesis, whenever a weight function of a WL statistic depends on the Kaplan-Meier estimator of $F(t)$ based on the combined sample of size n, we shall employ definition (2.6). More generally, whenever we make any sort of reference to this estimator of $F(t)$, we shall assume definition (2.6) is in effect. (A similar convention will be invoked for the Kaplan-Meier estimator of $F_i(t)$ based on the data from sample i,$i = 1, 2$.) Variants of (2.6), which can be used for prescoring purposes, are the Peto-Peto (1972) estimator

$$\tilde{F}_{PP}(t) = \frac{1}{2} \left\{ \tilde{F}_{KM2}(t^-) + \tilde{F}_{KM2}(t^+) \right\}, \tag{2.7}$$

Prentice's (1978) moment estimator defined by

$$\tilde{F}_P(t) = 1 - \prod_{j:\ j \leq N(t)} \left[\frac{R_j}{R_j + 1} \right], \tag{2.8}$$

and Altshuler's (1970) estimator

$$\tilde{F}_A(t) = 1 - \exp\left(- \sum_{j:\ j \leq N(t)} \frac{1}{R_j} \right). \tag{2.9}$$

We comment here that of the above four estimators of $F(t)$, only (2.6) prescores T_i^o as i/n with censoring absent; hence, of the four censored data counterparts, only the one implementing $\tilde{F}_{KM2}(t)$ reduces to S_n with no censoring present.

Another way in which we can extend S_n to accommodate censoring is via the methodology of Gu, Lai, and Lan (1991): To T_i^o, we once again assign the score $Q_i = \phi(\tilde{F}_{KM2}(t_i^o))$. To obtain scores corresponding to the censored failure times in $[T_i^o, T_{i+1}^o)$, first let $T_{(i,1)}, \ldots, T_{(i,M_i)}$ be, in some arbitrary order, the censored lifetimes falling in this interval. Since, for uncensored data,

$$\phi(i/n) \xrightarrow{P} \phi(F(t_i^o))$$

under H_0, the score assigned to $T_{(i,j)}$, denoted by $Q^*_{i,j}$, is such that under H_0

$$Q^*_{i,j} \xrightarrow{P} E(\phi(F(T))|T \geq t_{(i,j)}) = \int_{F(t_{(i,j)})}^{1} \phi(u)\, du/(1 - F(t_{(i,j)})),$$

where T is a random variable having d.f. $F(t)$, and where $t_{(i,j)}$ is the realized value of $T_{(i,j)}$. Hence, we should define Q^*_{ij} as

$$Q^*_{i,j} = \int_{\tilde{F}_{KM2}(t_{(i,j)})}^{1} \phi(u)\, du/(1 - \tilde{F}_{KM2}(t_{(i,j)})) = \Phi(\tilde{F}_{KM2}(t_{(i,j)})),$$

where

$$\Phi(u) = \int_{u}^{1} \phi(s)\, ds/(1-u).$$

Note that $\Phi(u)$ is in fact the mean value of $\phi(s)$ over $[u, 1]$. The resulting censored data generalization of S_n is

$$S_n^* = \sum_{i=1}^{K} \left[D_{1i}\phi(\tilde{F}_{KM2}(t_i^o)) + \sum_{j=1}^{M_i} D_{1,(i,j)}\Phi(\tilde{F}_{KM2}(t_{(i,j)})) \right]$$

$$+ \sum_{j=1}^{M_0} D_{1,(0,j)}\Phi(\tilde{F}_{KM2}(t_{(0,j)}))$$

$$= \sum_{i=1}^{K} \left[D_{1i}\phi(\tilde{F}_{KM2}(T_i^o)) + M_{1i}\Phi(\tilde{F}_{KM2}(T_i^o)) \right],$$

where $D_{1,(i,j)}$ is 1 or 0 according to whether the item censored at $T_{(i,j)}$ is from sample 1 or 2. Now S_n^*, in general, cannot be written as a WL statistic; however, as will be demonstrated in chapter 4, S_n^* is asymptotically equivalent to

$$\sum_{i=1}^{K} \left[\phi(\tilde{F}_{KM2}(T_i^o)) - \Phi(\tilde{F}_{KM2}(T_i^o)) \right] (D_{1i} - R_{1i}R_i^{-1}),$$

assuming ϕ satisfies certain conditions. Moreover, this asymptotic equivalence remains valid if $\tilde{F}_{KM2}(t)$ is replaced by any one of (2.7), (2.8), or (2.9).

2.5 Examples of Some Classical Tests.

2.5.1 The Log-Rank Test.

The log-rank test has weights $W_i = 1$, and scores

$$Q_i = 1 - \sum_{j=1}^{i} (1/R_j),$$

$$Q_i^* = - \sum_{j=1}^{i} (1/R_j),$$

$i = 1, \ldots, K$. Mantel (1966) and Cox (1972) developed this SWL statistic in the form

$$\frac{\sum_{i=1}^{K} (D_{1i} - R_{1i}R_i^{-1})}{\sqrt{\sum_{i=1}^{K} (R_{1i}R_i^{-2}R_{2i})}}. \tag{2.10}$$

Mantel did so using the contingency table approach, while Cox derived (2.10) as a score test within the context of the proportional hazards model.

Peto and Peto (1972) generated the log-rank statistic in the sum of scores representation using maximum likelihood techniques. Prentice (1978) yielded the log-rank procedure in the form

$$\frac{\sum_{i=1}^{K}(D_{1i}Q_i + M_{1i}Q_i^*)}{\sqrt{\sum_{i=1}^{K}(R_{1i}R_i^{-2}R_{2i})}}$$

as a score test arising from the accelerated lifetime model.

In the absence of censoring, the log-rank procedure reduces to the Savage (or exponential scores) test (Savage, 1956).

2.5.2 The Generalized Wilcoxon Tests.

The first known censored data counterpart of the Wilcoxon procedure was constructed by Gehan (1965). The Gehan test has weights $W_i = R_i$ and scores $Q_i = (R_i - i)/n$, $Q_i^* = -i/n$ $(i = 1, \ldots, K)$. Actually, Gehan wrote the numerator of his test in the form (2.3) with $W(t) = R(t)/n$. Mantel (1967) then expressed Gehan's representation as a sum of scores; afterwards, Tarone and Ware (1977) wrote Mantel's form as a WL statistic.

Gehan made the assumption that both samples have equal censoring patterns and proposed, as a variance estimator, a complicated version of V_p different from either of its two given forms in section 2.2. Mantel then wrote Gehan's form of V_p as

$$\frac{n_1 n_2}{n(n-1)} \sum_{i=1}^{K} \left(Q_i^2 + M_i Q_i^{*2}\right).$$

Breslow (1970) suggested, under a random censorship model, a variance estimator that is valid even when the censoring patterns are unequal and which, for large samples, is approximately equal to V_{c_P}. Finally, Tarone and Ware (1977) suggested a variance estimator which has the exact form of V_{c_P}.

The second generalized Wilcoxon procedure to be generated was the Efron (1967) test. In this case,

$$W_i = n^{-1} n_1 n_2 I(R_{1i} R_{2i} > 0) \left(1 - \tilde{F}_{1KM}(t_i^{o-})\right) \times \left(1 - \tilde{F}_{2KM}(t_i^{o-})\right) R_i R_{1i}^{-1} R_{2i}^{-1}, \tag{2.11}$$

where $\tilde{F}_{jKM}(t)$ is the Kaplan-Meier estimator of $F_j(t)$ based on

$$(X_{j1}, \Delta_{j1}), \ldots (X_{jn_j}, \Delta_{jn_j}); \quad j = 1, 2.$$

In terms of the sum of scores format,

$$Q_i = W_i - n^{-1} n_1 n_2 \sum_{j=1}^{i} \left[\left\{1 - \tilde{F}_{1KM}(t_j^{o-})\right\} \left\{1 - \tilde{F}_{2KM}(t_j^{o-})\right\} \times R_{1j}^{-1} R_{2j}^{-1} I(R_{1j} R_{2j} > 0) \right]$$

and

$$Q_i^* = -n^{-1} n_1 n_2 \sum_{j=1}^{i} \left[\left\{1 - \tilde{F}_{1KM}(t_j^{o-})\right\} \left\{1 - \tilde{F}_{2KM}(t_j^{o-})\right\} R_{1j}^{-1} \times R_{2j}^{-1} I(R_{1j} R_{2j} > 0) \right].$$

In actual fact, the two-sample test proposed by Efron was not based on a WL statistic but rather on

$$2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} V_{ij} - 1,$$

where $V_{ij} = 1,\ 0$ whenever $U_{ij} = 1,\ -1$ for Gehan's statistic, and where V_{ij} is an estimator of

$$\Pr[T_{1i} < T_{2j} | (X_{2i}, \Delta_{2i}), (X_{2j}, \Delta_{2j})]$$

whenever $U_{i,j} = 0$ for Gehan's statistic. (Efron assumed here that a random censorship model is in effect.) With V_{ij} defined as such,

$$\begin{aligned} V &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} V_{ij} \\ &= \int_0^{\infty} \left(1 - \tilde{F}_{2KM}(t^-)\right) I(R_2(t) > 0)\, d((1 - \tilde{F}_{1KM}(t)) I(R_1(t^+) > 0)) \end{aligned}$$

(Miller, 1981, p. 106). Under the condition that

$$\sup\{t : \Pr(R_i(t) > 0)\} = \sup\{t : F_i(t) < 1\}, \quad i = 1, 2,$$

V can thus be considered as an estimator of $\Pr(T_1 < T_2)$, where T_1 and T_2 are independent random variables with d.f.'s $F_1(t)$ and $F_2(t)$ respectively. So, under H_0, V should approximately equal 1/2. Although

$$n^{-1} n_1 n_2 (2V - 1)$$

cannot generally be written as a WL statistic, it is asymptotically equivalent to

$$\sum_{i=1}^{K} W_i (D_{1i} - R_{1i} R_i^{-1})$$

with W_i defined as in (2.11), under the condition that there are no ties between the X_{ji}'s (Gill, 1980).

Because of its dependence on $R_1(t)$ and $R_2(t)$, the weight function of Gehan's and Efron's statistic is inexorably linked to the intensity of censoring in both samples. This property is highly undesirable since censoring intensity provides little information–due to independent censoring being in effect about differences or lack thereof between $F_1(t)$ and $F_2(t)$. A

Wilcoxon analogue, which is less affected in this regard, has the scores

$$Q_i = 1 - 2\tilde{F}_{PP}(t_i^o), \qquad Q_i^* = -\tilde{F}_{PP}(t_i^o)$$

(Peto and Peto 1972). The resulting linear generalized rank statistic, however, cannot generally be expressed as a WL statistic. Such is not the case, though, with the statistic comprising scores

$$Q_i = 1 - 2\tilde{F}_P(t_i^o), \qquad Q_i^* = -\tilde{F}_P(t_i^o)$$

(Prentice 1978). The corresponding WL statistic has weight

$$W_i = 1 - \tilde{F}_P(t_i^o) = \prod_{j=1}^{i} \left(\frac{R_j}{R_j + 1} \right).$$

Remark that the random variable $\tilde{F}_P(t_i^o)$ is weakly influenced by the rate at which censoring events occur in each sample. Prentice originally developed his SWL statistic in the form

$$\frac{\sum_{i=1}^{K} (Q_i D_{1i} + Q_i^* M_{1i})}{\sqrt{\sum_{i=1}^{K} \left\{ W_i(1 - A_i)B_i - (A_i - W_i)B_i(W_i B_i + 2\sum_{j=1+i}^{K} W_j B_j) \right\}}}$$

as a score test arising from a log-linear regression model with the error variable having a logistic distribution. Here,

$$A_i = \prod_{j=1}^{i} \left(\frac{R_j + 1}{R_j + 2} \right), \qquad B_i = 2D_{1i} + M_{1i}.$$

For future reference, the weight $W_i = 1 - \tilde{F}_{KM2}(t_i^o)$ shall be known as the Peto-Peto weight, even though the actual statistic Peto and Peto developed is not expressible as a WL statistic.

2.5.3 The Tarone-Ware Class of Tests.

Tarone and Ware (1977) proposed a class of procedures that include, besides the log-rank and Gehan's test, SWL statistics which offer a compromise between the two. The weight function associated with this class is $W(t) = g(R(t)/n)$, where g is a nonrandom function defined on $[0, 1]$. Tarone and Ware conjectured that if $g(u)$ takes on intermediate values between u and 1 for all $u \in [0, 1]$, the resulting test will maintain good power across a wider range of alternatives than the other two procedures.

The variance estimator employed in the development of this class of procedures was V_{cp}.

2.5.4 The Harrington-Fleming Class of Tests.

Harrington and Fleming (1982) suggested a family of tests with weight functions defined as

$$W(t) = \left[1 - \tilde{F}_{KM2}(t)\right]^{\rho}, \tag{2.12}$$

where $\rho \geq 0$ and nonrandom. Note that (2.12) generalizes the log-rank and Peto-Peto weight functions. The originators of this class of SWL statistics utilized V_{cp} as the variance estimator.

CHAPTER 3
THE PROPORTIONAL HAZARDS MODEL AND SWL STATISTICS

3.1 Introduction.

Consider n items (not necessarily involved in a two-sample scenario) to have been placed on test in a survival study at time 0, and suppose associated with the i^{th} item is a column vector of p fixed covariates[5]

$$\mathbf{z}_i = (z_{1i}, \ldots, z_{pi})^T, \quad i = 1, \ldots, n.$$

Let T_i be the failure time variate corresponding to item i, and let $h_0(t)$ be an unspecified hazard function for the standard set of conditions $\mathbf{z}_i = \mathbf{0}$. The proportional hazards model, as proposed by Cox (1972), defines the conditional hazard function for item i, given $\mathbf{z}_i$, as

$$\lim_{h \to 0+} \frac{\Pr(T_i < t + h | T_i \geq t, \mathbf{z}_i)}{h} = h_i(t|\mathbf{z}_i) = \exp(\boldsymbol{\beta}\mathbf{z}_i) h_0(t), \tag{3.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is a row vector of p regression parameters. The nonrandom conditional density function and conditional survival function[6] of T_i, given $\mathbf{z}_i$, are thus respectively

$$f_i(t|\mathbf{z}_i) = \exp(\boldsymbol{\beta}\mathbf{z}_i) h_0(t) [S_0(t)]^{\exp(\boldsymbol{\beta}\mathbf{z}_i)}$$

---

[5] For the remainder of this thesis, the term "fixed," when describing a covariate, shall mean "independent of time and nonrandom." Moreover, the word "covariate" shall from hereon be used interchangeably with the terms "explanatory variable" and "regressor variable."

[6] By "survival function", we mean the complement of the d.f. in question.

and

$$S_i(t|\mathbf{z}_i) = [S_0(t)]^{\exp(\boldsymbol{\beta}\mathbf{z}_i)},$$

where

$$S_0(t) = \exp[-\int_0^t h_0(u)\,du].$$

Thus, the survival function for an item with covariate, **z**, is obtained by raising the baseline survival function $S_0(t)$ to a power. The family of models produced by this process is sometimes referred to as the family of Lehmann alternatives (Lehmann, 1953).

The proportional hazards model adapts readily to the inclusion of random, time dependent covariates. There are several types of such regressor variables; the one of concern to us–introduced by Oakes (1981)–is the evolutionary covariate. Let $H(t)$ denote the history of failures, censorings, and of all other random features of the study up to but not including time t. Then we shall call $Z(t)$ an evolutionary covariate if it is a function of $H(t)$ only. Thus $Z(t)$ could be the number of items at risk at t^-, the number of failures before t, or, in a comparison of two groups, the number of failures in one group before time t. The proportional hazards model, with a single evolutionary explanatory variable $Z_i(t)$, becomes therefore

$$\lim_{h \to 0+} \frac{\Pr(T_i \in [t, t+h) | T_i \geq t,\ V_i(t))}{h}$$
$$= h_i(t|V_i(t)) = h_0(t)\exp(\beta Z_i(t)), \quad i = 1, \ldots, n, \qquad (3.2)$$

where $V_i(t) = \{Z_i(s) : s \leq t\}$, and where "$h_i(t|V_i(t))$" is read as "the conditional hazard function of item i, given $V_i(t)$."

The proportional hazards model, with either fixed or evolutionary covariates, can be applied to the two-sample problem in the following manner.

Let n_1, n_2 be the sizes of samples 1 and 2 respectively. Let the items of sample 1 be denoted by labels 1, 2, ..., n_1, and let those of sample 2 be denoted by labels $n_1 + 1$, $n_1 + 2$, ..., n, where $n = n_1 + n_2$. Denote the hazard functions corresponding to samples 1 and 2 by, respectively, $h_1^*(t)$ and $h_2^*(t)$. Let $h_2^*(t) = h_0(t)$, where $h_0(t)$ is completely unspecified, and let $h_1^*(t) = h_0(t)e^{\beta}$, where β is some unknown parameter. If we assign to item i the fixed regressor variable z_i defined by

$$z_i = \begin{cases} 1 & \text{if } i \in \text{sample 1}, \\ 0 & \text{if } i \in \text{sample 2}, \end{cases} \tag{3.3}$$

then the conditional hazard of T_i, given z_i, can be directly obtained from the model

$$h_i(t|z_i) = e^{\beta z_i} h_0(t). \tag{3.4}$$

Consider, now, the above two-sample setting with the following modification. Let $W(t)$ be an evolutionary covariate, and let $h_i^*(t|U(t))$ denote the conditional hazard function for sample i, given $U(t)$, where $U(t) = \{W(s) : s \leq t\}$. Suppose

$$h_1^*(t|U(t)) = h_0(t)e^{\beta W(t)},$$

and suppose

$$h_2^*(t|U(t)) = h_2^*(t) = h_0(t),$$

where $h_0(t)$ is free of $U(t)$ and unspecified. If we associate with item i the stochastic explanatory variable $Z_i(t)$ defined as

$$Z_i(t) = \begin{cases} W(t) & \text{if } i \in \text{sample 1}, \\ 0 & \text{if } i \in \text{sample 2}, \end{cases} \tag{3.5}$$

then the conditional hazard of T_i, given $V_i(t)$, is described by model (3.2) (Lustbader, 1980; Oakes, 1981).

In the former and latter two-sample scenarios, the null hypothesis

$$H_0: \; h_2^*(t) = h_1^*(t|U(t)) = h_1^*(t) = h_0(t)$$

can thus be equivalently formulated as

$$H_0: \; \beta = 0$$

in, respectively, (3.4) and (3.2).

The aim of this chapter, therefore, is to demonstrate how SWL statistics can be derived within the context of a proportional hazards model that incorporates either a fixed or an evolutionary covariate. Towards this goal, we will always first construct a likelihood function–not restricting ourselves to the two-sample problem–from which valid inference about β can be made. After obtaining an appropriate likelihood, we will appeal to either (3.3) or (3.5) and then generate a test of H_0 via the first and second derivatives of the log likelihood.

3.2 Case of Fixed Covariates.

3.2.1 Likelihood Considerations.

3.2.1.1 Partial Likelihood.

We first present the method of partial likelihood–due to Cox (1975)–in a generalized setting and so do not limit ourselves to failure time data. Suppose the data, denoted simply as $\mathbf{Y}$, have joint density or probability function $f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_1$ is the vector of parameters of interest and $\boldsymbol{\theta}_2$ is the vector of nuisance parameters. One or more of the components of $\boldsymbol{\theta}_2$ may even be nuisance functions. Suppose that $\mathbf{Y}$ can be transformed into the sequence of pairs of variables

$$(X_1, S_1, X_2, S_2, \ldots, X_m, S_m) \tag{3.6}$$

in a one-to-one manner, where the number of pairs of terms m may in some cases be random. The full likelihood of (3.6) can be written as

$$\prod_{j=1}^{m} f_{X_j|\mathbf{X}^{(j-1)},\mathbf{S}^{(j-1)}}(x_j|\mathbf{x}^{(j-1)},\mathbf{s}^{(j-1)};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$$
$$\times \prod_{j=1}^{m} f_{S_j|\mathbf{X}^{j},\mathbf{S}^{(j-1)}}(s_j|\mathbf{x}^{(j)},\mathbf{s}^{(j-1)};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2), \quad (3.7)$$

where $\mathbf{X}^{(j)} = (X_1,\ldots,X_j)$, $\mathbf{S}^{(j)} = (S_1,\ldots,S_j)$, and $\mathbf{X}^{(0)},\mathbf{S}^{(0)} = \emptyset$.

If the second product of (3.7) is a function of $\boldsymbol{\theta}_1$ only, it is called the *partial likelihood* of $\boldsymbol{\theta}_1$ based on $\mathbf{S} = (S_1,\ldots,S_m)$ in the sequence $\{X_j, S_j\}$. In certain applications, one can argue that any information on $\boldsymbol{\theta}_1$ in the first product is inextricably linked with information on $\boldsymbol{\theta}_2$, and so for simplification we take for inference the partial likelihood which involves only $\boldsymbol{\theta}_1$. There will typically be some loss of information involved in using a partial likelihood; in many situations, however, heuristic arguments can be put forth which suggest little is lost in ignoring the first term of (3.7).

Consider now a survival study in which n items are put on test and the data for the i^{th} item, with lifetime variate T_i, are (Y_i,Δ_i,z_i). Here, Y_i is the survival time variate, Δ_i is a censoring indicator variable ($\Delta_i = 0$ if $Y_i < T_i$; $\Delta_i = 1$ if $Y_i = T_i$), and z_i is a fixed scalar covariate. Moreover, suppose the conditional hazard function of T_i, given z_i, is determined by model (3.4). As far as unknown parameters are concerned, the full likelihood of

$$\{(Y_i,\Delta_i,z_i):\ i=1,\ldots,n\}$$

is dependent on β, $h_0(t)$, and possibly one or more nuisance functions associated with the censoring mechanism. The problem at hand then is to make useful inference about β in the presence of the other unknown quantities.

Suppose in the above study that K items, with labels $S_1, \ldots, S_K$, give rise to ordered uncensored failure times $T_1^o < T_2^o < \cdots < T_K^o$ (set $T_0^o = 0$ and $T_{K+1}^o = +\infty$). Suppose further that M_i items, labeled $S_{i,1}, \ldots, S_{i,M_i}$, are censored in $[T_i^o, T_{i+1}^o)$ at times $T_{i,1}^o \leq T_{i,2}^o \leq \cdots \leq T_{i,M_i}$. For $i = 0, \ldots, K$, therefore, take

$$X_{i+1} = \left\{ z_l, S_{i,j}, T_{i,j}^o, T_{i+1}^o : j = 1, \ldots, M_i; l = 1, \ldots, n \right\}.$$

Finally, set $S_{K+1} = n+1$. The complete data, as described in the previous paragraph, can be rewritten as

$$(X_1, S_1, X_2, S_2, \ldots, X_K, S_K, X_{K+1}, S_{K+1}),$$

the likelihood of which is

$$\text{Lik} = \prod_{i=1}^{K+1} f_{X_i|\mathbf{X}^{(i-1)},\mathbf{S}^{(i-1)}}(x_i|\mathbf{x}^{(i-1)}, \mathbf{s}^{(i-1)}) \prod_{i=1}^{K} \Pr(S_i = (i)|\mathbf{X}^{(i)}, \mathbf{S}^{(i-1)}).$$

Note that $\Pr(S_{K+1} = n+1|\mathbf{X}^{(K+1)}, \mathbf{S}^{(K)}) = 1$. Now, let $H_i = (\mathbf{X}^{(i)}, \mathbf{S}^{(i-1)})$, and let $H(t)$ record the history of the study up to but not including t as well as all covariate values; hence $H_i = (H(T_i^o), T_i^o)$, and so Lik can be written as

$$\begin{aligned} &\prod_{i=1}^{K+1} f_{X_i|H_{i-1},S_{i-1}}(x_i|h_{i-1}, (i-1)) \prod_{i=1}^{K} \Pr(S_i = (i)|H_i) \\ &\quad = \prod_{i=1}^{K+1} f_{X_i|H_{i-1},S_{i-1}}(x_i|h_{i-1}, (i-1)) \prod_{i=1}^{K} \Pr(S_i = (i)|H(T_i^o), T_i^o), \end{aligned} \tag{3.8}$$

where $H_0 = \emptyset$.

We now make two assumptions which will enable us to evaluate the second product of (3.8):

Assumption 1: Any item which is at risk at t^- cannot be censored in the interval $[t, t+dt)$. Censoring can only occur at $t + dt$.

Assumption 2: If item i is at risk at t^-, then

$$f_{T_i|H(t)}(t|h(t)) = h_i(t|z_i),$$

where "$f_{T_i|H(t)}(t|h(t))$" means the conditional density function of T_i, given $H(t) = h(t)$, evaluated at specifically t. In words, this assumption means that the conditional hazard function of T_i, given z_i, is unaffected by

(1) the fate of the other items in $(0, t)$,

(2) the value of the other covariates, and

(3) the fact that item i was not censored in $(0, t)$.

The last of these three is a direct consequence of independent censoring being in effect.

Now let $\mathcal{R}(t)$ denote the risk set at time t, which is the set of items at risk at t^-. Then from assumptions 1 and 2,

$$\begin{aligned}
\Pr[T_i^o \in [t, t+dt)|H(t)] &= \sum_{l \in \mathcal{R}(t)} \Pr[T_l \in [t, t+dt), \bigcap_{j \in \mathcal{R}(t)-l} (T_l < T_j)|H(t)] \\
&= \sum_{l \in \mathcal{R}(t)} \int_t^{t+dt} f_{T_l|H(t)}(t_l|h(t)) \prod_{j \in \mathcal{R}(t)-l} [\Pr(T_j > t_l|H(t))]\, dt_l \\
&= \sum_{l \in \mathcal{R}(t)} [f_{T_l|H(t)}(t|h(t)) \prod_{j \in \mathcal{R}(t)-l} [\Pr(T_j > t|H(t))]\, dt] \\
&= \sum_{l \in \mathcal{R}(t)} h_l(t|z_l)\, dt. \qquad (3.9)
\end{aligned}$$

We thus have from assumptions 1 and 2, and from (3.9)

$$\Pr(S_i = (i)|H(T_i^o), T_i^o = t_i^o) = \Pr(T_{(i)} = t_i^o|H(T_i^o), T_i^o = t_i^o)$$

$$
\begin{aligned}
&= \frac{f_{H(T_{(i)}),T_{(i)}}(h(t_i^o), t_i^o)}{f_{H(T_i^o),T_i^o}(h(t_i^o), t_i^o)} \\
&= \frac{\Pr(T_{(i)} \in [t_i^o, t_i^o + dt_i^o)|H(t_i^o))}{\Pr(T_i^o \in [t_i^o, t_i^o + dt_i^o)|H(t_i^o))} \\
&= \frac{h_{(i)}(t_i^o|z_{(i)})dt_i^o}{\sum_{l \in \mathcal{R}(t_i^o)} h_l(t_i^o|z_l)dt_i^o} \\
&= \frac{\exp(\beta z_{(i)})}{\sum_{l \in \mathcal{R}(t_i^o)} \exp(\beta z_l)}.
\end{aligned}
$$

Therefore, the second product of (3.8) is

$$
L_1(\beta) = \prod_{i=1}^{K} \frac{\exp(\beta z_{(i)})}{\sum_{l \in \mathcal{R}(t_i^o)} \exp(\beta z_l)}, \tag{3.10}
$$

which is the famous "partial likelihood function" of Cox (1972) with fixed, scalar covariates.

The total likelihood, Lik, can be determined using product integrals:

$$
\text{Lik} = \Pr[H(0) = h(0)] \mathcal{P}_0^\infty \Pr[H(t + dt) = h(t + dt)|H(t)],
$$

where the second term is

$$
\lim_{\substack{m \to \infty \\ \Delta\tau_i \to 0}} \prod_{i=1}^{m} \Pr(H(\tau_{i-1} + \Delta\tau_i) = h(\tau_{i-1} + \Delta\tau_i)|H(\tau_{i-1})),
$$

where the τ_i's are nonrandom, and where $\tau_0 = 0 < \tau_1 < \cdots < \tau_m < \infty$, $\Delta\tau_i = \tau_i - \tau_{i-1}$, and $\tau_m \to \infty$ as $m \to \infty$. But

$$
\begin{aligned}
\Pr[H(t + dt) = h(t + dt)|H(t)] &= \Pr[D_t = d_t, C_t = c_t|H(t)] \\
&= \Pr[D_t = d_t|H(t)] \Pr[C_t = c_t|H(t), D_t],
\end{aligned} \tag{3.11}
$$

where D_t, C_t are the sets of labels associated with individuals that, respectively, fail and are censored in $[t, t + dt)$. It then follows that

$$
\Pr[D_t = d_t|H(t)] = \prod_{l \in d_t} h_l(t|z_l)dt \prod_{l \in \mathcal{R}(t) - d_t} [1 - h_l(t|z_l)dt], \tag{3.12}
$$

where assumptions 1 and 2 have been utilized. Using (3.11) and (3.12), and setting $\Pr[H(0) = h(0)] = 1$, we have

$$\text{Lik} = \mathop{\mathcal{P}}_0^\infty \left\{ \prod_{l \in d_t} h_l(t|z_l)\, dt \prod_{l \in \mathcal{R}(t) - d_t} [1 - h_l(t|z_l)\, dt] \right\} \mathop{\mathcal{P}}_0^\infty \Pr[C_t = c_t | H(t), D_t].$$

The first product integral, apart from differential elements, reduces to

$$\prod_{j=1}^{K} h_{(j)}(t_j^o | z_{(j)}) \exp\left[-\int_0^\infty \sum_{l \in \mathcal{R}(u)} h_l(u|z_l)\, du \right]$$
$$= \prod_{i=1}^{K} \frac{\exp(\beta z_{(i)})}{\sum_{l \in \mathcal{R}(t_i^o)} \exp(\beta z_l)}$$
$$\times \left[\prod_{i=1}^{K} h_0(t_i^o) \sum_{l \in \mathcal{R}(t_i^o)} \exp(\beta z_l) \right] \exp\left[-\int_0^\infty h_0(u) \sum_{l \in \mathcal{R}(u)} \exp(\beta z_l)\, du \right]$$

(Kalbfleisch and Prentice, 1980, p. 121). Hence,

$$\prod_{i=1}^{K+1} f_{X_i | H_{i-1}, S_{i-1}}(x_i | h_{i-1}, (i-1))$$
$$= \prod_{i=1}^{K} \left[h_0(t_i^o) \sum_{l \in \mathcal{R}(t_i^o)} \exp(\beta z_l) \right] \exp\left[-\int_0^\infty h_0(u) \sum_{l \in \mathcal{R}(u)} \exp(\beta z_l)\, du \right]$$
$$\times \mathop{\mathcal{P}}_0^\infty \Pr[C_t = c_t | H(t), D_t], \tag{3.13}$$

which is the portion of the full likelihood being ignored when using strictly $L_1(\beta)$ to make inference about β.

Intuitively, it would appear that if $h_0(t)$ consists of many unknown parameters, the first and second factors of (3.13) contain relatively little information about β. Moreover, if the third factor of (3.13) is free of β (i.e.; if the censoring mechanism is noninformative), no information about this parameter can be extracted from this term. Realistic examples of informative, but independent censoring schemes are difficult to construct. An artificial example is a random censorship model where the potential censoring time for each individual is determined by the failure time of another individual, with the same covariate, who is not included in the actual life-testing experiment and whose conditional hazard function is described by (3.4).

Kalbfleisch and Prentice (1980, p. 109-110) compared the Fisher information in $L_1(\beta)$ to that in Lik, assuming noninformative censoring, for the model

$$h_0(t) = \exp\left[g_1(t)\gamma_1 + \cdots + g_m(t)\gamma_m\right]\lambda_0(t), \tag{3.14}$$

where $\lambda_0(t)$ as well as $g_i(t)$ $(i = 2, \ldots, m)$ are completely known, where $g_1(t) = 1$, and where $(\gamma_1, \ldots, \gamma_m)$ is a vector of unknown parameters. They found that L_1 will be asymptotically fully efficient with respect to Lik for the estimation of β if, for some $\boldsymbol{\theta}$ and $\forall t \in \sup\{t : \Pr(\#\mathfrak{R}(t) > 0) > 0\}$,

$$\lim_{n\to\infty} \frac{E(\sum_{i\in\mathfrak{R}(t)} z_i \exp(z_i\beta))}{E(\sum_{i\in\mathfrak{R}(t)} \exp(z_i\beta))} = \boldsymbol{\theta}(g_1(t), \ldots, g_m(t))^T, \tag{3.15}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ is free of t. Note that if $\beta = 0$ and the censoring mechanism operates in the same manner for all n items, this ratio of expectations is free of t and (3.15) is satisfied; hence, under these circumstances, the score statistic $L_1'(0)$ will have full Pitman efficiency relative to the likelihood-ratio statistic (Kalbfleisch and Prentice, 1980, p. 106).

3.2.1.2 Marginal Likelihood Approach of Kalbfleisch and Prentice.

Another manner in which L_1 can be derived is by modification of the marginal distribution of the ranks with censoring absent. Suppose for the moment, then, that all n items of the above survival study are observed to fail. Let $T_1, \ldots, T_n$ be the unordered lifetimes, and let $z_1, \ldots, z_n$ be the corresponding covariates. Denote by **O** the order statistic $(T_1^o, \ldots, T_n^o)$, and denote by **R** the "rank statistic" $(\mathfrak{I}_1, \ldots, \mathfrak{I}_n)$, where $\mathfrak{I}_i$ is the label attached to T_i^o. More correctly, **R** is the vector of anti-ra . For convenience, though, we will continue to refer to **R** by the former term or by the term "rank vector."

Define $U_i = g^{-1}(T_i)$ $(i = 1, \ldots, n)$, where $g \in G$, the group of strictly increasing differentiable transformations of $(0, \infty)$ onto $(0, \infty)$. Then the conditional hazard of U_i, given z_i, is

$$\lambda_0(u) \exp[z_i \beta],$$

where $\lambda_0(u) = h_0[g(u)]g'(u)$. Thus, if the data were presented in the form

$$(U_1, z_1), \ldots, (U_n, z_n),$$

the inference problem about β would remain unchanged provided $h_0(t)$ were completely unknown. In effect, the estimation problem for β is invariant under the group G of transformations on the T_i's. We also note that when a member of G operates on $T_1, \ldots, T_n$, it acts transitively on the order statistic **O** while leaving **R** invariant. Finally, the homomorphic group H, acting on the parameter space, is transitive on $h_0(\cdot)$ and leaves the regression parameter β invariant. Therefore, by the definition of Barnard (1963), the rank statistic **R** is marginally sufficient for the estimation of β, that is 'sufficient for β in the absence of knowledge of $h_0(t)$.' For inference about β, the probability function of **R** is available and is given by

$$\Pr\{\mathbf{R} = [(1), \ldots, (n)]\} = \int_0^\infty \int_{t_1^o}^\infty \int_{t_2^o}^\infty \cdots \int_{t_{n-1}^o}^\infty \prod_{i=1}^n f_{(i)}(t_i^o | z_{(i)})\, dt_n^o \cdots dt_1^o$$

$$= \frac{\exp\left[\sum_{i=1}^n z_i \beta\right]}{\prod_{i=1}^n \left[\sum_{j=i}^n \exp(\beta z_{(j)})\right]}. \tag{3.16}$$

By the terminology of Fraser (1968), and Kalbfleisch and Sprott (1970), (3.16) is called the marginal likelihood of β.

If censoring is present, some modification of the above argument is required. When a censored sample is obtained, the rank vector of the underlying T_i's is only partially observed. For example, suppose that items 1, 2, 3, 4 yield survival times 40, 10, 20*, 30*, respectively, where the asterisk indicates a censored lifetime. The rank statistic, on the basis of this data, is known to be one of the following:

$$(2,1,3,4); (2,1,4,3); (2,3,1,4); (2,3,4,1); (2,4,1,3); (2,4,3,1).$$

In order to make an inference about β, we calculate the probability that the underlying rank vector is any one of these (Kalbfleisch and Prentice, 1973). This probability is the sum of six terms of the type (3.16). Remark that this approach ignores the ordering of censored lifetimes between successive failures; however, the fact that independent censoring is in effect suggests that little information is lost in this restriction (see pp. 11-12).

Suppose then that K items, with labels $(1), \ldots, (K)$, give rise to ordered uncensored event times $T_1^o < T_2^o < \cdots < T_K^o$, and suppose further that M_j items, labeled $(j,1), \ldots, (j, M_j)$ in some arbitrary order, are censored in

$$[T_j^o, T_{j+1}^o) \quad (j = 0, \ldots, K),$$

where $T_0^o = 0$ and $T_{K+1}^o = \infty$. The rank vector generalized to censored data, as defined by Prentice (1978), is

$$\mathbf{R}_G = ((1), \ldots, (K); \{(j,1), \ldots, (j, M_j)\}, j = 0, \ldots, K),$$

which is in fact an equivalent form of the generalized rank vector as defined in the introduction. The marginal likelihood of β is computed as the probability of obtaining all possible underlying rank vectors in the uncensored experiment which are consistent with $\mathbf{R}_G$. This probability (defining $T_{(0)} = 0$) is

$$\Pr\left\{T_{(1)} < T_{(2)} < \cdots < T_{(K)}; (T_{(i)} < T_{(i,j)}), i = 0, \ldots, K; j = 1, \ldots, M_i | \mathbf{R}_G\right\}$$
$$= \int_0^\infty \int_{v_1}^\infty \cdots \int_{v_{K-1}}^\infty \prod_{j=1}^{K} \left\{ f_{(j)}(v_j|z_{(j)}) \prod_{l=1}^{M_j} S_{(j,l)}(v_j|z_{(j,l)}) \right\} dv_K \ldots dv_1$$
$$= \int_0^\infty \int_{v_1}^\infty \cdots \int_{v_{K-1}}^\infty$$
$$\prod_{i=1}^{K} \left(e^{\beta z_{(i)}} h_0(v_i) \exp\left[- \left\{ e^{\beta z_{(i)}} + \sum_{j=1}^{M_i} e^{\beta z_{(i,j)}} \right\} \int_0^{v_i} h_0(u)\, du \right] \right) dv_k \cdots dv_1$$
$$= \prod_{i=1}^{K} \left[\frac{\exp(\beta z_{(i)})}{\sum_{l \in \mathcal{R}(T_i^o)} \exp(\beta z_l)} \right],$$

which is Cox's partial likelihood function. Here, $f_i(t|z_i)$ and $S_i(t|z_i)$ are, respectively, the nonrandom conditional density function and conditional survival function of T_i, given z_i.

Recall from subsection 3.2.1.1 that the likelihood of

$$\{(Y_i, \Delta_i, z_i) : \; i = 1, \ldots, n\}$$

–regardless of the conditional d.f. of T_i $(i = 1, \ldots, n)$, given z_i is

$$\text{Lik} = Q \prod_{j=1}^{n} [h_j(Y_j|z_j)]^{\Delta_j} \exp\left[- \int_0^\infty \sum_{l \in \mathcal{R}(u)} h_l(u|z_l)\, du \right]$$
$$= Q \prod_{i=1}^{n} \left([f_i(Y_i|z_i)]^{\Delta_i} [S_i(Y_i|z_i)]^{1-\Delta_i} \right),$$

where Q is the contribution to the likelihood provided by times of censorings and labels of items censored. For type II and progressive type II censoring,

Q is 1 and $\prod_{i=1}^{K} \binom{R_i - 1}{M_i}^{-1}$ respectively, where $R_i = \sum_{j=i}^{K}(M_j + 1)$. Recall that, for progressive type II censoring, M_i and R_i are predetermined; thus, for these two censoring schemes,

$$\text{Lik } \alpha \prod_{i=1}^{n} \left[[f_i(Y_i|z_i)]^{\Delta_i} [S_i(Y_i|z_i)]^{1-\Delta_i} \right]$$
$$= \prod_{j=1}^{K} \left\{ f_{(j)}(T_j^o|z_{(j)}) \prod_{l=1}^{M_j} S_{(j,l)}(T_j^o|z_{(j,l)}) \right\}.$$

Hence, for the above two censoring mechanisms, the probability function of $\mathbf{R}_G$ is given by

$$f_{\mathbf{R}_G}(\mathbf{r}_G) = \int_{T_1^o < \cdots < T_K^o} \cdots \int \text{Lik } dT_K^o \cdots dT_1^o,$$

which, for the special case of model (3.4) equals $CL_1(\beta)$, for some constant C.

3.2.2 The Log-rank Test.

Consider now the two-sample scenario described in section 3.1 as related to model (3.4). Invoking covariate definition (3.3), $\log(L_1(\beta))$ becomes

$$L(\beta) = \log \left[\prod_{i=1}^{K} \left(\frac{\exp(\beta D_{1i})}{R_{1i} e^{\beta} + R_{2i}} \right) \right] = \sum_{i=1}^{K} \left[\beta D_{1i} - \log(R_{1i} e^{\beta} + R_{2i}) \right],$$

where $D_{1i} = 1$ or 0 according as the failure at T_i^o occurs in the first or second group, where R_{ji} is the risk set size of group j at time T_i^o, and where $R_i = R_{1i} + R_{2i}$. The score statistic for $H_0 : \beta = 0$ is therefore

$$U = L'(0) = \sum_{i=1}^{K} \left(D_{1i} - \frac{R_{1i}}{R_i} \right),$$

which is the log-rank statistic. The expectation and variance of U, under H_0, were evaluated in chapter 2 using 2×2 contingency tables (set $W(t)$ of chapter 2 equal to 1). We now evaluate these two employing the properties of a score statistic.

Applying the notation of subsection 3.2.1.1,

$$L(\beta) = \sum_{i=1}^{K} l_i(\beta),$$

where $l_i(\beta) = \log\left(\Pr(S_i = (i)|H_i)\right) = \log\left[\exp(\beta D_{1i})/(R_{1i}e^{\beta} + R_{2i})\right]$. Now, $\Pr(S_i = (i)|H_i)$ is a conditional probability function over $\mathcal{R}(T_i^o)$, and it depends on the parameter β. It follows then just as for any probability or density function that

$$E_{H_0}\left[l_i'(0)|H_i\right] = 0$$

and

$$\mathrm{Var}_{H_0}\left[l_i'(0)|H_i\right] = -E_{H_0}\left[l_i''(0)|H_i\right],$$

for $i = 1, \ldots, K$. Hence, unconditionally under H_0,

$$E_{H_0}\left(l_i'(0)\right) = 0$$

and

$$\begin{aligned}\mathrm{Var}(l_i'(0)) &= E\left(\mathrm{Var}(l_i'(0)|H_i)\right) + \mathrm{Var}\left(E(l_i'(0)|H_i)\right) \\ &= -E\left(E(l_i''(0)|H_i)\right).\end{aligned}$$

The properties of iterated expectations yield one further result. Since H_i and S_i are included in H_j if $i < j$, we have under H_0,

$$\begin{aligned}E\left[l_i'(0)l_j'(0)\right] &= E\left[E(l_i'(0)l_j'(0)|H_j)\right] \\ &= E\left[l_i'(0)E(l_j'(0)|H_j)\right] \\ &= 0.\end{aligned}$$

Hence,

$$\text{Cov}_{H_0}(l_i'(0), l_j'(0)) = 0.$$

Treating K as a fixed quantity if it is in fact a random variable, we have, under H_0, from the above results

$$E(U) = \sum_{i=1}^{K} E(l_i'(0)) = 0$$

and

$$\begin{aligned}
\text{Var}(U) = \text{Var}\left[\sum_{i=1}^{K} l_i'(0)\right] &= \sum_{i=1}^{K} \text{Var}(l_i'(0)) \\
&= E\left\{\sum_{i=1}^{K} E\left[\left(\frac{R_{1i}}{R_i}\right)\left(\frac{R_{2i}}{R_i}\right)\Big|H_i\right]\right\} \\
&= E\left[-L''(0)\right],
\end{aligned}$$

where

$$-L''(0) = \sum_{i=1}^{K} \frac{R_{1i}R_{2i}}{R_i^2}.$$

$-L''(0)$ is, of course, the conditional permutation variance estimator for the log-rank statistic. For large n_1, n_2, and K, $\text{Var}_{H_0}(U)$ can be adequately replaced by $-L''(0)$; thus, the SWL statistic we have generated for testing $H_0 : \beta = 0$ is

$$\frac{U}{\sqrt{-L''(0)}} = \frac{\sum_{i=1}^{K}(D_{1i} - R_{1i}/R_i)}{\sqrt{\sum_{i=1}^{K}\left(\frac{R_{1i}}{R_i}\right)\left(\frac{R_{2i}}{R_i}\right)}},$$

which is the log-rank test.

3.3 Case of Evolutionary Covariates.

3.3.1 Likelihood Considerations.

Since the covariates of concern now are time dependent, the proportional hazards model no longer is invariant under the group of differentiable, strictly increasing transformations; thus, the marginal-likelihood-of-the-ranks approach is inappropriate here. The method of partial likelihood, however, is once again applicable.

As in subsection 3.2.1.1, consider n items to have been placed on test at time 0, and let Y_i, T_i, Δ_i be the survival time variate, failure time variate, and censoring indicator variable corresponding to item i. Moreover, suppose that the evolutionary covariate $Z_i(t)$ is associated with item i. Let $V_i(t)$ denote the covariate path up to and including time t,

$$\{Z_i(u) : \ 0 < u \leq t\}.$$

Then the data for the ith individual are $(Y_i, \Delta_i, V_i(Y_i))$, $i = 1, \ldots, n$. The conditional hazard function of T_i, given $V_i(t)$, is denoted by

$$h_i(t|V_i(t)) = \lim_{h \to 0^+} \Pr(T_i \in [t, t+h)|V_i(t), T_i \geq t)/h,$$

and as a special case of interest we suppose

$$h_i(t|V_i(t)) = \exp(\beta Z_i(t))h_0(t). \tag{3.17}$$

As in subsection 3.2.1.1, we consider K items labeled $S_1, \ldots, S_K$ to give rise to ordered uncensored lifetimes $T_1^o < T_2^o < \cdots < T_K^o$. The remaining $n - K$ lifetimes are right censored. For $i = 0, \ldots, K$, let

$$X_{i+1} = \left\{S_{i,j}, T_{i,j}^0, Z_l(u), T_{i+1}^o \ : \ j = 1, \ldots, M_i; l \in \mathcal{R}(u); T_i^o < u \leq T_{i+1}^o\right\},$$

where $S_{i,j}$, $T_{i,j}^o$, and M_i have the same definition as in 3.2.1.1. In addition, set $S_{K+1} = n + 1$. The data in its original form,

$$(Y_1, \Delta_1, V_1(Y_1)), \ldots, (Y_n, \Delta_n, V_n(Y_n)),$$

is thus equivalent to

$$X_1, S_1, \ldots, X_{K+1}, S_{K+1},$$

the likelihood of which is

$$\text{Lik} = \prod_{i=1}^{K+1} f_{X_i|H_{i-1},S_{i-1}}(x_i|h_{i-1,(i-1)}) \prod_{i=1}^{K} \Pr(S_i = (i)|H(T_i^o), T_i^o, \mathcal{L}(T_i^{o+})).$$

Here, H_i is as previously defined, $H(t)$ is the collection of all censorings, failures, and covariate values in $(0, t)$, while $\mathcal{L}(t)$ describes the path of all covariates up to but not including time t.

The second product of Lik, under assumptions similar to those of p. 38, is

$$\begin{aligned} L_2(\beta) &= \prod_{j=1}^{K} \left[\frac{h_{(j)}(T_j^o|V_{(j)}(T_j^o))}{\sum_{l\in\mathcal{R}(T_j^o)} h_l(T_j^o|V_l(T_j^o))} \right] \\ &= \prod_{i=1}^{K} \left[\frac{\exp(\beta Z_{(i)}(T_i^o))}{\sum_{l\in\mathcal{R}(T_i^o)} \exp(\beta Z_l(T_i^o))} \right], \end{aligned}$$

which is a partial likelihood.

The derivation of the full likelihood, Lik, proceeds as in subsection 3.2.1.1, but the factorization (3.11) of

$$\Pr[H(t+dt) = h(t+dt)|H(t)]$$

is replaced with

$$\begin{aligned} &\Pr[\mathcal{L}(t+dt) = \ell(t+dt)|H(t)] \Pr[D_t = d_t|H(t), \mathcal{L}(t+dt)] \\ &\times \Pr[C_t = c_t|H(t), \mathcal{L}(t+dt), D_t]. \end{aligned} \tag{3.18}$$

The product integral of the first term of (3.18) equals 1, while that of the second term yields

$$\prod_{i=1}^{K} [\exp(\beta Z_{(i)}(T_i^o)) h_0(T_i^o)] \exp\left[-\int_0^\infty \sum_{l\in\mathcal{R}(t)} (\exp(\beta Z_l(t))) h_0(t)\, dt \right] \tag{3.19}$$

under assumptions similar to those stated on p. 38 (Kalbfleisch and Prentice, 1980, p. 127). If censoring is noninformative, the only portion of Lik that is of concern to us is (3.19).

Kalbfleisch and Prentice (1980, p. 141) compared the asymptotic efficiency of $L_2(\beta)$ relative to (3.19) for the parametric model

$$h_0(t) = \lambda_0(t) \exp\left(\sum_{i=1}^{m} (g_i(t)\gamma_i)\right),$$

where $g_1(t) = 1$, and where $\lambda_0(t)$ as well as the remaining $g_i(t)$ are completely specified. They showed that $L_2(\beta)$ is fully efficient if

$$\lim_{n\to\infty} \frac{E\left\{\sum_{i\in\mathcal{R}(t)} Z_i(t)\exp(\beta Z_i(t))\right\}}{E\left\{\sum_{i\in\mathcal{R}(t)} \exp(\beta Z_i(t))\right\}}$$

can be expressed as an exact linear combination of the $g_i(t)$'s.

3.3.2 Generation of SWL Statistics Excluding the Log-Rank Test.

Suppose that the life-testing experiment of subsection 3.3.1 is, more specifically, a two-sample scenario with the covariates defined as in (3.5). We therefore have

$$L(\beta) = \log(L_2(\beta)) = \sum_{i=1}^{K} \{D_{1i}\beta W(T_i^o) - \log[R_{1i}\exp(\beta W(T_i^o)) + R_{2i}]\}.$$

A test for $H_0 : \beta = 0$ can be based on

$$L'(0) = \sum_{i=1}^{K} W_i(D_{1i} - R_{1i}/R_i),$$

where $W_i = W(T_i^o)$. Using arguments similar to those presented in subsection 3.2.2, we can show that, under H_0,

$$E(L'(0)) = 0$$

and

$$\mathrm{Var}(L'(0)) = E(-L''(0)),$$

where $-L''(0) = \sum_{i=1}^{K} \frac{R_{1i} R_{2i} W_i^2}{R_i^2}$. Since, under H_0, $\mathrm{Asvar}_{H_0}(n^{-1/2} L'(0))$ can be consistently estimated by $-L''(0)/n$ (Andersen and Gill, 1982), an appropriate test for $\beta = 0$ under asymptotic conditions is

$$U_n = \frac{L'(0)}{\sqrt{-L''(0)}} = \frac{\sum_{i=1}^{K} W_i (D_{1i} - R_{1i} R_i^{-1})}{\sqrt{\sum_{i=1}^{K} \left(\frac{R_{1i}}{R_i}\right) \left(\frac{R_{2i}}{R_i}\right) W_i^2}}.$$

If we restrict $W(t)$ to being a function of only

$$\left(\left\{D_{1i}, R_{1i}, R_{2i} :\ i \leq N(t^-)\right\}, R_1(t), R_2(t), N(t^-)\right)$$

$\subset H(t)$, U_n is an SWL statistic. Here,

$$R_j(t) = \sum_{i=1}^{n} I(Y_i \geq t, i \in \text{group } j),$$

for $j = 1, 2$, while

$$N(t) = \sum_{i=1}^{n} I(Y_i \leq t, \Delta_i = 1).$$

CHAPTER 4
THE ACCELERATED FAILURE TIME MODEL AND SWL STATISTICS

4.1 Introduction.

Consider a life-testing experiment (not necessarily dealing with the comparison of two samples) in which the failure time variate for the i^{th} item, T_i, is modelled as

$$W_i = \log T_i = \boldsymbol{\beta}\mathbf{z}_i + E_i, \quad i = 1, \ldots, n. \tag{4.1}$$

Here, $\mathbf{z}_i$ is a column vector of p fixed covariates associated with item i, $\boldsymbol{\beta}$ is a row vector of regression coefficients, and E_i is an error variable with nonrandom density f and absolutely continuous, strictly increasing d.f. F. We refer to (4.1) as the *accelerated failure time model.* The data for the i^{th} item are $(Y_i, \Delta_i, \mathbf{z}_i)$, where Y_i is the log survival time and Δ_i is the censoring indicator variable ($\Delta_i = 1$ if $Y_i = W_i$, $\Delta_i = 0$ if $Y_i < W_i$). The principle objective here, of course, is to make inference about $\boldsymbol{\beta}$ from the available data.

Consider, now, a two-sample scenario such that the items of sample 1 are denoted by labels $1, 2, \ldots, n_1$ while those of sample 2 are denoted by labels $n_1 + 1, n_1 + 2, \ldots, n_1 + n_2 = n$, and such that the log event time for item i, $i = 1, \ldots, n$, is defined by (4.1) with

$$\mathbf{z}_i = z_i = \begin{cases} 1 & \text{if } i \in \text{group } 1, \\ 0 & \text{if } i \in \text{group } 2, \end{cases}$$

and with $\boldsymbol{\beta} = \beta$. The null hypothesis is, therefore, $H_0 : \beta = 0$. For the remainder of this chapter, we shall concern ourselves with model (4.1) as it applies to the two-sample problem.

The numerator of the optimal parametric test for $\beta = 0$ is derived, under an assumed f, as follows. Let $W_1^o < \cdots < W_K^o$ represent the ordered uncensored W_j's in the combined sample of size n. In some arbitrary order, let $W_{i,1}^o, \ldots, W_{i,M_i}^o$ denote the censored W_j's in $[W_i^o, W_{i+1}^o)$, for $i = 0, 1, \ldots, K$, with $W_0^o = -\infty$ and $W_{K+1}^o = +\infty$. Also, let (i) and (i,j) represent item labels corresponding to W_i^o and $W_{i,j}^o$ respectively. The log likelihood, under a noninformative, independent censoring scheme, is

$$L(\beta) = \sum_{i=0}^{K} \left(\log f(W_i^o - \beta z_{(i)}) + \sum_{j=1}^{M_i} \log[1 - F(W_{i,j}^o - \beta z_{(i,j)})] \right)$$

(see p. 44). By convention, $\log f(-\infty) = 0$ and any summation $a_1 + \cdots + a_m$ has value zero if $m = 0$. A test for $H_0 : \beta = 0$ utilizes the score statistic

$$l'(0) = \sum_{i=0}^{K} (z_{(i)} Q_i + \sum_{j=1}^{M_i} z_{(i,j)} Q_{i,j}^*), \tag{4.2}$$

where $z_{(0)} Q_0 = 0$, where

$$Q_i = -f'(W_i^o)/f(W_i^o)$$

is a score corresponding to W_i^o, $i = 1, \ldots, K$, and where

$$Q_{i,j}^* = f(W_{i,j}^o)/[1 - F(W_{i,j}^o)]$$

is a score corresponding to $W_{i,j}^o$, $i = 0, 1, \ldots, K$; $j = 1, \ldots, M_i$.

A usual uncertainty concerning the choice of f, and the possibility that a few outlying W_i^o's and/or $W_{i,j}^o$'s may have a dominating effect on (4.2) are important reasons for seeking alternatives to tests based on (4.2). With uncensored data, rank procedures generally possess greater robustness against a wrong choice of f and greater outlier resistance than do their parametric analogues

$$\frac{\sum_{i=1}^{n} z_{(i)} Q_i}{\sqrt{\sum_{i=1}^{n} (z_i^2 E_{H_0}([f'(W_i)/f(W_i)]^2))}}.$$

In section 4.2, we review the construction and properties of these rank procedures since the development of their censored data counterparts is essentially the same. In section 4.3, we generate censored data test statistics by means of the generalized rank vector. Ultimately, we shall demonstrate that these generalized rank procedures can be expressed as SWL statistics.

4.2 Rank Tests with Uncensored Data.

Suppose, in the above notation, that there is no possibility of censoring, so that all $M_i = 0$, and the total sample size is $n = K$. We denote the rank vector by

$$\mathbf{R} = [\mathfrak{I}_1, \ldots, \mathfrak{I}_n],$$

where $\mathfrak{I}_i$ is the label attached to W_i^o. Thus, letting $\mathbf{r} = [(1), \ldots, (n)]$, we have

$$\Pr(\mathbf{R} = \mathbf{r}) = \int \cdots \int_{W_1^o < \cdots < W_n^o} \prod_{i=1}^{n} f(W_i^o - \beta z_{(i)})\, dW_i^o. \tag{4.3}$$

A locally most powerful rank test of $\beta = 0$ can be based on the score statistic from (4.3). Straightforward calculations give

$$V_n^* = \left. \frac{d \log \Pr(\mathbf{R} = \mathbf{r})}{d\beta} \right|_{\beta=0} = \sum_{i=1}^{n} z_{(i)} Q_i, \tag{4.4}$$

where Q_i is a nonrandom score attached to W_i^o and equal to

$$n! \int_{W_1^o < \cdots < W_n^o} \cdots \int \left(-\frac{f'(W_i^o)}{f(W_i^o)} \right) \prod_{j=1}^{n} \{ f(W_j^o)\, dW_j^o \}$$

$$= n! \int_{U_1 < \cdots < U_n} \cdots \int \phi(U_i) \prod_{j=1}^{n} dU_j$$

$$= E\{\phi(U_i)\}. \tag{4.5}$$

Here, $U_i = F(W_i^o)$ is the i^{th} order statistic in a uniform (0,1) sample of size n, and $\phi(u)$, for $0 < u < 1$, is given by

$$\phi(u) = \phi(u, f) = -f'\{F^{-1}(u)\}/f\{F^{-1}(u)\}.$$

Note that

$$\sum_{i=1}^{n} Q_i = \sum_{i=1}^{n} E_{H_0}(-f'(W_i^o)/f(W_i^o)) = \sum_{i=1}^{n} E_{H_0}(-f'(W_i)/f(W_i)) = 0.$$

Hence,

$$E_{H_0}(V_n^*) = \left(\frac{n_1}{n}\right) \sum_{i=1}^{n} Q_i = 0.$$

The fact that U_i has expectation $i(n+1)^{-1}$, for $i = 1, \ldots, n$, leads to an asymptotically equivalent system of scores

$$Q_i = \phi\{i(n+1)^{-1}\}. \tag{4.6}$$

Some interesting special cases of (4.5) and (4.6) are as follows. A logistic density, $f(t) = e^t(1+e^t)^{-2}$, gives Wilcoxon (1945) scores for both (4.5) and (4.6), $Q_i = 2i(n+1)^{-1} - 1$. A standard normal density gives normal scores for (4.5), $Q_i = E_{H_0}(W_i^o)$ (Fisher and Yates, 1963), and van der Waerden (1953) scores for (4.6), $Q_i = G^{-1}\{i(n+1)^{-1}\}$, where G is the standard normal distribution function. Similarly, an extreme value density, $f(t) = \exp(t - e^t)$, yields for (4.5) the exponential scores

$$Q_i = n^{-1} + (n-1)^{-1} + \cdots + (n-i+1)^{-1} - 1$$

of Savage (1956), while the double exponential density, $f(t) = e^{-|t|}/2$, gives for (4.6) sign (median) scores $Q_i = \text{sign}\{2i - (n+1)\}$, with $Q_i = 0$ if $i = (n+1)/2$.

The exact null distribution of V_n^* can be determined without knowledge of f since each of the $n!$ possible realizations of $\mathbf{R}$ are equally likely under H_0. Except in the simplest of problems, though, the computation of this distribution is very laborious. An alternative test procedure results from the fact that $V_n^*/\sqrt{\text{Var}_{H_0}(V_n^*)}$ is asymptotically a standard normal variate under $\beta = 0$ and under mild restrictions on the explanatory variables (Hájek and Sĭdák, 1967, p. 159). Here,

$$\text{Var}_{H_0}(V_n^*) = \sum_{i=1}^{n} Q_i^2 \left[\frac{n_1 n_2}{n(n-1)}\right].$$

Results with contiguous alternatives (Hájek and Sĭdák, 1967, p.268) show that the Pitman asymptotic efficiency of

$$n^{-1/2}V_n^* \Big/ \sqrt{\text{Asvar}_{H_0}(n^{-1/2}V_n^*)}$$

(based on the assumed score generating density f) relative to the optimum parametric test (based on the actual density f_0) will quite generally be given by

$$\frac{\left\{\int_0^1 \phi(u,f)\phi(u,f_0)\,du\right\}^2}{\int_0^1 \phi^2(u,f)\,du \int_0^1 \phi^2(u,f_0)\,du}. \tag{4.7}$$

It is presumed that the Fisher information terms in the denominator are finite. Expression (4.7) reveals that the Pitman ARE is one if f and f_0 agree up to location and scaling. Under $f \neq f_0$, (4.7) typically indicates substantial improvement over the parametric analogue. For example, the normal scores test, under mild conditions on f_0 (Puri and Sen, 1971, p.118), has efficiency equal or greater than that of the least squares test. Under Cauchy sampling, for instance, the latter procedure has efficiency zero, while its rank counterpart has efficiency 0.43. Rank tests themselves differ somewhat in efficiency properties. For example, under Cauchy sampling, the Wilcoxon procedure has efficiency 0.61, while the sign test has an even higher efficiency of 0.81. It is thus important to consider the class of plausible sampling density functions in selecting a rank test.

4.3 Censored Data Analogues of Rank Tests.

4.3.1 Construction of Test Statistics.

To begin with, assume that if a failure and one or more censorings occur at the same instant, the failure is immediately followed by the censorings. Now, recall that Y_i is the log survival time corresponding to item i (i=1, ... ,n); thus, let $Y_1^o < \cdots < Y_r^o$ be the ordered distinct Y_i's, let $\mathfrak{D}_i$ be the item which fails at Y_i^o, and let $\mathfrak{C}_i$ be the set of items which are censored at Y_i^o. $\mathfrak{D}_i(\mathfrak{C}_i) = \emptyset$ if there is (are) no failure (censorings) at Y_i^o. The most comprehensive extension of the rank vector to right censored data considers the set

$$\mathbf{L} = [\mathfrak{D}_1, \mathfrak{C}_1. \ldots, \mathfrak{D}_r, \mathfrak{C}_r]$$

(Peto, 1972). This statistic is the maximal invariant statistic under monotone increasing transformations on the Y_i's, but its sampling distribution depends on functions linked to the censoring mechanism (Crowley, 1974). As a consequence, if the censoring mechanism cannot be precisely identified, **L** cannot be utilized for generating tests of H_0. Moreover, even if the censoring scheme can be exactly defined, **L** will usually not yield an easily derivable two-sample procedure. An alternative approach, as discussed in subsection 3.2.1.2, views the rank vector of the underlying W_i's, which is only partially observed owing to the censoring, to be of primary interest. The "rank vector probability," in this scenario, is taken to be the probability of obtaining all possible rank vectors in the uncensored experiment which are consistent with

$$\mathbf{R}_G = [(1), \ldots, (K); \{(i,1), \ldots, (i, M_i)\}, i = 0, \ldots, K],$$

where (i), (i,j), and M_i have the same definition as in section 4.1. $\mathbf{R}_G$ is, of course, the generalized rank vector. The above probability is given by

$$p(\beta) = \int_{v_1 < \cdots < v_K} \cdots \int \prod_{i=1}^{K} \left\{ f(v_i - \beta z_{(i)}) \prod_{j=1}^{M_i} [1 - F(v_i - \beta z_{(i,j)})] \; dv_i \right\}. \qquad (4.8)$$

Note that at $\beta = 0$, (4.8) can be integrated directly without specifying f and F. The value obtained in this case is

$$\prod_{i=1}^{K} (R_i)^{-1},$$

where $R_i = \sum_{j=i}^{K} (M_j + 1)$.

As in (4.4), a score statistic for testing $\beta = 0$ may be obtained from (4.8) giving

$$V_n = \left. \frac{d(\log p(\beta))}{d\beta} \right|_{\beta=0} = \sum_{i=1}^{K} (z_{(i)} Q_i + M_{1i} Q_i^*), \qquad (4.9)$$

where $M_{1i} = z_{(i,1)} + \cdots + z_{(i,M_i)}$, Q_i is a score corresponding to W_i^o, and Q_i^* is a score corresponding to each of $W_{i,1}^o, \ldots, W_{i,M_i}^o$. (We set $Q_0^* = 0$.) Assuming sufficient regularity that differentiation and integration may be interchanged, the uncensored and censored data scores are respectively

$$Q_i = \int_{v_1 < \cdots < v_K} \cdots \int \left\{ -\frac{d \log f(v_i)}{dv_i} \right\} \prod_{j=1}^{K} \{ R_j (1 - F(v_j))^{M_j} f(v_j)\, dv_j \}$$

$$Q_i^* = \int_{v_1 < \cdots < v_K} \cdots \int \left\{ -\frac{d \log(1 - F(v_i))}{dv_i} \right\} \prod_{j=1}^{K} \{ R_j (1 - F(v_j))^{M_j} f(v_j)\, dv_j \}.$$

The statistic (4.9) is of the same form as its parametric counterpart (4.2), though unlike (4.2), the same score is assigned to $W_{i,1}^o, \ldots, W_{i,M_i}^o$. Johnson and Mehrotra (1972) derived (4.9) for the special case of simple type II censoring.

As in (4.5), theses scores can be expressed in terms of functions on $(0,1)$. Set $u_i = F(v_i)$, $i = 1, \ldots, K$, and define for $0 < u < 1$

$$\phi(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}, \quad \Phi(u) = (1-u)^{-1} f(F^{-1}(u)).$$

(Remark that $\Phi(u) = \int_u^1 \phi(s)\, ds/(1-u)$.) The scoring system can now be written

$$Q_i = \int_{u_1 < \cdots < u_K} \cdots \int \phi(u_i) \prod_{j=1}^{K} [R_j (1-u_j)^{M_j}\, du_j],$$

$$Q_i^* = \int_{u_1 < \cdots < u_K} \cdots \int \Phi(u_i) \prod_{j=1}^{K} [R_j (1-u_j)^{M_j}\, du_j]. \qquad (4.10)$$

In order to list some specific scoring schemes, let

$$J(g(u_i)) = \int_{u_1 < \cdots < u_K} \cdots \int g(u_i) \prod_{j=1}^{K} [R_j (1-u_j)^{M_j}\, du_j]$$

for an arbitrary function g. A simple calculation gives

$$J((1-u_i)^m) = \prod_{j=1}^{i} \left(\frac{R_j}{R_j + m} \right), \quad (m = 1, 2, \dots). \tag{4.11}$$

Letting

$$\hat{F}_P(w) = 1 - \prod_{j: W_j^o \leq w} \left(\frac{R_j}{R_j + 1} \right),$$

which is Prentice's moment estimator of $F(w)$, under $\beta = 0$, based on the data from both samples, we note that $\hat{F}_P(W_i^o) = J(u_i)$. Now return to the scores (4.10). A logistic score generating density gives $\phi(u) = 2u - 1$, $\Phi(u) = u$ so that from (4.11),

$$Q_i = -1 + 2\hat{F}_P(W_i^o), \quad Q_i^* = \hat{F}_P(W_i^o). \tag{4.12}$$

These are Prentice's scores (Prentice, 1978).

An extreme value density yields $\phi(u) = -\log(1-u) - 1$, $\Phi(u) = -\log(1-u)$. Direct integration gives

$$J\{\log(1-u)\} = -\sum_{j=1}^{i} R_j^{-1},$$

so that

$$Q_i = \left(\sum_{j=1}^{i} R_j^{-1} \right) - 1, \quad Q_i^* = \sum_{j=1}^{i} R_j^{-1} \tag{4.13}$$

which are log-rank scores.

Note that $-V_n$ corresponds to the score statistic arising from the model

$$W_i^* = h(T_i) = -\beta z_i + E_i$$

(utilizing the density function and d.f. corresponding to either T_i or W_i^*; $i = 1, \dots, n$), where h is a nonrandom, absolutely continuous, nondecreasing function from $[0, a)$ onto $(-\infty, +\infty)$, and where $a \in (0, +\infty)$ or $a = +\infty$. However, since V_n and $-V_n$ have identical efficiency properties for a given alternative, either of the two can be utilized for testing $\beta = 0$ in the above model. Now, let $h_0(t)$ be an unspecified, nonrandom, nonnegative function defined on $[0, +\infty)$ such that

$$\lim_{u \to a^-} H_0(u) = +\infty,$$

where

$$H_0(u) = \int_0^u h_0(t)\, dt,$$

and where $a \in (0, +\infty)$ or $a = +\infty$. Then, if $h(u) = \log(H_0(u))$ and if E_i has an extreme value d.f., the conditional hazard function of T_i given z_i $(i = 1, \ldots, n)$, with $h(T_i)$ defined as above, is described by model (3.4); hence, scores (4.13) (with or without "−" in front of each) should be used for testing $H_0 : \beta = 0$ in (3.4). As shall be demonstrated in subsection 4.3.3, $-V_n$ with scores (4.13) can be written as

$$\sum_{i=1}^{K} (z_{(i)} - R_{1i} R_i^{-1}), \tag{4.14}$$

where $R_{1i} = \sum_{j=i}^{K} (M_{1j} + z_{(j)})$. Recall that in subsection 3.2.2 we derived (4.14) as a score statistic within the context of model (3.4). The results, therefore, of the previous paragraph are consistent with those of subsection 3.2.2.

In subsection 4.3.3, we shall show that the scores (4.10) satisfy condition (2.2); consequently, (4.9) can be expressed as a WL statistic, and so

$$E_{H_0}(V_n) = 0.$$

Suppose, for the moment, that the censoring mechanism is identical for all n items. Then, to determine a variance estimator for V_n under H_0, we may employ a permutation model where the n scores are held fixed (each of which is treated as distinct), but where each of the $n!$ possible labellings of the n scores is equally likely. Such a model yields a variance of

$$V_p = (n-1)^{-1} \left[\sum_{i=1}^{K} (Q_i^2 + M_i Q_i^{*2}) \right] \left(\frac{n_1 n_2}{n} \right),$$

which is of course the permutation variance estimator. Note that V_p is the exact variance of V_n, under H_0, for simple and progressive type II censoring since the scores, the M_i's, and K are nonrandom in both of these situations.

In contrast to V_p, the Fisher information based on $p(\beta)$,

$$\mathcal{V}_0 = - \left. \frac{d^2 \log p(\beta)}{d\beta^2} \right|_{\beta=0}, \tag{4.15}$$

provides a variance estimator that is generally appropriate. We now demonstrate that, under H_0, $\mathcal{V}_0$ is unbiased for $\mathrm{Var}_{H_0}(V_n)$ both with and without censoring.

Consider first th' uncensored rank probability $\Pr(\mathbf{R} = \mathbf{r})$ of (4.3). For this case, we have

$$\begin{aligned} 0 &= E_{H_0} \left[\left. \left(\frac{d^2 \Pr(\mathbf{R} = \mathbf{r})}{d\beta^2} (\Pr(\mathbf{R} = \mathbf{r}))^{-1} \right) \right|_{\beta=0} \right] \\ &= E_{H_0} \left[n! \left. \frac{d^2 \Pr(\mathbf{R} = \mathbf{r})}{d\beta^2} \right|_{\beta=0} \right]. \end{aligned} \tag{4.16}$$

Moreover, with censoring absent, V_n reduces to V_n^* of (4.4), while $\mathcal{V}_0$ reduces to

$$V_n^{*2} - n! \left. \frac{d^2 \Pr(\mathbf{R} = \mathbf{r})}{d\beta^2} \right|_{\beta=0}. \tag{4.17}$$

Hence, from (4.16), (4.17), and from the fact that $E_{H_0}(V_n^*) = 0$, we have

$$E_{H_0}(\mathcal{V}_0) = E_{H_0}(V_n^{*2}) = \mathrm{Var}_{H_0}(V_n^*).$$

Now return to the censored data generalized rank probability $p(\beta)$ of (4.8). Expression (4.15) can be rewritten

$$\mathcal{V}_0 = V_n^2 - (R_1 R_2 \cdots R_K) \left. \frac{d^2 p(\beta)}{d\beta^2} \right|_{\beta=0}.$$

Since $E_{H_0}(V_n^2) = \mathrm{Var}_{H_0}(V_n)$, it is only necessary to show that the expectation of the second term, under H_0, is zero. For this, note that

$$\begin{aligned} U &= R_1 \cdots R_K \left. \frac{d^2 p(\beta)}{d\beta^2} \right|_{\beta=0} \\ &= R_1 \cdots R_K \sum_{\mathbf{r}_i \in S} \left. \frac{d^2 \Pr(\mathbf{R} = \mathbf{r}_i)}{d\beta^2} \right|_{\beta=0} \\ &= \frac{\sum_{\mathbf{r}_i \in S} n! \left. \frac{d^2 \Pr(\mathbf{R}=\mathbf{r}_i)}{d\beta^2} \right|_{\beta=0}}{(n!/R_1 \cdots R_K)}, \end{aligned}$$

where $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{n!}\}$ is the set of all possible underlying uncensored rank vectors, and where S is the set of underlying rank vectors consistent with the generalized rank vector. There are $n!/(R_1 \cdots R_K)$ vectors in S; hence, U is in fact the average of

$$\left. \frac{n! d^2 \Pr(\mathbf{R} = \mathbf{r})}{d\beta^2} \right|_{\beta=0}$$

over all of these vectors. Prentice (1978) uses this result and an inductive argument to demonstrate that, for simple or progressive type I censoring,

$$E_{H_0}(U | B_i) = 0$$

for $i = 1, 2$. The condition B_i means that the underlying uncensored rank vector is in D_i, where $D_1 = \{(d_1, \ldots, d_n) | \text{all } d_i > w^*\}$, $D_2 = D_1^c$, $w^* = \min(w_1^*, \ldots, w_n^*)$, and w_i^* is the realized potential censoring time for item i $(i = 1, \ldots, n)$. It follows that $E_{H_0}(U) = 0$ as desired.

Of course, further work must be done to demonstrate that $E_{H_0}(U) = 0$ for independent censoring mechanisms in general. Moreover, further research is required to establish, for independent censoring mechanisms, the consistency of $n^{-1}\mathcal{V}_0$ for $\text{Asvar}_{H_0}(n^{-1/2}V_n)$ when $\beta = 0$. Prentice (1978), though, declares this to be so for simple and progressive type II censoring.

Consider now the calculation of $\mathcal{V}_0$. After straightforward differentiation of $p(\beta)$, $\mathcal{V}_0$ can be written as

$$\sum_{i=1}^{K} \left[z_{(i)}^2 J\{\psi_1(u_i)\} + M_{1i} J\{\psi_2(u_i)\} \right] - \left\{ J(B^2) - V_n^2 \right\},$$

where

$$\psi_1(u) = \left[-d^2 \log f(\tau)/d\tau^2 \right]_{\tau = F^{-1}(u)},$$
$$\psi_2(u) = \left[-d^2 \log(1 - F(\tau))/d\tau^2 \right]_{\tau = F^{-1}(u)},$$

and

$$B = \sum_{i=1}^{K} \left\{ z_{(i)}\phi(u_i) + M_{1i}\Phi(u_i) \right\}.$$

$\mathcal{V}_0$ can be calculated explicitly in the aforementioned special cases. A logistic density f gives the variance estimator

$$\sum_{i=1}^{K} \left\{ A_i(1 - A_i^*)X_{(i)} - (A_i^* - A_i)X_{(i)} \left(A_i X_{(i)} + 2 \sum_{j=i+1}^{K} (A_j X_{(j)}) \right) \right\},$$

where

$$A_i = \prod_{j=1}^{i} \left(\frac{R_j}{R_j + 1} \right), \quad A_i^* = \prod_{j=1}^{i} \left(\frac{R_j + 1}{R_j + 2} \right), \quad X_{(i)} = 2z_{(i)} + M_{1i},$$

$i = 1, \ldots, K$. The extreme value distribution gives a variance estimator of

$$\sum_{i=1}^{K} \{ R_i^{-1} R_{1i} - R_i^{-2} R_{1i}^2 \} = \sum_{i=1}^{K} (R_i^{-2} R_{1i} R_{2i}) , \tag{4.18}$$

where $R_{2i} = R_i - R_{1i}$. Remark that the extreme value distribution, once again, yields (4.18) as a variance estimator if "$-\beta$" is replaced with "$+\beta$" in $p(\beta)$. The integral (4.8), though, with "$-\beta$ " replaced by "$+\beta$ " and with $f(x) = \exp(x - e^x)$, is tractable and leads to

$$\prod_{i=1}^{K} \left[e^{\beta z_{(i)}} \left(\sum_{l \in \mathcal{R}(W_i^o)} e^{\beta z_l} \right)^{-1} \right],$$

which is Cox's partial likelihood function. Here

$$\mathcal{R}(W_i^o) = \{ (j), (j,k) : \ j \geq i; k = 1, \ldots, M_j \} .$$

Recall that, in subsection 3.2.2, we generated (4.18) as a Fisher information based on Cox's partial likelihood.

Assuming that V_0/n is consistent for Asvar$(n^{-1/2} V_n)$ under H_0, the hypothesized asymptotic null normality of the proposed test statistic $V_n/\sqrt{V_0}$ cannot be proven with V_n in the given form. It can be shown, however, that V_n is expressible as a WL statistic. With V_n written as such, we can confirm its predicted asymptotic null distribution.

4.3.2 An Asymptotically Equivalent Test Statistic.

The formulae (4.10) are inconvenient for many score generating densities, f, so that approximate scores of the type (4.6) may be preferred. A first-order approximation to $\phi(u_i)$ about $U_{io} = \hat{F}_P(W_i^o)$ gives

$$J(\phi(u_i)) \simeq \phi(U_{io}) + J(u_i - U_{io})\phi'(U_{io}) = \phi(U_{io}).$$

A similar approximation to $\Phi(u_i)$ suggests the scoring scheme

$$Q_i = \phi(\hat{F}_P(W_i^o)), \quad Q_i^* = \Phi(\hat{F}_P(W_i^o)). \tag{4.19}$$

We now show, via the methodology of Cuzick (1985), that $V_n/\sqrt{V_0}$ and $S_n/\sqrt{V_0}$, where $S_n = \sum_{i=1}^{K}\left[z_{(i)}\phi(U_{io}) + M_{1i}\Phi(U_{io})\right]$, are asymptotically equivalent test statistics. We list, however, only highlights of the proof.

Firstly, assume that the score function ϕ is twice continuously differentiable on $(0,1)$ and that

$$|u^{-1}\phi'(u)| + |\phi''(u)| \le Lu^{-\alpha} \tag{4.20}$$

for some $0 \le \alpha < 5/2$ and $0 \le L < \infty$. Now, expand $\phi(u_i)$ around U_{io} to see that for any $i \le K$

$$\begin{aligned} |\phi(U_{io}) - J\{\phi(u_i)\}| &= \left| J\left\{\frac{(U_{io} - u_i)^2}{2}\phi''(\Delta u_i + (1-\Delta)U_{io})\right\}\right| \\ &\le J\left\{\frac{(U_{io} - u_i)^2}{2}\left|\phi''(\Delta u_i + (1-\Delta)U_{io})\right|\right\}, \end{aligned} \tag{4.21}$$

where $0 < \Delta \le 1$. It follows from (4.20) that

$$\begin{aligned} |\phi''(\Delta u_i + (1-\Delta)U_{io})| &\le L(\Delta u_i + (1-\Delta)U_{io})^{-\alpha} \\ &\le L(\Delta u_i)^{-\alpha} \\ &\le L\Delta^{-\alpha}(u_i^{-\alpha} + U_{io}^{-\alpha}) \end{aligned}$$

for some $0 \leq L < \infty$ and some $0 \leq \alpha < 5/2$. Thus, (4.21) is bounded by a constant times

$$U_{io}^{-\alpha} J\{(U_{io} - u_i)^2\} + J\{u_i^{-\alpha}(U_{io} - u_i)^2\}. \tag{4.22}$$

Cuzick then shows that both terms in (4.22) are less than a constant times $R_i^{-1}(1-U_{io})^{2-\alpha}$, with the first term being so for $R_i \geq 1$ and with the second one being so for $R_i \geq 6$. It is easily checked that Φ also satisfies (4.20), and so a similar argument can be used to bound the difference between $J(\Phi(u_i))$ and $\Phi(U_{io})$. Thus, letting

$$R(w) = \sum_{i=1}^{n} I(Y_i \geq w),$$

we have

$$\begin{aligned}
& n^{-1/2}\left| \sum_{i: R_i \geq 6} \left[z_{(i)}\phi(U_{io}) + M_{1i}\Phi(U_{io})\right] - \sum_{i: R_i \geq 6} \left[z_{(i)}J(\phi(u_i)) + M_{1i}J(\Phi(u_i))\right]\right| \\
&= n^{-1/2}\left| \sum_{i: R_i \geq 6} z_{(i)}\left[\phi(U_{io}) - J(\phi(u_i))\right] + \sum_{i: R_i \geq 6} M_{1i}\left[\Phi(U_{io}) - J(\Phi(u_i))\right]\right| \\
&\leq n^{-1/2}(\text{constant}) \sum_{i: R_i \geq 6} \left[R_i^{-1}(1 - U_{io})^{2-\alpha}(1 + M_i)\right] \\
&\leq -n^{-1/2}(\text{constant}) \int_{w: R(w) \geq 6} \frac{(1 - \hat{F}_P(w))^{2-\alpha}}{(R(w)/n)}\, d(R(w)/n)
\end{aligned} \tag{4.23}$$

$\xrightarrow{P} 0$ as $n \to \infty$ if $\alpha \leq 2$. If $\alpha > 2$, then, because $1 - \hat{F}_P(w) \geq R(w)/n$, (4.23) is bounded by a constant times

$$-n^{-1/2} \int_{w: R(w) \geq 6} \left(\frac{R(w)}{n}\right)^{1-\alpha} d(R(w)/n) \xrightarrow{P} 0$$

as $n \to \infty$. For both $\alpha \leq 2$ and $\alpha > 2$, the finite number of terms when $R(w) < 6$ are easily seen to be negligible. Assuming that $\mathcal{V}_0/n$ is bounded away from zero in probability as $n \to \infty$, it follows that

$$\left| n^{-1/2}(V_n - S_n)/\sqrt{n^{-1}\mathcal{V}_0} \right| \xrightarrow{P} 0,$$

as required.

In that same paper, Cuzick (1985) shows that asymptotic equivalence of the two tests remains so if either the Kaplan-Meier estimator, or Altshuler's estimator, or the Peto-Peto estimator is used in place of U_{io}.

Consider now the following examples of the scoring system (4.19). An extreme value density for f yields scores

$$Q_i = -\log(1 - \hat{F}_P(W_i^o)) - 1, \quad Q_i^* = -\log(1 - \hat{F}_P(W_i^o)).$$

A logistic density gives scores

$$Q_i = -1 + 2\hat{F}_P(W_i^o), \quad Q_i^* = \hat{F}_P(W_i^o),$$

which are identical to the exact scores (4.12). A standard normal density gives

$$Q_i = G^{-1}(\hat{F}_P(W_i^o)), \quad Q_i^* = \frac{g(G^{-1}(\hat{F}_P(W_i^o)))}{1 - \hat{F}_P(W_i^o)},$$

where $g(t)$ is the standard normal density and $G(t)$ is the corresponding d.f.. A double exponential density yields

$$Q_i = \begin{cases} -1 & \text{if } \hat{F}_P(W_i^o) \leq 1/2, \\ 1 & \text{if } \hat{F}_P(W_i^o) > 1/2, \end{cases} \quad Q_i^* = \begin{cases} \frac{\hat{F}_P(W_i^o)}{1-\hat{F}_P(W_i^o)} & \text{if } \hat{F}_P(W_i^o) \leq 1/2, \\ 1 & \text{if } \hat{F}_P(W_i^o) > 1/2. \end{cases}$$

Here we have defined $f'(0) = 1/2$, for $f'(0)$, in actual fact, does not exist. Finally, the family of densities

$$f_\rho(x) = \begin{cases} e^x \exp(-e^x) & \text{if } \rho = 0, \\ (1 + \rho e^x)^{-(1+\rho)/\rho} e^x & \text{if } \rho > 0, \end{cases}$$

gives

$$Q_i = \begin{cases} -\log(1-\hat{F}_P(W_i^o)) - 1 & \text{if } \rho = 0, \\ (1/\rho) - \left(\frac{\rho+1}{\rho}\right)\left(1-\hat{F}_P(W_i^o)\right)^\rho & \text{if } \rho > 0; \end{cases}$$

$$Q_i^* = \begin{cases} -\log(1-\hat{F}_P(W_i^o)) & \text{if } \rho = 0, \\ (1/\rho)\left[1-(1-\hat{F}_P(W_i^o))^\rho\right] & \text{if } \rho > 0, \end{cases}$$

which are scores corresponding to the class of tests proposed by Harrington and Fleming (1982) (note that $\lim_{\rho\to 0^+} f_\rho(x) = f_0(x)$).

4.3.3 Representation as an SWL Statistic.

From the results of subsection 2.3.1, the score statistic (4.9) can be written as

$$\sum_{i=1}^{K}[(Q_i - Q_i^*)(z_{(i)} - R_{1i}R_i^{-1})], \tag{4.24}$$

if the scores satisfy

$$Q_i - Q_i^* = R_i(Q_{i-1}^* - Q_i^*), \quad i = 1, \ldots, K. \tag{4.25}$$

These equations indeed hold as Mehrotra, Michalek, and Mihalko (1982) demonstrated.

First, note that upon direct integration on $u_{i+1}, u_{i+2}, \ldots, u_K$, Q_i and Q_i^* may be written

$$Q_i = \int_{u_1<} \cdots \int_{<u_i} \phi(u_i)R_i(1-u_i)^{R_i-1} \prod_{j=1}^{i-1}\left\{R_j(1-u_j)^{M_j}\,du_j\right\}\,du_i, \tag{4.26}$$

$$Q_i^* = \int_{u_1<\cdots} \cdots \int_{<u_i} \Phi(u_i)R_i(1-u_i)^{R_i-1} \prod_{j=1}^{i-1}\left\{R_j(1-u_j)^{M_j}\,du_j\right\}\,du_i, \tag{4.27}$$

where the product term in both of these equations is unity if $i = 1$. If $R_i > 1$, then the integral on u_i in (4.27),

$$\int_{u_{i-1}}^{1} \Phi(u_i)R_i(1-u_i)^{R_i-1}\,du_i,$$

where $u_0 = 0$, can be integrated by parts using $U = f(F^{-1}(u_i))$ and $dV = R_i(1-u_i)^{R_i-2}\,du_i$. This integration yields

$$H(u_{i-1}) = \left(\frac{R_i}{R_i-1}\right) f(F^{-1}(u_{i-1}))(1-u_{i-1})^{R_i-1} - \left(\frac{1}{R_i-1}\right)\int_{u_{i-1}}^{1} R_i(1-u_i)^{R_i-1}\phi(u_i)\,du_i;$$

hence,

$$Q_i^* = \int\cdots\int_{u_1<\cdots<u_{i-1}} H(u_{i-1})R_{i-1}(1-u_{i-1})^{M_{i-1}} \times \prod_{j=1}^{i-2}\left\{R_j(1-u_j)^{M_j}\,du_j\right\}\,du_{i-1} = \frac{R_i}{R_i-1}Q_{i-1}^* - \frac{Q_i}{R_i-1},$$

as required. If $R_i = 1$, then $i = K$ and $M_K = 0$. Integration on u_K in (4.26) with these values shows that

$$\int_{u_{K-1}}^{1}\phi(u_K)\,du_K = \Phi(u_{K-1})(1-u_{K-1}) = \int_{u_{K-1}}^{1}\Phi(u_{K-1})\,du_K;$$

hence, $Q_K = Q_{K-1}^*$, which is the special case of (4.25) for $R_i = 1$. Therefore, since Q_i and Q_i^* are solely dependent on $R_1, R_2, \ldots, R_i$, we conclude that $V_n/\sqrt{V_0}$ is expressible as an SWL statistic.

Note from the above results that the rank statistic for uncensored data,

$$V_n^* = \sum_{i=1}^{K} z_{(i)}E(\phi(U_{(i)})) = \sum_{i=1}^{K}\left[z_{(i)}E(\phi(U_{(i)})) + M_{1i}E(\Phi(U_{(i)}))\right],$$

can thus be written as

$$\sum_{i=1}^{K}\left[E(\phi(U_{(i)}) - \Phi(U_{(i)}))\left(z_{(i)} - R_{1i}/R_i\right)\right],$$

where $M_{1i} = 0$, and where $U_{(i)}$ is the i^{th} order statistic in a uniform $(0,1)$ sample of size n.

Under the conditions of (4.20), the statistic (4.24), via a methodology similar to that of subsection 4.3.2, can be shown to be asymptotically equivalent to

$$S_n^* = \sum_{i=1}^{K} [\phi(U_{io}) - \Phi(U_{io})] (z_{(i)} - R_{1i} R_i^{-1})$$

(Cuzick, 1985); consequently, under these conditions and under the assumption that V_0/n is bounded away from zero in probability as $n \to \infty$, $S_n^*/\sqrt{V_0}$ is asymptotically equivalent to

$$\frac{\sum_{i=1}^{K} [z_{(i)}\phi(U_{io}) + M_{1i}\Phi(U_{io})]}{\sqrt{V_0}}.$$

Moreover, asymptotic equivalence between these two tests holds for any pair of score functions $\phi(u)$, $\Phi(u)$ such that ϕ satisfies both (4.20) and the condition

$$\int_0^1 \phi(u)\, du = 0,$$

and such that $\Phi(u)$ is defined as

$$\Phi(u) = \int_u^1 \phi(s)\, ds/(1-u).$$

CHAPTER 5
LARGE AND SMALL-SAMPLE BEHAVIOR OF SWL STATISTICS

5.1 Introduction.

In this chapter, we shall be concerned with asymptotic and small-sample properties of SWL statistics. Our examination of asymptotic characteristics (section 5.2) assumes that a random censorship model is in effect with the potential censoring times of each sample identically distributed. For a discussion of large-sample behavior as pertaining to the whole class of independent censoring mechanisms, see Gill (1980) and Anderson et al. (1982). In subsections 5.2.1 and 5.2.2, we show that the limiting distribution of a WL statistic, under respectively the null hypothesis and a sequence of contiguous alternatives, is a normal distribution. The variances in these two cases are identical, while the means are different. We, as well, shall derive consistent estimators of the variance.

In subsection 5.2.3, we first generate an optimal limiting weight function for a special class of contiguous sequences of alternatives. Subsequently, for a particular member of this class, we suggest that an SWL statistic can be asymptotically fully efficient if and only if the censoring distribution of sample 1 is identical to that of sample 2. We then conclude 5.2.3 by constructing a test which should be especially powerful against a parametric alternative that can be reduced to a location family after a suitable transformation.

In subsection 5.2.4, we present WL statistics whose asymptotic null-hypothesis variance is free of both the failure time and censoring distributions. Subsection 5.2.5 is concerned with aspects related to consistency of the tests. We conclude our examination of asymptotic properties in subsection 5.2.6 with a comparison of the efficacy of censored data extensions of the Wilcoxon and Savage statistics against specified alternatives.

In section 5.3, we are concerned with small-sample behavior of SWL statistics. Subsection 5.3.1 deals with the estimation of null distribution of specific members of this class of procedures, while subsection 5.3.2 examines power properties of selected tests.

5.2 Asymptotic Properties.

5.2.1 Limiting Null Distribution of a WL Statistic.

The following derivation is based on the methodology of Tsiatis (1982), who considers the asymptotic joint distribution of sequentially computed WL statistics.

Suppose n items are put on test in a two-sample scenario, with n_i ($i = 1, 2$) being the number of items comprising sample i. Without loss of generality, assume sample 1 consists of item labels $1, \ldots, n_1$, and that sample 2 consists of the remaining labels.

The censoring mechanism to be implemented here and throughout section 5.2 is the random censorship model with the potential censoring times of each sample identically distributed. Thus, let the nonnegative random variables T_i, C_i denote respectively the failure time and potential censoring time corresponding to item i, and let $F_j(t)$, $G_j(t)$ be respectively the failure time and censoring d.f. associated with sample j. The null hypothesis of interest is $H_0: F_1(t) = F_2(t) = F(t)$. It is assumed now and throughout 5.2 that

(1) if $G_i(t)$ $(i = 1, 2)$ is *not* absolutely continuous, it is left continuous, and

(2) $F_i(t) < 1$ for every $t \in [0, \infty)$ $(i = 1, 2)$.

Let Y_i and Δ_i be, respectively, the survival time variable and censoring indicator variable corresponding to item i ($\Delta_i = 1$ if $T_i < C_i$, $\Delta_i = 0$ if otherwise). Then the complete data for the study is represented by the n independently distributed random vectors (Y_i, Δ_i), $i = 1, \ldots, n$ Note, however, that the sets

$$\{(Y_1, \Delta_1), \ldots, (Y_{n_1}, \Delta_{n_1})\}, \quad \{(Y_{n_1+1}, \Delta_{n_1+1}), \ldots, (Y_n, \Delta_n)\}$$

each consists of identically distributed random vectors.

For testing H_0, we use the WL statistic

$$S_n = \sum_{i=1}^{n_1} \hat{Q}(Y_i)\Delta_i \left\{1 - \frac{R_1(Y_i)}{R(Y_i)}\right\} + \sum_{i=n_1+1}^{n} \hat{Q}(Y_i)\Delta_i \left\{-\frac{R_1(Y_i)}{R(Y_i)}\right\},$$

where

$$R(t) = \sum_{j=1}^{n} I(Y_j \geq t)$$

and

$$R_1(t) = \sum_{j=1}^{n_1} I(Y_j \geq t).$$

The weight function $\hat{Q}(t)$ is assumed under H_0 to converge in probability to a function $Q(t)$ uniformly on $[0, \infty)$. $Q(t)$ is dependent on t through $\bar{F}(t)$, $\bar{G}_1(t)$, and $\bar{G}_2(t)$, where $\bar{F}(t) = 1 - F(t)$, $\bar{G}_i(t) = 1 - G_i(t)$ $(i = 1, 2)$.

The key to deriving the asymptotic null distribution of S_n is to approximate it by a sum of independently distributed random variables. A routine application of the Lindberg-Feller version of the central limit theorem will then be used to obtain the asymptotic results.

We first note that S_n can be written as

$$\sum_{i=1}^{n_1} \int_0^\infty \hat{Q}(t) \left\{1 - \frac{R_1(t)}{R(t)}\right\} dN_i(t) + \sum_{i=n_1+1}^{n} \int_0^\infty \hat{Q}(t) \left\{-\frac{R_1(t)}{R(t)}\right\} dN_i(t), \tag{5.1}$$

where $N_i(t) = I(Y_i \leq t, \Delta_i = 1)$. By adding and subtracting terms, we can rewrite (5.1) as

$$S_n = \sum_{i=1}^{n_1} \int_0^\infty \hat{Q}(t) \left\{1 - \frac{R_1(t)}{R(t)}\right\} dJ_i(t) + \sum_{i=n_1+1}^{n} \int_0^\infty \hat{Q}(t) \left\{-\frac{R_1(t)}{R(t)}\right\} dJ_i(t), \tag{5.2}$$

where $J_i(t) = N_i(t) - \int_0^t \lambda(s) I(Y_i \geq s)\, ds$, and where $\lambda(t)$ is the hazard function associated with $F(t)$.

All subsequent results are now under the null hypothesis. By using the law of large numbers, we can establish that as $n \to \infty$,

$$R_1(t)/n_1 \xrightarrow{P} \pi_1(t)$$

and

$$R(t)/n \xrightarrow{P} p_1\pi_1(t) + p_2\pi_2(t),$$

where

$$\pi_i(t) = \bar{F}(t)\bar{G}_i(t) = \Pr(Y_j \geq t),$$

for $j \in$ sample i $(i = 1, 2)$. Here, we have assumed that

$$0 < \lim_{n\to\infty} (n_i/n) = p_i < 1,$$

a condition which shall be maintained throughout section 5.2. Hence,

$$\frac{R_1(t)}{R(t)} \xrightarrow{P} \mu(t) = \frac{p_1\pi_1(t)}{p_1\pi_1(t) + p_2\pi_2(t)}. \tag{5.3}$$

Now, the statistic S_n given in (5.2) can be written as $\tilde{S}_n + E_n$, where

$$\tilde{S}_n = \sum_{i=1}^{n_1} \int_0^\infty Q(t)\{1 - \mu(t)\}\, dJ_i(t) + \sum_{i=n_1+1}^{n} \int_0^\infty Q(t)\{-\mu(t)\}\, dJ_i(t)$$

and

$$\begin{aligned} E_n = & -\sum_{i=1}^{n} \int_0^\infty Q(t)\{(R_1(t)/R(t)) - \mu(t)\}\, dJ_i(t) \\ & + \sum_{i=1}^{n_1} \int_0^\infty \{\hat{Q}(t) - Q(t)\}\{1 - \mu(t)\}\, dJ_i(t) \\ & + \sum_{i=n_1+1}^{n} \int_0^\infty \{\hat{Q}(t) - Q(t)\}\{-\mu(t)\}\, dJ_i(t) \\ & - \sum_{i=1}^{n} \int_0^\infty \{\hat{Q}(t) - Q(t)\}\{(R_1(t)/R(t)) - \mu(t)\}\, dJ_i(t). \end{aligned}$$

Since $R_1(t)/R(t) \xrightarrow{P} \mu(t)$ and $\hat{Q}(t) \xrightarrow{P} Q(t)$, we can show, via the results of Tsiatis (1981b, Lemma 3.1) and Breslow and Crowley (1974, Theorem 4), that $n^{-1/2}E_n$ is a second-order term that is asymptotically negligible. Hence, the asymptotic distribution of the statistic $n^{-1/2}S_n$ is the same as that of $n^{-1/2}\tilde{S}_n$.

The approximate statistic $\tilde{S}_n$ can be written as

$$\begin{aligned} \tilde{S}_n = & \sum_{i=1}^{n_1} \left[\Delta_i Q(Y_i)\{1 - \mu(Y_i)\} - \int_0^{Y_i} Q(t)\{1 - \mu(t)\}\lambda(t)\, dt\right] \\ & + \sum_{i=n_1+1}^{n} \left[\Delta_i Q(Y_i)\{-\mu(Y_i)\} - \int_0^{Y_i} Q(t)\{-\mu(t)\}\lambda(t)\, dt\right]. \end{aligned} \tag{5.4}$$

Although (5.4) is complex, it is nonetheless a sum of independently distributed random variables, and so its asymptotic distribution can be obtained by application of the Lindberg-Feller form of the central limit theorem. As will be demonstrated below, $E(n^{-1/2}\tilde{S}_n) = 0$ and

$$\lim_{n\to\infty} \text{Var}(n^{-1/2}\tilde{S}_n) = \sigma^2 = \int_0^\infty Q^2(t)\Phi(t)\lambda(t)\,dt,$$

where

$$\Phi(t) = \frac{p_1 p_2 \pi_1(t)\pi_2(t)}{p_1\pi_1(t) + p_2\pi_2(t)}.$$

Thus, $n^{-1/2}\tilde{S}_n \xrightarrow{D} N(0,\sigma^2)$ as $n \to \infty$.

We now confirm the above-mentioned expectation and asymptotic variance of $n^{-1/2}\tilde{S}_n$. To do this, we nust first evaluate the first and second moments of

$$\left[\Delta_i Q(Y_i)\{1-\mu(Y_i)\} - \int_0^{Y_i} Q(t)\{1-\mu(t)\}\lambda(t)\,dt\right] \tag{5.5}$$

and

$$\left[\Delta_i Q(Y_i)\{-\mu(Y_i)\} - \int_0^{Y_i} Q(t)\{-\mu(t)\}\lambda(t)\,dt\right]. \tag{5.6}$$

Let $\tilde{Q}(t) = Q(t)\{1-\mu(t)\}$. Then the expectation of (5.5) is

$$\int_0^\infty \tilde{Q}(t)\lambda(t)\pi_1(t)\,dt \tag{5.7A}$$

$$+\int_0^\infty \left\{\int_0^t \tilde{Q}(u)\lambda(u)\,du\right\} d\pi_1(t). \tag{5.7B}$$

Integrating (5.7B) by parts, we note that this equals the negative value of (5.7A). Hence, the expectation of (5.5) equals zero. In a similar manner, we can show that the expectation of (5.6) equals zero. The first moment of $n^{-1/2}\tilde{S}_n$, therefore, is zero.

The second moment of (5.5) can be expressed as

$$E\left\{\tilde{Q}^2(Y_i)\Delta_i\right\} \tag{5.8A}$$

$$-2E\left\{\tilde{Q}(Y_i)\Delta_i\int_0^{Y_i}\tilde{Q}(u)\lambda(u)\,du\right\} \tag{5.8B}$$

$$+E\left[\left\{\int_0^{Y_i}\tilde{Q}(u)\lambda(u)\,du\right\}^2\right]. \tag{5.8C}$$

For simplicity of notation, we shall denote

$$\int_0^t \tilde{Q}(u)\lambda(u)\,du$$

by $\phi(t)$. Expression (5.8A) is equal to

$$A=\int_0^\infty \tilde{Q}^2(u)\lambda(u)\pi_1(u)\,du.$$

Expression (5.8B) is equal to

$$B=-2\int_0^\infty \tilde{Q}(u)\phi(u)\lambda(u)\pi_1(u)\,du.$$

Expression (5.8C) is calculated in two regions, namely when $\{\Delta_i=1\}$ and $\{\Delta_i=0\}$. In the region where $\{\Delta_i=1\}$, (5.8C) is equal to

$$C=\int_0^\infty \phi^2(u)\lambda(u)\pi_1(u)\,du.$$

In the region where $\{\Delta_i=0\}$, (5.8C) is equal to

$$D=-\int_0^\infty \phi^2(u)\bar{F}(u)\,d\bar{G}_1(u).$$

After integrating D by parts, we note that $B+C+D=0$. Hence, the second moment of (5.5) is equal to

$$A=\sigma_1^2=\int_0^\infty Q^2(t)\{1-\mu(t)\}^2\lambda(t)\pi_1(t)\,dt.$$

Similarly, the second moment of (5.6) is

$$\sigma_2^2 = \int_0^\infty Q^2(t)\mu^2(t)\lambda(t)\pi_2(t)\,dt.$$

Therefore,

$$\begin{aligned}
\sigma^2 &= \text{Asvar}(n^{-1/2}\tilde{S}_n) \\
&= \lim_{n\to\infty} \text{Var}(n^{-1/2}\tilde{S}_n) \\
&= \lim_{n\to\infty} \hat{p}_1\sigma_1^2 + \hat{p}_2\sigma_2^2 \\
&= p_1\sigma_1^2 + p_2\sigma_2^2 \\
&= \int_0^\infty Q^2(t)\lambda(t)\Phi(t)\,dt \qquad (5.9) \\
&= \int_0^\infty J^2(\bar{F}(t),\bar{G}_1(t),\bar{G}_2(t))p_1p_2\frac{\bar{G}_1(t)\bar{G}_2(t)}{p_1\bar{G}_1(t)+p_2\bar{G}_2(t)}f(t)\,dt,
\end{aligned}$$

where $\hat{p}_i = n_i/n$ $(i = 1,2)$, where $f(t) = \bar{F}(t)\lambda(t)$, and where $J(u_1,u_2,u_3)$ is a nonrandom function such that

$$J:\ [0,1]\times[0,1]\times[0,1]\to(-\infty,+\infty)$$

and such that $Q(t) = J(\bar{F}(t),\bar{G}_1(t),\bar{G}_2(t))$.

If $\bar{F}(t)$, $\bar{G}_1(t)$, and $\bar{G}_2(t)$ are unspecified, then σ^2 is nonevaluable unless its integrand is free of $\bar{G}_1(t),\bar{G}_2(t)$, and unless $\sup\{t:\ \bar{G}_1(t)\bar{G}_2(t)>0\}$ is known (see subsection 5.2.4). On the other hand, if $\bar{F}(t)$ is unknown but $\bar{G}_1(t)$, $\bar{G}_2(t)$ are discrete and specified (for example, simple type I censorship), then

$$\begin{aligned}
\sigma^2 &= \sum_{i=1}^m \int_{c_{i-1}^*}^{c_i^*} J^2(\bar{F}(t),p_{1i}^*,p_{2i}^*)p_1p_2\frac{p_{1i}^*p_{2i}^*}{p_1p_{1i}^*+p_2p_{2i}^*}f(t)\,dt \\
&= -\sum_{i=1}^m \int_{u_{i-1}^*}^{u_i^*} J^2(u,p_{1i}^*,p_{2i}^*)p_1p_2\frac{p_{1i}^*p_{2i}^*}{p_1p_{1i}^*+p_2p_{2i}^*}\,du,
\end{aligned}$$

where $0 = c_0^* < c_1^* < \cdots < c_m^*$ are the ordered distinct values of the set of points of discontinuity arising from either $\bar{G}_1(t)$ or $\bar{G}_2(t)$ or both, and where $p_{ji}^* = \bar{G}_j(c_i^*)$ $(j = 1, 2)$ and $u_i^* = \bar{F}(c_i^*)$. Under the given censoring conditions, therefore, σ^2 is calculable.

Using methods similar to Tsiatis (1981A), a consistent estimator of σ^2, denoted by $\hat{\sigma}_{cp}^2$, can be obtained by replacing the quantities in (5.9) by their appropriate estimators. Therefore,

$$\begin{aligned}\hat{\sigma}_{cp}^2 &= \int_0^\infty \frac{\hat{Q}^2(t)\hat{p}_1\hat{p}_2\hat{\pi}_1(t)\hat{\pi}_2(t)}{\hat{p}_1\hat{\pi}_1(t) + \hat{p}_2\hat{\pi}_2(t)}\, d\hat{\Lambda}(t) \\ &= \sum_{i=1}^n \frac{\hat{Q}^2(Y_i)\Delta_i R_1(Y_i)R_2(Y_i)}{nR^2(Y_i)},\end{aligned} \tag{5.10}$$

where $R_2(t) = R(t) - R_1(t)$ and where

$$\hat{\Lambda}(t) = \sum_{i=1}^n \int_0^t \frac{dN_i(u)}{R(u)} \xrightarrow{P} \int_0^t \lambda(u)\, du$$

as $n \to \infty$. $\hat{\sigma}_{cp}^2$ is, of course, the conditional permutation variance estimator for $n^{-1/2}S_n$. We thus have that

$$n^{-1/2}S_n / \sqrt{\hat{\sigma}_{cp}^2} \xrightarrow{D} N(0, 1) \quad \text{as } n \to \infty.$$

The permutation variance estimator for $n^{-1/2}S_n$ is

$$\begin{aligned}\hat{\sigma}_p^2 &= \frac{\hat{p}_1\hat{p}_2}{(n-1)} \sum_{i=1}^n \Delta_i \hat{Q}^2(Y_i) \frac{(R(Y_i) - 1)}{R(Y_i)} \\ &= \frac{\hat{p}_1\hat{p}_2}{(n-1)} \int_0^\infty \hat{Q}^2(t)(R(t) - 1)\, d\hat{\Lambda}(t) \\ &\xrightarrow{P} \sigma_p^2 = p_1 p_2 \int_0^\infty Q^2(t)\left(p_1\pi_1(t) + p_2\pi_2(t)\right)\lambda(t)\, dt.\end{aligned}$$

If, in fact, $\bar{G}_1(t) = \bar{G}_2(t)$ for all t, then

$$\pi_1(t) = \pi_2(t) = \pi(t) \quad \forall t \in [0, \infty),$$

and so

$$\sigma_p^2 = \sigma^2 = \int_0^\infty Q^2(t) p_1 p_2 \pi(t) \lambda(t)\, dt.$$

Thus, when $\bar{G}_1(t) = \bar{G}_2(t)$ for all t, $\hat{\sigma}_p^2$ consistently estimates σ^2 and

$$n^{-1/2} S_n / \sqrt{\hat{\sigma}_p^2} \xrightarrow{D} N(0,1).$$

Jennrich (1983) shows that if

(i) $\pi_1(t) \leq \pi_2(t)$, for all t, and $p_1 \leq p_2$, or

(ii) $\pi_1(t) \geq \pi_2(t)$, for all t, and $p_1 \geq p_2$,

then $\sigma^2 \leq \sigma_p^2$, with equality holding only when $\pi_1(t) = \pi_2(t)$ for all t. Therefore, under condition (i) or (ii) and under H_0,

$$n^{-1/2} S_n / \sqrt{\hat{\sigma}_p^2}$$

is asymptotically conservative compared to $n^{-1/2} S_n / \sqrt{\hat{\sigma}_{cp}^2}$.

5.2.2 Limiting Distribution of a WL Statistic Under a Sequence of Contiguous Alternatives.

Suppose we are given a sequence (for $n = 1, 2, \ldots$) of two-sample set-ups, the n^{th} one having the form described in subsection 5.2.1 with a total of $n = n_1 + n_2$ observations:

$$(Y_1^n, \Delta_1^n), \ldots, (Y_{n_1}^n, \Delta_{n_1}^n), (Y_{n_1+1}^n, \Delta_{n_1+1}^n), \ldots, (Y_n^n, \Delta_n^n).$$

Here, $n_i/n \to p_i$ as $n \to \infty$, where $0 < p_i < 1$ $(i = 1, 2)$. Letting $G_i^n(t)$ denote the censoring d.f. for sample i of the n^{th} set-up, we assume that $G_i^n(t) = G_i(t)$ for every n and t. Let F_i^n be the failure time d.f. corresponding to sample i of the n^{th} set-up, and suppose for every n and t that

$$F_1^n(t) \geq F_2^n(t),$$

with, for each n, strict inequality at some $t \in \{t : \bar{G}_1(t)\bar{G}_2(t) > 0\}$. Suppose further that for each i

$$F_i^n(t) \to F(t) \quad \text{uniformly in } t \in [0, \infty) \tag{5.11}$$

as $n \to \infty$ for some d.f. $F(t)$. We refer to the sequence $\{F_1^n, F_2^n\}$ as a *sequence of contiguous alternatives*. The null hypothesis sequence is of the form $\{F_1^n, F_2^n\}$, with $F_1^n = F_2^n = F$ for every n and t, and we assume that large positive values of the SWL statistic in question lead to the rejection of H_0. Finally, we assume that the convergence (5.11) is such that for some real-valued functions γ_i $(i = 1, 2)$,

$$\sqrt{n\hat{p}_1\hat{p}_2}\left(\frac{\lambda_i^n(t)}{\lambda(t)} - 1\right) \to \gamma_i(t) \quad \text{as } n \to \infty \tag{5.12}$$

uniformly on each closed subinterval of $[0, +\infty)$, where $\lambda_i^n(t), \lambda(t)$ are the hazard functions associated with $F_i^n(t)$ and $F(t)$ respectively, and we define $\gamma = \gamma_1 - \gamma_2$ (Gill, 1980).

Now, consider the WL statistic S_n of subsection 5.2.1 with representation (5.1). Letting

$$K(t) = \hat{Q}(t)R_1(t)R_2(t)(R(t))^{-1},$$

$$M_1(t) = \sum_{i=1}^{n_1}\left[N_i(t) - \int_0^t \lambda_1^n(u)I(Y_i \geq u)\,du\right],$$

and

$$M_2(t) = \sum_{i=n_1+1}^{n}\left[N_i(t) - \int_0^t \lambda_2^n(u)I(Y_i \geq u)\,du\right], \tag{5.13}$$

(5.1) can be rewritten as

$$\int_0^\infty \frac{K(t)}{R_1(t)}\,dM_1(t) - \int_0^\infty \frac{K(t)}{R_2(t)}\,dM_2(t) + \int_0^\infty K(t)\left(\frac{\lambda_1^n(t)}{\lambda(t)} - 1\right)\lambda(t)\,dt$$
$$- \int_0^\infty K(t)\left(\frac{\lambda_2^n(t)}{\lambda(t)} - 1\right)\lambda(t)\,dt.$$

Gill (1980) shows by martingale methods that, under the given sequence of alternatives,

$$n^{-1/2}\int_0^\infty K(t)\left[\frac{dM_1(t)}{R_1(t)}-\frac{dM_2(t)}{R_2(t)}\right]\xrightarrow{D} N(0,\sigma^2)\quad \text{as } n\to\infty,$$

where σ^2 is defined as in (5.9). Thus, from the latter result, (5.12), and the fact that, under $\{F_1^n, F_2^n\}$, $R_i(t)/n_i$ and $\hat{Q}(t)$ converge uniformly on $[0,\infty)$ in probability to $\pi_i(t)$ and $Q(t)$ respectively, we have under $\{F_1^n, F_2^n\}$ that

$$n^{-1/2}S_n \xrightarrow{D} N(\mu,\sigma^2)$$

where

$$\mu=\int_0^\infty p_1^{-1/2}p_2^{-1/2}Q(t)\Phi(t)\gamma(t)\lambda(t)\,dt. \tag{5.14}$$

Under the alternative hypothesis, $\hat{\sigma}_p^2$ (when $\pi_1(t)=\pi_2(t)$) and $\hat{\sigma}_{cp}^2$ are consistent for σ^2.

5.2.3 Asymptotic Relative Efficiencies.

Consider the sequence of alternatives dealt with in subsection 5.2.2, and let V be σ^2 or a consistent estimator thereof under H_0. Then the Pitman efficacy of $n^{-1/2}S_n/\sqrt{V}$ (or of $n^{-1/2}S_n$) for this sequence of alternatives, assuming (5.12) holds, is given by

$$e=\frac{\mu^2}{\sigma^2}=\frac{\left\{\int_0^\infty Q(t)\gamma(t)\Phi(t)\lambda(t)\,dt\right\}^2}{p_1p_2\int_0^\infty Q^2(t)\Phi(t)\lambda(t)\,dt} \tag{5.15}$$

which, in view of Schwartz's inequality, is maximized by letting $Q(t)=\gamma(t)$. With such a choice of $Q(t)$,

$$e=e_{\max}=\int_0^\infty p_1^{-1}p_2^{-1}\gamma^2(t)\Phi(t)\lambda(t)\,dt.$$

As an application of the above results, suppose the weight function $\hat{Q}(t)$ is defined as

$$\hat{Q}(t) = \hat{Q}_{TW}(t) = g_0(R(t)/n),$$

where g_0 is some nonrandom function with domain $[0, 1]$. Recall that $\hat{Q}_{TW}(t)$ belongs to the Tarone-Ware class of weight functions. Then the efficacy of $n^{-1/2}S_n$, with $\hat{Q}(t) = \hat{Q}_{TW}(t)$, (denoted by e_{TW}) is given by (5.15) with

$$Q(t) = Q_{TW}(t) = g_0[p_1\pi_1(t) + p_2\pi_2(t)].$$

If, on the other hand,

$$\hat{Q}(t) = \hat{Q}_{KM}(t) = g_0(\hat{\bar{F}}(t)),$$

where $\hat{\bar{F}}(t)$ is the Kaplan-Meier estimator of $\bar{F}(t)$ based on the pooled sample of size n, then the efficacy of $n^{-1/2}S_n$ (denoted by e_{KM}) is given by (5.15) with $Q(t) = g_0(\bar{F}(t))$. Now, suppose specifically that g_0 is defined as

$$g_0(u) = \gamma(\bar{F}^{-1}(u)),$$

where we have assumed that $\bar{F}(t)$ is strictly decreasing on $[0, \infty)$. (Note here that $g_0(u)$ does not depend on any function related to the censoring mechanism.) Then the optimal weight function for S_n is $\hat{Q}_{KM}(t)$, and, by Schwartz's inequality, $e_{TW} < e_{KM} = e_{\max}$ unless $Q_{TW}(t)/\gamma(t)$ is constant, which is the case when there is no censoring. If, however, g_0 is given by

$$g_0(u) = \gamma(S^{-1}(u)),$$

where $S(t) = p_1\pi_1(t) + p_2\pi_2(t)$, and where $\pi_1(t)$ and/or $\pi_2(t)$ are strictly decreasing, then $e_{TW} = e_{\max}$. Remark though that, in this case, $g_0(u)$ depends on the censoring d.f.'s (Gu, Lai, and Lan 1991).

Suppose now that $\{F_\theta : \theta \in \Theta\}$ is some family of distribution functions on $[0, \infty)$ indexed by a parameter θ taking values in a real interval Θ. Denote by, $\lambda_\theta(t)$, the hazard function associated with $F_\theta(t)$. Suppose further that $F_i^n(t)$ and $F(t)$ of subsection 5.2.2 are such that

$$\begin{aligned} F_i^n(t) &= F_{\theta_i^n}(t), \quad i = 1, 2; n = 1, 2, \ldots \\ F(t) &= F_{\theta_0}(t) \end{aligned} \tag{5.16}$$

for some θ_0 and $\theta_i^n \in \Theta$. Therefore, defining θ_i^n, for some constant $c \neq 0$, by

$$\theta_i^n = \theta_0 + (-1)^{i+1} c \sqrt{\frac{\hat{p}_{i'}}{n\hat{p}_i}}, \quad i \neq i', \tag{5.17}$$

and assuming $\lambda_\theta(t)$ is differentiable with respect to θ at $\theta = \theta_0$, we have

$$\begin{aligned} \gamma_i(t) &= \lim_{n\to\infty} \sqrt{n\hat{p}_1\hat{p}_2} \left(\frac{\lambda_i^n(t)}{\lambda(t)} - 1 \right) \\ &= \lim_{n\to\infty} \left\{ \sqrt{n\hat{p}_1\hat{p}_2} \left(\frac{\lambda_{\theta_i^n}(t) - \lambda_{\theta_0}(t)}{\lambda_{\theta_0}(t)} \right) \frac{(\theta_i^n - \theta_0)}{(\theta_i^n - \theta_0)} \right\} \\ &= \lim_{n\to\infty} \left\{ (-1)^{i+1} c \sqrt{n\hat{p}_1\hat{p}_2} \sqrt{\frac{\hat{p}_{i'}}{n\hat{p}_i}} \right\} \left(\frac{1}{\lambda_{\theta_0}(t)} \right) \left. \frac{d\lambda_\theta(t)}{d\theta} \right|_{\theta=\theta_0} \\ &= (-1)^{i+1} p_{i'} c \left. \frac{d \log \lambda_\theta(t)}{d\theta} \right|_{\theta=\theta_0} . \end{aligned}$$

Hence,

$$\gamma(t) = \gamma_1(t) - \gamma_2(t) = c \left. \frac{d \log \lambda_\theta(t)}{d\theta} \right|_{\theta=\theta_0} .$$

This suggests we should try to find an SWL statistic whose weight function converges under H_0 to

$$Q(t) \alpha \left. \frac{d \log \lambda_\theta(t)}{d\theta} \right|_{\theta=\theta_0} \tag{5.18}$$

whatever the value of θ_0 and the definition of $G_1(t)$ and $G_2(t)$; such a test should have efficacy

$$e = e_{\max} = c^2 \int_0^\infty p_1^{-1} p_2^{-1} \left(\left. \frac{d \log \lambda_\theta(t)}{d\theta} \right|_{\theta=\theta_0} \right)^2 \Phi(t) \lambda_{\theta_0}(t)\, dt \qquad (5.19)$$

and be optimal in the class of SWL statistics for the family $\{F_\theta(t) : \theta \in \Theta\}$.

We comment here that testing

$$H_0 : \; F_1^n(t) = F_2^n(t) = F_{\theta_0}(t) \quad \text{for all } n \text{ and } t$$

against the given sequence of alternatives is equivalent to testing

$$H_0 : \; c^* = 0$$

against

$$H_1 : \; c^* = c$$

in the model

$$\theta_i^n = \theta_0 + (-1)^{i+1} c^* \sqrt{\frac{\hat{p}_{i'}}{n \hat{p}_i}}, \quad i \neq i'; \; i = 1, 2, \qquad (5.20)$$

where $c^* \in I$ is an unknown parameter, and where the interval I is such that, for all $c^* \in I$,

$$F_{\theta_1^n}(t) \geq F_{\theta_2^n}(t) \quad \text{for all } t \text{ and } n,$$

with, for each n, strict inequality at some $t \in \{t : \bar{G}_1(t)\bar{G}_2(t) > 0\}$.

Gill (1980) shows, under H_1, that the likelihood-ratio test is asymptotically normally distributed with mean $\sqrt{c^2 \sigma_L^2}$ and variance 1, where

$$\sigma_L^2 = \int_0^\infty (p_1 \pi_2(t) + p_2 \pi_1(t)) \left(\left. \frac{d \log \lambda_\theta(t)}{d\theta} \right|_{\theta=\theta_0} \right)^2 \lambda_{\theta_0}(t)\, dt.$$

Hence, as far as H_1 is concerned, the Pitman ARE of the optimal test in the class of SWL statistics with respect to the most powerful test is

$$\frac{\int_0^\infty \left(\frac{d \log \lambda_\theta(t)}{d\theta}\Big|_{\theta=\theta_0} \right)^2 \Phi(t)\lambda_{\theta_0}(t)\, dt}{p_1 p_2 \int_0^\infty (p_1\pi_2(t) + p_2\pi_1(t)) \left(\frac{d \log \lambda_\theta(t)}{d\theta}\Big|_{\theta=\theta_0} \right)^2 \lambda_{\theta_0}(t)\, dt} \leq 1,$$

with equality holding if and only if $G_1(t) = G_2(t)$ for all t.

It still remains to show, however, that an SWL statistic can be constructed for which (5.18) holds and hence (5.19) does too. We shall only do this in the special situation in which

$$F_\theta(t) = \Psi(g(t) + \theta), \quad t \in [0, \infty), \ \theta \in \Theta = (-\infty, +\infty), \tag{5.21}$$

where g is a nonrandom, absolutely continuous, nondecreasing function from $[0, \infty)$ onto $(-\infty, \infty)$, and where Ψ is an absolutely continuous d.f. with nonrandom, positive density ψ on $(-\infty, \infty)$ such that ψ', the derivative of ψ, exists and is continuous at all but finitely many points. We define $\beta(t) = \psi(t)/(1 - \Psi(t))$ and $l(t) = \log \beta(t)$, and note that

$$l'(t) = (\psi'(t)/\psi(t)) + \beta(t)$$

exists where $\psi'(t)$ does. We suppose that, except possibly on arbitrarily small neighbourhoods of at most finitely many points of $(-\infty, \infty)$, $l'(t)$ is of bounded variation on $(-\infty, \infty)$. Finally, according to some convention, $l'(t)$ is assigned finite values at the points $\pm\infty$ and at the points where $\psi'(t)$ does not exist.

The family defined by (5.21) is termed a "time transformed location family." For this particular case, the parameter c^* in (5.20) is an element of $[0, +\infty)$, and so $c > 0$.

Now, $F_\theta(t)$ has density $g'(t)\psi(g(t)+\theta)$, and so it has hazard rate $\lambda_\theta(t) = \beta(g(t)+\theta)g'(t)$. Since

$$\frac{d\log\lambda_\theta(t)}{d\theta} = l'(g(t)+\theta) = l'(\Psi^{-1}(F_\theta(t))),$$

we define our optimal weight function by

$$\hat{Q}_{opt}(t) = l'(\Psi^{-1}(\hat{F}_{\theta_0}(t))),$$

where $\hat{F}_{\theta_0}(t)$ is the Kaplan-Meier estimator of $F_{\theta_0}(t)$ based on the pooled sample of size n. Of course, $\hat{F}_{\theta_0}(t)$ can be replaced by either Altshuler's estimator or Prentice's estimator or the Peto-Peto estimator. Remark that neither θ_0 nor any of $g(t), \bar{G}_1(t)$, and $\bar{G}_2(t)$ enters into the specification of $\hat{Q}_{opt}(t)$ as we required.

Suppose for the moment that for all n and t,

$$F_1^n(t) = F_1(t) = \Psi(g(t)+\theta_0+\alpha), \quad F_2^n(t) = F_2(t) = \Psi(g(t)+\theta_0-\alpha),$$

where α is a parameter known to be greater than or equal to zero. Note here that, although $F_1(t)$ and $F_2(t)$ do not vary with n, they have the same basic parametric form as

$$F_1^n(t) = \Psi(g(t)+\theta_1^n), \quad F_2^n(t) = \Psi(g(t)+\theta_2^n) \tag{5.22}$$

respectively, where θ_i^n is defined as in (5.20), and where $c^* \in [0,\infty)$. The pair of parametric failure time models in (5.22), as a unit, is equivalent to the regression model

$$g(T_i) = E_i - \alpha z_i - \theta_0; \quad i = 1,\ldots,n,$$

where T_i is the failure time variate associated with item i, E_i is an error variable with d.f. Ψ, and

$$z_i = \begin{cases} 1 & \text{if } i \in \text{sample 1}, \\ -1 & \text{if } i \in \text{sample 2}. \end{cases}$$

The null hypothesis here is $H_0 : \alpha = 0$. Now, it can be shown that the score statistic generated from this regression model, using the generalized rank vector, is precisely $-2V_n$, where V_n is as in subsection 4.3.1 with the exception that the scores employ "Ψ," "ψ" in place of "F," "f" respectively. This score statistic, in turn, can be written as a WL statistic and is asymptotically equivalent to $2S_n$ with $\hat{Q}(t) = \hat{Q}_{opt}(t)$ (see subsection 4.3.3). The results of subsection 5.2.3 reveal, therefore, that $-2V_n$ (or $-V_n$) is the optimal WL statistic, under asymptotic conditions, against the alternative specified by (5.22).

To conclude this subsection, consider the alternative hypothesis H_1, not necessarily restricting ourselves to the parametric family (5.21). A classical family of rank statistics for uncensored data has the form

$$L_n = \sum_{i=1}^{n_1} \int_0^\infty \phi(\bar{\tilde{F}}_{\theta_0}(t)) \left\{ 1 - \frac{R_1(t)}{R(t)} \right\} dN_i(t) + \sum_{i=n_1+1}^{n} \int_0^\infty \phi(\bar{\tilde{F}}_{\theta_0}(t)) \left\{ -\frac{R_1(t)}{R(t)} \right\} dN_i(t),$$

where ϕ is a nonrandom function defined on $[0,1]$ such that

$$\int_0^1 \phi(u)\,du = 1,$$

and where

$$\bar{\tilde{F}}_{\theta_0}(t) = \frac{R(t)}{n}$$

is an estimator of $\bar{F}_{\theta_0}(t) = 1 - F_{\theta_0}(t)$ based on the combined sample. Since, under H_0,

$$\bar{\bar{F}}_{\theta_0}(t) \xrightarrow{P} \bar{F}_{\theta_0}(t) \quad \text{as } n \to \infty,$$

$n^{-1/2}L_n$ is asymptotically fully efficient against H_1 if and only if

$$\phi(\bar{F}_{\theta_0}(t)) = \left.\frac{d(\log \lambda_\theta(t))}{d\theta}\right|_{\theta=\theta_0}. \tag{5.23}$$

Now, we note that $\hat{\bar{F}}_{\theta_0}(t^-) = 1 - \hat{F}_{\theta_0}(t^-)$ reduces to $\bar{\bar{F}}_{\theta_0}(t)$ with censoring absent, and that

$$\hat{\bar{F}}_{\theta_0}(t^-) \xrightarrow{P} \bar{F}_{\theta_0}(t) \quad \text{as } n \to \infty.$$

Therefore, assuming condition (5.23) holds and that $G_1(t) = G_2(t)$ for all t, a WL statistic which is asymptotically fully efficient against H_1 and which reduces to $n^{-1/2}L_n$ in the absence of censoring has weight function

$$\hat{Q}(t) = \hat{Q}_E(t) = \phi(\hat{\bar{F}}_{\theta_0}(t^-)).$$

As far as parametric family (5.21) is concerned, under the assumption that $\phi(u) = l'[\Psi^{-1}(1-u)]$,

$$\hat{Q}_E(t) = l'[\Psi^{-1}(1 - \hat{\bar{F}}_{\theta_0}(t^-))].$$

5.2.4 Asymptotically Distribution-Free WL Statistics.

Assume in this subsection that the alternative hypothesis is fixed (the alternative does not vary with n), so that the null hypothesis is $H_0: F_1(t) = F_2(t) = F(t)$.

Recall from subsection 5.2.1 that

$$\sigma^2 = \text{Asvar}(n^{-1/2}S_n)$$
$$= \int_0^\infty J^2(\bar{F}(t), \bar{G}_1(t), \bar{G}_2(t)) p_1 p_2 \frac{\bar{G}_1(t)\bar{G}_2(t)}{p_1\bar{G}_1(t) + p_2\bar{G}_2(t)} f(t)\, dt,$$

where the nonrandom function J was there defined. A sufficient condition for σ^2 to be free of $\bar{F}(t)$, $\bar{G}_1(t)$, $\bar{G}_2(t)$, and $f(t)$ is that

$$Q(t) = J(\bar{F}(t), \bar{G}_1(t), \bar{G}_2(t))$$
$$= J^*(\bar{F}(t)) I(\bar{G}_1(t)\bar{G}_2(t) > 0) \sqrt{\frac{p_1\bar{G}_1(t) + p_2\bar{G}_2(t)}{\bar{G}_1(t)\bar{G}_2(t)}}, \tag{5.24}$$

where J^* is some nonrandom function defined on $[0, 1]$, and where $J^*(\bar{F}(t))$ is the null hypothesis limit of $\hat{Q}(t)$ with censoring absent (Leurgans 1984). For such a $Q(t)$,

$$\sigma^2 = \sigma^2_{DF} = \int_{u^*}^{1} J^{*2}(u) p_1 p_2 \, du,$$

where $u^* = \bar{F}(t^*)$ and $t^* = \sup\{t : \bar{G}_1(t)\bar{G}_2(t) > 0\}$. A weight function for which equation (5.24) is satisfied is given by

$$\hat{Q}(t) = \hat{Q}_{DF}(t) = \tilde{Q}(t) \sqrt{\frac{R_1(t)}{n\hat{\bar{F}}_1(t^-)} + \frac{R_2(t)}{n\hat{\bar{F}}_2(t^-)}} \sqrt{\frac{n_1}{R_1(t)} \frac{n_2}{R_2(t)}}$$
$$\times I(R_1(t)R_2(t) > 0) \sqrt{\hat{\bar{F}}_1(t^-)\hat{\bar{F}}_2(t^-)}$$
$$= \sqrt{n\hat{p}_1\hat{p}_2}\tilde{Q}(t) \sqrt{\frac{\hat{\bar{F}}_2(t^-)}{R_2(t)} + \frac{\hat{\bar{F}}_1(t^-)}{R_1(t)}} I(R_1(t)R_2(t) > 0), \tag{5.25}$$

where $\hat{\bar{F}}_i(t)$ is the Kaplan-Meier estimator of $\bar{F}_i(t) = 1 - F_i(t)$ based on sample i ($i = 1, 2$), where $\tilde{Q}(t)$ is a random weight function determined by

$$\left(\left\{R_1(T_i^o), R_2(T_i^o), dN_1^*(T_i^o) : i \leq N(t^-)\right\}, R_1(t), R_2(t), N(t^-)\right)$$

such that, under H_0, $\tilde{Q}(t) \xrightarrow{P} J^*(\bar{F}(t))$, and where

$$N_1^*(t) = \sum_{i=1}^{n_1} I(Y_i \leq t, \Delta_i = 1)$$

and $T_1^o < T_2^o < \cdots < T_K^o$ are the ordered uncensored failure times in the pooled sample of size n. Of course, one obvious possibility for $\tilde{Q}(t)$ is $J^*(\hat{\bar{F}}(t))$, where $\hat{\bar{F}}(t)$ is the Kaplan-Meier estimator of $\bar{F}(t)$ based on the combined sample. The asymptotic null distribution of a WL statistic with weight function $\hat{Q}_{DF}(t)$ will be free of $F(t)$, $G_1(t)$, and $G_2(t)$ (that is, distribution-free[7]). In particular, with $\hat{Q}(t)$ defined as in (5.25)

$$n^{-1/2}S_n \xrightarrow{D} N(0, \sigma_{DF}^2) \quad \text{as } n \to \infty.$$

Now, suppose for the moment that censoring is absent, and consider the rank statistic L_n of subsection 5.2.3 with "$\bar{\bar{F}}_{\theta_0}(t)$" there replaced by the notation "$\bar{\bar{F}}(t)$." Remark that L_n remains unchanged with $\phi(\bar{\bar{F}}(t))$ substituted with $\phi(\bar{\bar{F}}(t))I(R_1(t)R_2(t) > 0)$. In addition, we note that, with censoring absent, the square-root terms on the right-hand side of equation (5.25) vanish. Hence, recalling the points made in 5.2.3 concerning the relationship between $\bar{F}_{\theta_0}(t)$, $\bar{\bar{F}}_{\theta_0}(t)$, and $\hat{\bar{F}}_{\theta_0}(t^-)$, a censored data extension of $n^{-1/2}L_n$ which is asymptotically distribution-free under H_0 is a WL statistic whose weight function is given by (5.25), with $\tilde{Q}(t) = \phi(\hat{\bar{F}}(t^-))$.

5.2.5 Consistency of SWL Statistics.

Let H_A be some fixed alternative hypothesis, and let V be σ^2 or a consistent estimator thereof under H_0. If V is a random variable, assume, under H_A, that V is bounded away from both zero and $+\infty$ in probability as $n \to \infty$. Then a one-sided SWL statistic, $n^{-1/2}S_n/\sqrt{V}$, which rejects H_0 whenever $n^{-1/2}S_n/\sqrt{V} > z_{1-\alpha}$ is consistent against H_A if, under H_A,

$$n^{-1/2}S_n \xrightarrow{P} +\infty \quad \text{as } n \to \infty. \tag{5.26}$$

[7] From hereon, any statistic whose distribution is free of all underlying d.f.'s shall be referred to as being *distribution-free*.

Now, under H_A for $i = 1, 2$,

$$R_i(t)/n_i \xrightarrow{P} \bar{F}_i(t)\bar{G}_i(t) = \pi_i^*(t)$$

and

$$\hat{Q}(t) \xrightarrow{P} Q^*(t),$$

where $Q^*(t)$ is dependent on t via the functions $\bar{F}_1(t)$, $\bar{F}_2(t)$, $\bar{G}_1(t)$, and $\bar{G}_2(t)$. Therefore, letting

$$N_2^*(t) = \sum_{i=n_1+1}^{n} I(Y_i \leq t, \Delta_i = 1),$$

we have under H_A,

$$\begin{aligned} n^{-1}S_n &= n^{-1} \int_0^\infty \frac{\hat{Q}(t)R_1(t)R_2(t)}{R(t)} \left(\frac{dN_1^*(t)}{R_1(t)} - \frac{dN_2^*(t)}{R_2(t)} \right) \\ &\xrightarrow{P} \int_0^\infty K^*(t)(\lambda_1(t) - \lambda_2(t))\, dt, \end{aligned}$$

where

$$K^*(t) = \frac{Q^*(t)p_1p_2\pi_1^*(t)\pi_2^*(t)}{p_1\pi_1^*(t) + p_2\pi_2^*(t)},$$

and where $\lambda_i(t)$ is the hazard function associated with sample i ($i = 1, 2$). Hence, a sufficient condition for (5.26) to hold is

$$\int_0^\infty K^*(t)(\lambda_1(t) - \lambda_2(t))\, dt > 0. \tag{5.27}$$

Note that S_n–in addition to $n^{-1/2}S_n/\sqrt{V}$– is consistent against H_A if (5.27) is satisfied.

We now establish consistency of particular subsets of the class of WL statistics against two types of alternative hypotheses.

(1) The alternative $H_1 : \lambda_1(t) \geq \lambda_2(t)$ for all t, and $F_1(t) \neq F_2(t)$ for some t, is called the *ordered hazards* alternative.

(2) The alternative $H_2 : \bar{F}_2(t) \geq \bar{F}_1(t)$ for every t, and $F_1(t) \neq F_2(t)$ for some t, is called the alternative of *stochastic ordering*.

It is clear that H_1 implies H_2. For both H_1 and H_2, we assume that there exists a $t_0 > 0$ such that for $i = 1, 2$, $\bar{F}_2(t_0) > \bar{F}_1(t_0)$ and $\bar{G}_i(t_0) > 0$.

With respect to H_1, (5.27) will hold if $Q^*(t) > 0$, $\forall t \in I = (0, t^*)$, where $t^* = \sup\{t : \bar{G}_1(t)\bar{G}_2(t) > 0\}$. Hence, a WL statistic which is consistent against H_1 has $\acute{Q}(t) > 0$, $\forall t \in (0, T)$, where $T = \sup\{t : R_1(t)R_2(t) > 0\}$. Examples of such WL statistics include those of the Tarone-Ware class with $g(u)$ of subsection 2.5.3 such that $g(u) > 0$, $\forall u \in (0,1)$, as well as all other WL statistics discussed in section 2.5. $Q^*(t)$ is

$$p_1\pi_1^*(t) + p_2\pi_2^*(t)$$

for the Gehan statistic,

$$\frac{I(G_1(t)G_2(t) > 0)\bar{F}_1(t)\bar{F}_2(t)\,(p_1\pi_1^*(t) + p_2\pi_2^*(t))}{\pi_1^*(t)\pi_2^*(t)}$$

for the Efron statistic, and 1 for the log-rank statistic.

With respect to H_2, a sufficient condition for inequality (5.27) to hold is: $Q^*(t) > 0$, $\forall t \in I$, and $Q^*(t)$ is decreasing on I (Fleming and Harrington, 1991, p. 267). To prove this, remark that inequality (5.27) is, by integration by parts, equivalent to

$$-\int_0^\infty (\Lambda_1(t) - \Lambda_2(t))\,dK^*(t) > 0, \tag{5.28}$$

where

$$\Lambda_i(t) = \int_0^t \lambda_i(u)\,du, \quad i = 1, 2.$$

Now,

$$dK^*(t) = \frac{p_1p_2\pi_1^*(t)\pi_2^*(t)}{p_1\pi_1^*(t) + p_2\pi_2^*(t)}dQ^*(t) + Q^*(t)p_1p_2 \times \frac{p_1\pi_1^{*2}(t)d\pi_2^*(t) + p_2\pi_2^{*2}(t)d\pi_1^*(t)}{[p_1\pi_1^*(t) + p_2\pi_2^*(t)]^2}.$$

Since $\bar{F}_1(t)$ is continuous on $[0,\infty)$ and since there exists a $t_0 \in I$ such that $\bar{F}_1(t_0) < \bar{F}_2(t_0)$, there must exist a closed interval $I^* \subset I$ such that $\bar{F}_1(t) < \bar{F}_2(t)$, $\forall t \in I^*$, with $\bar{F}_1(t)$, and hence $\pi_1^*(t)$, strictly decreasing on I^*. Thus, assuming $Q^*(t) > 0$, $\forall t \in I$, and that $Q^*(t)$ is decreasing on I, we have, denoting the first and second summands of $dK^*(t)$ by $dK_1^*(t)$ and $dK_2^*(t)$ respectively, that

$$-\int_0^\infty (\Lambda_1(t) - \Lambda_2(t))\, dK_1^*(t) \geq 0,$$
$$-\int_{I^*} (\Lambda_1(t) - \Lambda_2(t))\, dK_2^*(t) > 0,$$

and

$$-\int_{I^{*c}\cap[0,\infty)} (\Lambda_1(t) - \Lambda_2(t))\, dK_2^*(t) \geq 0.$$

Hence, inequality (5.28) follows accordingly. Examples of weight functions with $Q^*(t)$ characterized as above are those of the log-rank, Gehan, Prentice, and Harrington-Fleming statistics. The Tarone-Ware class of WL statistics will be consistent against H_2 if $g(u) > 0$, for every $u \in (0,1)$, and if $g(u)$ is increasing on $(0,1)$. With regards to Efron's statistic,

$$K^*(t) = K_E^*(t) = p_1 p_2 \bar{F}_1(t)\bar{F}_2(t) I(\bar{G}_1(t)\bar{G}_2(t) > 0),$$

which is decreasing on I and strictly decreasing on I^*. Hence, (5.28) is satisfied with $K^*(t) = K_E^*(t)$, and so Efron's statistic is consistent against H_2.

Assume for the moment that censoring is absent, and consider the statistic L_n of subsection 5.2.4 as well as the alternative hypothesis H_A mentioned at the beginning of subsection 5.2.5. Under H_A,

$$n^{-1} L_n \xrightarrow{P} \int_0^\infty K_u^*(t)\,(d\Lambda_1(t) - d\Lambda_2(t)),$$

where

$$K_u^*(t) = \frac{p_1 p_2 \bar{F}_1(t)\bar{F}_2(t)}{p_1\bar{F}_1(t) + p_2\bar{F}_2(t)}\phi(p_1\bar{F}_1(t) + p_2\bar{F}_2(t)).$$

Now, suppose censoring is present, and suppose

$$\int_0^{t^*} K_u^*(t)\,(d\Lambda_1(t) - d\Lambda_2(t)) > 0.$$

Then a given WL statistic will be consistent against H_A if, $\forall t \in I$,

$$K^*(t) = K_u^*(t)$$
$$\iff Q^*(t) = I(\bar{G}_1(t)\bar{G}_2(t) > 0)\frac{(p_1\pi_1^*(t) + p_2\pi_2^*(t))\phi(p_1\bar{F}_1(t) + p_2\bar{F}_2(t))}{(p_1\bar{F}_1(t) + p_2\bar{F}_2(t))\bar{G}_1(t)\bar{G}_2(t)}.$$

Therefore, a censored data extension of L_n which is consistent against H_A is a WL statistic with weight function

$$\begin{aligned}\hat{Q}(t) &= \hat{Q}_c(t)\\ &= I(R_1(t)R_2(t) > 0)\\ &\quad\times\frac{n\hat{p}_1\hat{p}_2\phi\left(\hat{p}_1\hat{\bar{F}}_1(t^-) + \hat{p}_2\hat{\bar{F}}_2(t^-)\right)R(t)\hat{\bar{F}}_1(t^-)\hat{\bar{F}}_2(t^-)}{R_1(t)R_2(t)\left(\hat{p}_1\hat{\bar{F}}_1(t^-) + \hat{p}_2\hat{\bar{F}}_2(t^-)\right)}.\end{aligned}$$

5.2.6 Comparison of the Efficacy of Three Classes of WL Statistics.

The discussions of subsections 5.2.3, 5.2.4, and 5.2.5 lead us to infer that, given a rank statistic of the form L_n, there exists at least three possible extensions to accommodate censoring, assuming $G_1(t) = G_2(t) = G(t)$: asymptotically efficient (with respect to the alternative hypothesis H_1 of subsection 5.2.3), asymptotically distribution-free, and consistent WL statistics. Leurgans (1983) investigated the extent to which losses in efficiency occur when a nonoptimal extension is utilized. Specifically, her study involved censored data generalizations of the Savage and Wilcoxon statistic. Before describing details regarding design and results of the study, we present the various censored data counterparts of these two statistics.

Savage's test, with exact variance, is well known to be the locally most powerful rank test for the exponential scale family, and, with asymptotic variance or consistent estimator thereof, is asymptotically fully efficient against H_1. For this particular SWL statistic, $\phi(u) = 1, \forall u \in [0,1]$. Hence, the asymptotically efficient extension of Savage's statistic has

$$\hat{Q}_E(t) = 1,$$

which is the weight function corresponding to the log-rank statistic.

A Savage statistic extension with a distribution-free asymptotic variance has weight function

$$\hat{Q}_{DF}(t) = \sqrt{n\hat{p}_1\hat{p}_2} I(R_1(t)R_2(t) > 0)\sqrt{\frac{\hat{\bar{F}}_1(t^-)}{R_1(t)} + \frac{\hat{\bar{F}}_2(t^-)}{R_2(t)}},$$

while a consistent censored data analogue has

$$\hat{Q}_C(t) = n\hat{p}_1\hat{p}_2 \frac{I(R_1(t)R_2(t) > 0)\hat{\bar{F}}_1(t^-)\hat{\bar{F}}_2(t^-)R(t)}{\left(\hat{p}_1\hat{\bar{F}}_1(t^-) + \hat{p}_2\hat{\bar{F}}_2(t^-)\right) nR_1(t)R_2(t)}.$$

The Wilcoxon procedure is the locally most powerful rank test against time-transformed location alternatives for the logistic distribution, and is asymptotically fully efficient. For this particular test, $\phi(u) = u$; thus, the efficient censored data counterpart of Wilcoxon's statistic has

$$\hat{Q}_E(t) = \hat{\bar{F}}_{\theta_0}(t^-),$$

which is the Peto-Peto weight function. The consistent extension has

$$\hat{Q}_C(t) = n\hat{p}_1\hat{p}_2 I(R_1(t)R_2(t) > 0)\frac{\hat{\bar{F}}_1(t^-)\hat{\bar{F}}_2(t^-)R(t)}{R_1(t)R_2(t)},$$

which is the weight function corresponding to Efron's test. The distribution-free extension has yet to be studied and uses

$$\hat{Q}_{DF}(t) = \sqrt{n\hat{p}_1\hat{p}_2}I(R_1(t)R_2(t) > 0)\hat{\bar{F}}(t^-)\sqrt{\frac{\hat{\bar{F}}_1(t^-)}{R_1(t)} + \frac{\hat{\bar{F}}_2(t^-)}{R_2(t)}}.$$

Gehan's statistic, which has $\hat{Q}(t) = R(t)/n$, is not a member of any of the three classes of WL statistics presented in this subsection.

The parametric families of lifetime d.f.'s considered in Leurgan's study comprised, in fact, two special cases of (5.21): (1) the exponential scale family

$$F_{1,\theta}(t) = \Psi_1(g_1(t) + \theta) = 1 - \exp(-e^{\theta}t),$$

where $\Psi_1(x) = 1 - \exp(-e^x)$ and $g_1(t) = \log(t)$. and (2) the family

$$F_{2,\theta}(t) = \Psi_2\{g_2(t) + \theta\} = (e^t - 1)/(e^{-\theta} - 1 + e^t),$$

where $\Psi_2(x) = (1 + e^{-x})^{-1}$ and $g_2(t) = \log(e^t - 1)$. The alternative hypothesis for each family was specified by H_1 of subsection 5.2.3 with $\theta_0 = 0$ and $c = 1$. Note that $F_{1,0}(t) = F_{2,0}(t) = 1 - e^{-t}$, which is the standard exponential distribution. Hence, since for all censored data generalizations the variance, σ^2, depends on the null hypothesis failure time distribution as well as $G(t)$, σ^2 is the same for both parametric families, given a WL statistic and $G(t)$.

Two types of censoring distributions were utilized: truncated exponential censoring and uniform censoring. The truncated exponential censoring distributions were

$$G(t) = \begin{cases} 1 - e^{-\nu t}, & t \leq \tau, \\ 1, & t > \tau. \end{cases}$$

for $\nu = 0, 1, 2$ and $\tau = 2, \infty$. The uniform censoring densities were $I(a < t < b)/(b - a)$, for the choices $(a, b) = (0, 2)$ and $(1, 2)$.

Leurgans calculated the efficacy (equation (5.15) with

$$\gamma(t) = d(\log \lambda_\theta(t))/d\theta|_{\theta=\theta_0})$$

of all seven WL statistics against the two alternatives for each type of censoring distribution. These calculations revealed the following:

(1) The consistent extensions (especially that of the Savage statistic) are more sensitive to heavy censoring than the other censored data counterparts.

(2) The distribution-free analogue has high efficiency relative to the optimal statistic when censoring is mild. Only when censoring events occur at twice the intensity of observed failures does the efficacy of the distribution-free statistic become substantially lower than that of the optimal extension.

(3) As the intensity of censoring events increases, the Pitman ARE of Gehan's statistic with respect to the Peto-Peto statistic against the logistic location alternative decreases.

(4) Gehan's statistic is less efficacious than the log-rank and Peto-Peto statistic against both alternatives for all types of censoring.

5.3 Small-Sample Properties.

5.3.1 Small-Sample Null Distribution of SWL Statistics.

From now throughout the remainder of this chapter, we employ the two-sample set-up that was described in subsection 5.2.1 as well as the notation of subsections 5.2.1, 5.2.4, and 5.2.5; however, unless otherwise specified, we do not restrict our censoring schemes to the censoring model of section 5.2.

In this and the following two paragraphs, we assume that the censoring mechanism operates in the same manner for all n items, that is, that the censoring patterns of the two samples are equal. Consider the WL statistic S_n of subsection 5.2.1. This statistic can be written as

$$\sum_{i=1}^{n_1} A_i, \tag{5.29}$$

where

$$A_i = \left[\hat{Q}(Y_i) - \int_0^{Y_i} \frac{\hat{Q}(t)}{R(t)}\,dN(t)\right]^{\Delta_i} \left[-\int_0^{Y_i} \frac{\hat{Q}(t)}{R(t)}\,dN(t)\right]^{1-\Delta_i}$$

is the score corresponding to item i ($i = 1, \ldots, n$), where $T_1^o < \cdots < T_K^o$ are the ordered uncensored failure times of the pooled sample of size n, and where

$$N(t) = \sum_{i=1}^{n} I(Y_i \le t, \Delta_i = 1).$$

Of course, (5.29) depends on the vector of pairs of random variables

$$\mathbf{P}_1 = ((N(Y_1), \Delta_1), \ldots, (N(Y_n), \Delta_n)),$$

which is the generalized rank vector. Now, let $\mathbf{P}_1^*$ be the vector $\mathbf{P}_1$ without item labels attached. Treating all identical elements of $\mathbf{P}_1^*$ as distinct entities, we note that, under H_0 and conditional on $\mathbf{P}_1^*$, all $\binom{n}{n_1}$ subsets of size n_1 from $\mathbf{P}_1$ have the same chance of belonging to sample 1. Hence, letting S equal the set of all $\binom{n}{n_1}$ such subsets, an exact test of H_0 in very small samples can be based on the nonrandom conditional probability function

$$h_1(a) = \Pr_{H_0}(S_n = a | \mathbf{P}_1^*) = N_{1s}(a) \Big/ \binom{n}{n_1},$$

where $N_{1s}(a)$ is the number of elements in S such that $S_n = a$. Remark that $h_1(a)$ is independent of $F(t)$ as well as all d.f.'s related to the censoring mechanism. Hence, the associated expectation and variance are distribution-free. In particular,

$$E_{H_0}(S_n|\mathbf{P}_1^*) = E_0 = \sum_{l\in(-\infty,+\infty)} \frac{lN_{1s}(l)}{\binom{n}{n_1}}$$

and

$$\mathrm{Var}_{H_0}(S_n|\mathbf{P}_1^*) = \sum_{l\in(-\infty,+\infty)} \frac{l^2 N_{1s}(l)}{\binom{n}{n_1}} - E_0^2.$$

Now, let

$$\mathbf{P}_2 = (K, A_1, \ldots, A_n),$$

and treat all identical A_i's as distinct entities. Then another distribution-free, nonrandom conditional probability function for S_n is

$$h_2(a) = \Pr_{H_0}(S_n = a|\mathbf{P}_2^*) = N_{2s}(a) \Big/ \binom{n}{n_1},$$

where $\mathbf{P}_2^*$ is the vector $\mathbf{P}_2$ without item labels attached to the A_i's, and where $N_{2s}(a)$ has an analogous definition as $N_{1s}(a)$. Recall from section 2.2 that $h_2(a)$ is in fact the probability function corresponding to the permutation distribution of S_n, while

$$\mathrm{Var}_{H_0}(S_n|\mathbf{P}_2^*) = \frac{n_1 n_2}{n(n-1)} \sum_{i=1}^{n} A_i^2$$

is the associated permutation variance. In the next paragraph, we give conditions under which $h_1(a) = h_2(a)$, $\forall a \in (-\infty, +\infty)$. Thus, if these conditions hold and if the censoring mechanism renders $\mathbf{P}_1^*$ nonrandom (for example, simple and progressive type II censoring), then obviously $h_2(a)$ is the exact probability function of S_n under H_0.

In addition to Δ_i, A_i is a function of

$$\left\{R_1(T^o_{j'}), R_2(T^o_{j'}), dN^*_1(T^o_j) :\ j \leq N(Y^{*-}_i),\ j' \leq N(Y^*_i)\right\},$$

where $Y^*_i = \sup\{T^o_j :\ T^o_j \leq Y_i; j = 1, \ldots, K\}$. Suppose, though, that A_i depends only on

$$\mathcal{B}_i = \left(\Delta_i, \{R(T^o_j) :\ j \leq N(Y^*_i)\}\right).$$

Then $\mathbf{P}^*_2$ is a function of strictly

$$(K, R(T^o_1), \ldots, R(T^o_K)),$$

which is an equivalent form of $\mathbf{P}^*_1$. Hence, if we suppose further that $f_{\mathbf{P}^*_1|\mathbf{P}_2}(\mathbf{p}^*_1|\mathbf{p}^*_2)$ is degenerate, then, $\forall a \in (-\infty, +\infty)$,

$$\begin{aligned}
\Pr{}_{H_0}(S_n = a|\mathbf{P}^*_1) &= \Pr{}_{H_0}(S_n = a|\mathbf{P}^*_1, \mathbf{P}^*_2) \\
&= \Pr{}_{H_0}(S_n = a|\mathbf{P}^*_2) \\
&\Rightarrow h_1(a) = h_2(a).
\end{aligned} \tag{5.30}$$

Examples of scores for which (5.30) holds include those of the log-rank test, Gehan's test, and Prentice's test. On the other hand, A_i of Efron's test is a function of

$$\left\{dN^*_1(T^o_j), R_1(T^o_{j'}) :\ j \leq N(Y^{*-}_i),\ j' \leq N(Y^*_i)\right\}$$

as well as $\mathcal{B}_i$, and so, in this case, equation (5.30) does not necessarily hold.

If the censoring mechanism is not identical for all n items, a test of H_0 via either $h_1(a)$ or $h_2(a)$ is not valid. Furthermore, even if the two groups have equal censoring patterns, calculation of $h_1(a)$ or $h_2(a)$ for moderate size samples is extremely laborious. An alternative approach for testing H_0 in small samples, independent of sample size and equality status of censoring patterns, relies on approximating the unconditional null distribution of a standardized version of S_n. We address this matter in subsections 5.3.1.1 and 5.3.1.2.

5.3.1.1 Goodness of Fit of Normal Distribution.

(a) Results of Monte Carlo Simulations.

If there is a sufficiently large number of failures in each sample, the null distribution of the SWL statistic $S_n/\sqrt{V}$ is well approximated by the standard normal distribution. Here, V is $\mathrm{Var}_{H_0}(S_n)$ (if computable) or a suitable estimator thereof under H_0. The adequacy of this approximation in relatively small samples, however, is an issue subject to much controversy. Throughout the last fifteen years, several Monte Carlo studies have been performed to assess the goodness of fit of the normal distribution as a function of variance estimator, censoring mechanism, censoring intensity, size of combined sample, equality status of sample sizes (that is, equality versus inequality of sample sizes), and equality status of censoring patterns of the two samples. In this subsection, we present the results of three significant studies: Latta (1981), Breslow et al. (1984), and Groggel et al. (1988). Rather than examine each Monte Carlo investigation individually, we describe the design and major outcomes of the three studies as a whole.

The WL statistics studied had $\hat{Q}(t) = 1$, $R(t)/n$, and $\tilde{\bar{F}}_p(t)$, where $\tilde{\bar{F}}_p(t)$ is Prentice's estimator of $\bar{F}(t)$ based on the pooled sample of size n. With respect to the first two of these WL statistics, the variance estimators considered were the conditional permutation variance and the permutation variance. As far as Prentice's statistic is concerned, the variance estimators utilized were the two aforementioned ones as well as $\mathcal{V}_0$ of chapter 4. The lifetime variates generated were either exponential ($\lambda = 1.0$), or Weibull (shape parameter $\gamma = 4.0$ and scale parameter $\lambda = 1.0$), or log-normal ($\mu = 0$, $\sigma^2 = 1$), or uniform $(0, 1)$ random variables. The censoring schemes considered were as follows:

(a) simple type I censorship;

(b) fixed censorship, with entry times uniformly distributed on $[0, a]$ and termination time of study at $t = a$;

(c) a variation of fixed censorship such that, for one sample, the entry times are uniformly distributed on $[0, a]$ with termination time $t = a$, while, for the other sample, censoring is absent;

(d) random censorship model of section 5.2 with $\bar{G}_1(t) = \bar{G}_2(t) = e^{-\beta t}$ for some $\beta > 0$;

(e) random censorship model of section 5.2 with $\bar{G}_i(t) = e^{-\beta t}$, for some $\beta > 0$, and with $\bar{G}_{i'}(t) = 1$ $(i \neq i')$;

(f) cases (a) and (d) present together;

(g) cases (a) and (e) present together;

(h) two-stage progressive type II censorship with $n_1 = n_2 = 50$ and 87 items censored at the first failure time (this case simulates early heavy censoring).

For each combination of lifetime distribution, sample sizes, and censoring scheme, the data

$$(Y_1, \Delta_1), \ldots, (Y_n, \Delta_n)$$

were generated independently a predetermined number of times (say N), thus yielding N values of all SWL statistics in question (two studies used 1000 repetitions, while the other used 5000). Afterwards, observed tail probabilities corresponding to several predetermined critical values were recorded. The nominal tail probabilities, therefore, are type I errors obtained via the standard normal distribution. We now summarize the major results of the studies:

(1) The null distribution of an SWL statistic, regardless of variance estimator, appears to be approximately normal as long as the sample sizes are equal, the censoring patterns are equal, and the censoring percentage is less than 50%.

(2) For censoring schemes (a), (b), (d), and (f), all SWL statistics give very conservative observed error levels (relative to the nominal levels) for 90% censoring in both samples with $n_1 = n_2 = 10$.

(3) For censoring situation (h), Gehan's test, with either the permutation variance or conditional permutation variance, provides type I error levels which are extremely conservative. Tests based on the other two WL statistics, though, have observed type I errors which agree well with the nominal ones.

(4) Discrepancies between observed and nominal error levels are generally greater for censoring situations (b) and (d) than for case (a).

(5) When the two samples differ significantly in terms of their expected number of uncensored failure times, all tests with conditional permutation variance have a skewed distribution, with greater tail probability for the critical region where we infer that the failure times of the sample with more uncensored observations are longer. In other words, the sample which was more likely to yield the greater number of uncensored lifetimes was much

more likely than the other sample to be declared as having the longer lifetimes. This bias is attributable to the fact that when the failure times of the sample with expectedly more uncensored observations appear to be longer than those of the other sample, the conditional permutation variance tends to be smaller, and hence the absolute value of the SWL statistic is larger, than when the opposite occurs. Breslow et al. (1984) provide, for the log-rank test, a scattergram plot of numerator versus denominator which illustrates the correlation between the two.

(6) With respect to censoring schemes (c) and (e), if $n_1 \leq n_2$ and sample 2 is uncensored, or if $n_1 \geq n_2$ and sample 1 is uncensored, then an SWL statistic with permutation variance generally yields conservative error levels. This observation is consistent with Jennrich's (1983) asymptotic results (see subsection 5.2.1).

(7) If the samples experience equal censoring patterns but the sample sizes differ, a test with a permutation variance generally outperforms one with the same WL statistic and a conditional permutation variance. For a high percentage of censoring, however, (about 90%) this discrepancy is very small. This difference in small-sample null distribution between the two SWL statistics can be plausibly explained as follows: As indicated in the discussion of result (5), an SWL statistic with conditional permutation variance for any particular weight function and not merely for those in question has, under inequality of expected number of uncensored failure times in each sample, a skewed distribution due to the correlation between numerator and denominator. As far as the three weight functions of interest are concerned, this correlation is ultimately attributable to the fact that

the conditional permutation variance uses all the information provided by

$$\{R_1(T_j^o), R_2(T_j^o) : j \leq K\}.$$

(Recall that S_n is dependent on $\{R_1(T_i^o), R_2(T_i^o), dN_1^*(T_i^o) : i \leq K\}$.) On the other hand, the permutation variance for each of the S_n's considered is a function of

$$\{R(T_j^o) : j \leq K\}$$

only, and so should be less strongly associated with the numerator under inequality of expected number of uncensored failure times. From the above arguments, therefore, we conclude that, under equal censoring patterns but unequal sample sizes, the observed type I error of an SWL statistic with permutation variance should be closer to the nominal value than the observed error level of a test with the same S_n, but with a conditional permutation variance.

(8) Whenever the sample sizes and/or censoring patterns are unequal, Prentice's test, with variance estimator $\mathcal{V}_0$ of chapter 4, provides error levels which are generally closer to the nominal values than those of the other SWL statistics.

(9) Finally, lifetime distribution does not have much effect on the performance of a given test, keeping all other parameters fixed.

(b) Alternative Variance Estimators.

In this subsection, we assume that the censoring model of section 5.2 is in effect. Furthermore, we maintain the same conventions concerning $\bar{F}(t)$, $\bar{G}_i(t)$ $(i = 1, 2)$ as were stated in subsection 5.2.1.

To eliminate the skewness in small-sample distribution induced by the conditional permutation variance when the expected number of uncensored observations in one sample differs from that of the other, one could instead use

$$\tilde{\sigma}^2 = \int_0^\infty E_{H_0}\left[\hat{Q}^2(t)R_1(t)R_2(t)R^{-2}(t)|\mathbf{A}(t)\right] dN(t).$$

Here, $\hat{Q}(t)$ is assumed to be a function of

$$\mathbf{B}(t) = (R(t), N(t^-), \{R(T_i^o): \ i \leq N(t^-)\})$$

only, while $\mathbf{A}(t)$ consists of $\mathbf{B}(t)$ as well as other necessary information from the study with the proviso that, under H_0, $dN(t)$ and $\mathbf{A}(t)$ are conditionally independent given $R(t)$. We thus have

$$\begin{aligned} E_{H_0}[dN(t)|\mathbf{A}(t)] &= E_{H_0}[dN(t)|\mathbf{A}(t), R(t)] \\ &= E_{H_0}[dN(t)|R(t)] \\ &= R(t)\lambda(t)\,dt \end{aligned}$$

(Brown, 1984), and so

$$\begin{aligned} E_{H_0}(\tilde{\sigma}^2) &= \int_0^\infty E_{H_0}\left[\frac{\hat{Q}^2(t)R_1(t)R_2(t)}{R(t)}\right]\lambda(t)\,dt \\ &= \int_0^\infty E_{H_0}\left[\frac{\hat{Q}^2(t)R_1(t)R_2(t)}{R^2(t)} E_{H_0}[dN(t)|R_1(t), R_2(t), \mathbf{B}(t)]\right] \\ &= E_{H_0}(V_{cp}) \\ &= \mathrm{Var}_{H_0}(S_n), \end{aligned}$$

where V_{cp} is the conditional permutation variance of S_n, and where the equality

$$E_{H_0}[dN(t)|R_1(t), R_2(t), \mathbf{B}(t)] = E_{H_0}[dN(t)|R(t)]$$

has been used (Brown, 1984).

Now, suppose $G_1(t) = G_2(t)$, $\forall t \in [0, \infty)$, with $n_1 \neq n_2$. Then, letting $\mathbf{A}(t) = \mathbf{B}(t)$ and noting that the conditional distribution of $R_1(t)$, given $R(t)$, is hypergeometric, we have

$$\begin{aligned} E_{H_0}[R_1(t)R_2(t)|\mathbf{A}(t)] &= E_{H_0}[R_1(t)R_2(t)|R(t)] \\ &= \frac{R(t)(R(t)-1)n_1n_2}{n(n-1)}. \end{aligned}$$

Hence,

$$\tilde{\sigma}^2 = \int_0^\infty \frac{n_1n_2\hat{Q}^2(t)(R(t)-1)}{n(n-1)R(t)}\,dN(t),$$

which is in fact the permutation variance of S_n.

If $G_1(t)$ and $G_2(t)$ are not identical, then $\tilde{\sigma}^2$, with $\mathbf{A}(t) = \mathbf{B}(t)$, depends on $(G_1(t), G_2(t), F(t))$ in a complex way. Specifically,

$$E_{H_0}\{R_1(t)R_2(t)|R(t) = r_t\} = \begin{cases} \dfrac{\sum_l l(r_t-l)\binom{n_1}{l}\binom{n_2}{r_t-l}c^l(t)}{\sum_l \binom{n_1}{l}\binom{n_2}{r_t-l}c^l(t)} & \text{if } t \in \mathcal{A}, \\ 0 & \text{if otherwise,} \end{cases} \tag{5.31}$$

where the sums are over $l = \alpha, \ldots, \beta$, and where

$$\mathcal{A} = \{t : \bar{G}_1(t)\bar{G}_2(t) > 0\},$$

$$\alpha = \max(0, r_t - n_2),$$

$$\beta = \min(r_t, n_1),$$

$$c(t) = \frac{\bar{G}_1(t)\{1 - \bar{F}(t)\bar{G}_2(t)\}}{\bar{G}_2(t)\{1 - \bar{F}(t)\bar{G}_1(t)\}}.$$

Remark, however, that if $\bar{G}_1(t) = \bar{G}_2(t)$ $\forall t \in \mathcal{A}$, then conditional expectation (5.31) is

$$\frac{r_t(r_t-1)n_1n_2}{n(n-1)} I(\bar{G}_1(t)\bar{G}_2(t) > 0),$$

and so

$$\tilde{\sigma}^2 = \int_{\mathcal{A}} \frac{n_1n_2\hat{Q}^2(t)(R(t)-1)}{n(n-1)R(t)}\,dN(t). \tag{5.32}$$

An example of a censoring model where the variance estimator (5.32) is appropriate is one in which all items are put on test at the same instant, yet termination of the study–the only cause of censoring—occurs at time T_1^* for group 1 and T_2^* for group 2 ($T_1^* \neq T_2^*$). If, on the other hand,

$$\bar{G}_1(t) \neq \bar{G}_2(t) \text{ for at least one } t \in \mathcal{A}, \tag{5.33}$$

then $\tilde{\sigma}^2$, with $\mathbf{A}(t) = \mathbf{B}(t)$, cannot be implemented since it depends on $F(t)$, which is assumed unknown, as well as on $G_1(t)$ and $G_2(t)$, which are most probably unknown. Under condition (5.33), therefore, rather than try to estimate $\tilde{\sigma}^2$, we opt for the approach of incorporating other elements into $\mathbf{A}(t)$ to render $\tilde{\sigma}^2$ distribution-free. The elements chosen are in accordance with the suggestions of Brown (1984).

Case 1: Case where all potential censoring time variates are observable

Suppose all potential censoring times, including those corresponding to uncensored lifetimes, are observable (for example, censoring scheme (c)). Define $L_j(t)$, for $j = 1, 2$, to be the random variable which indicates the number of group j items with potential censoring time $\geq t$, and let $L(t) = L_1(t) + L_2(t)$. There are $\binom{L(t)}{R(t)}$ possible risk sets of size $R(t)$ at time t, each of which is equally likely under H_0 to be observed. Thus, letting

$$\mathbf{A}(t) = (\mathbf{B}(t), L_1(t), L_2(t)),$$

we have

$$\Pr_{H_0}(R_1(t) = r_1(t) | \mathbf{A}(t)) = \frac{\binom{L_1(t)}{r_1(t)}\binom{L_2(t)}{R(t)-r_1(t)}}{\binom{L(t)}{R(t)}},$$

which is the probability function corresponding to the hypergeometric distribution. Remark that, in agreement with our original definition of $\mathbf{A}(t)$,

$$E_{H_0}(dN(t)|\mathbf{A}(t)) = E_{H_0}(dN(t)|R(t)).$$

It easily follows that

$$E_{H_0}\left[R_1(t)R_2(t)|\mathbf{A}(t)\right] = \frac{L_1(t)L_2(t)R(t)(R(t)-1)}{L(t)(L(t)-1)}.$$

Therefore,

$$\tilde{\sigma}^2 = \int_0^\infty \frac{\hat{Q}^2(t)L_1(t)L_2(t)(R(t)-1)}{L(t)(L(t)-1)R(t)}\,dN(t). \tag{5.34}$$

Now,

$$L_j(t)/n_j \xrightarrow{P} \bar{G}_j(t)$$

as $n \to \infty$ $(j = 1, 2)$. Moreover,

$$L(t)/n \xrightarrow{P} p_1\bar{G}_1(t) + p_2\bar{G}_2(t).$$

Hence, under H_0,

$$\begin{aligned} n^{-1}\tilde{\sigma}^2 &\xrightarrow{P} \int_0^\infty Q^2(t)\Phi(t)\lambda(t)\,dt \\ &= \text{Asvar}_{H_0}(n^{-1/2}S_n). \end{aligned}$$

Case 2: Case where not all potential censoring time variates are observable

Suppose, now, that the potential censoring times corresponding to uncensored lifetimes are unobservable and that both $G_1(t)$, $G_2(t)$ are completely unspecified. (Such a censoring mechanism would, for example, be present in a clinical trial where one cause of censoring for both samples is withdrawal from the study due to severe side effects of the treatment.); hence, $L_i(t)$ $(i = 1, 2)$ cannot be observed exactly $\forall t \in (T_{is}, +\infty)$, where T_{is} is the smallest uncensored failure time in sample i. We know, however, that

$$L_1(t) = R_1(t) + \sum_{j=1}^{n_1} I(C_j \geq t, N_j(t^-) = 1), \tag{5.35}$$

where $N_j(t) = I(Y_j \le t, \Delta_j = 1)$ $(j = 1, \ldots, n)$. Now, although the second summand of (5.35) cannot be observed exactly $\forall t \in (T_{1s}, +\infty)$, it can be approximated in this interval by

$$E\left[\sum_{j=1}^{n_1} I(C_j \ge t, N_j(t^-) = 1) | \{T_i^o, (i), C_{(i)} \ge T_i^o : i \le N(t^-)\}\right]$$

$$= E\left[\sum_{j \;\; T_j^o < t} \left(\Pr(C_{(j)} \ge t | C_{(j)} \ge T_j^o, (j), T_j^o\right) dN_1^*(T_j^o))\right]$$

$$= E\left[\int_0^{t^-} \frac{\bar{G}_1(t)}{\bar{G}_1(u)} \, dN_1^*(u)\right], \tag{5.36}$$

where (i) is the item label corresponding to T_i^o. Expectation (5.36) can be estimated by

$$\int_0^{t^-} \frac{\hat{\bar{G}}_1(t)}{\hat{\bar{G}}_1(u)} \, dN_1^*(u),$$

where $\bar{G}_1(t)$ is the Kaplan-Meier estimator of $\bar{G}_1(t)$. Therefore, $L_1(t)$ can be approximated by

$$\hat{L}_1(t) = R_1(t) + \int_0^{t^-} \frac{\hat{\bar{G}}_1(t)}{\hat{\bar{G}}_1(u)} \, dN_1^*(u).$$

Similarly, $\hat{L}_2(t)$ is defined. Substituting $\hat{L}_1(t)$, $\hat{L}_2(t)$, $\hat{L}(t) = \hat{L}_1(t) + \hat{L}_2(t)$ for $L_1(t)$, $L_2(t)$, and $L(t)$ in equation (5.34), we obtain

$$\tilde{\sigma}_A^2 = \int_0^{\infty} \frac{\hat{Q}^2(t)\hat{L}_1(t)\hat{L}_2(t)(R(t) - 1)}{\hat{L}(t)(\hat{L}(t) - 1)R(t)} \, dN(t),$$

which is an approximation to $\tilde{\sigma}^2$ defined by (5.34).

Now, under H_0, as $n \to \infty$

$$\begin{aligned}
\hat{L}_j(t)/n_j \xrightarrow{P} \pi_j(t) + \bar{G}_j(t) \int_0^{t^-} f(u) \, du \\
&= \pi_j(t) + \bar{G}_j(t)F(t) \\
&= \bar{G}_j(t), \quad j = 1, 2;
\end{aligned}$$

and

$$\hat{L}(t)/n \xrightarrow{P} p_1 \bar{G}_1(t) + p_2 \bar{G}_2(t),$$

where $f(u) = \lambda(u)\bar{F}(u)$. Therefore, under H_0,

$$n^{-1}\tilde{\sigma}_A^2 \xrightarrow{P} \text{Asvar}_{H_0}(n^{-1/2}S_n).$$

Note that $\tilde{\sigma}_A^2$, like $\tilde{\sigma}^2$ of case 1, is not exclusively a function of the generalized rank vector.

We conclude subsection 5.3.1.1 (b) by commenting that

$$\begin{aligned}\text{Var}_{H_0}(S_n) &= E_{H_0}\left[\int_0^\infty \frac{\hat{Q}^2(t)R_1(t)R_2(t)}{R^2(t)}\, dN(t)\right] \\ &= \int_0^\infty E_{H_0}\left[\frac{\hat{Q}^2(t)R_1(t)R_2(t)}{R(t)}\right]\lambda(t)\, dt, \qquad (5.37)\end{aligned}$$

even if $\hat{Q}(t)$ is unrestricted and is allowed to be determined by

$$(\{R_1(T_i^o), R_2(T_i^o), dN_1^*(T_i^o) :\ i \leq N(t^-)\}, R_1(t), R_2(t), N(t^-)).$$

If the expectation in the integrand of (5.37) is dependent upon $G_1(t)$, $G_2(t)$, $F(t)$, and if these d.f.'s are unspecified, then this integral is nonevaluable. On the other hand, if this expectation is determined by $F(t)$ only, then (5.37) is free of $F(t)$ and hence calculable (assuming $\sup\{t :\ \bar{G}_1(t)\bar{G}_2(t) > 0\}$ is specified), regardless of whether or not the above d.f.'s are known. Finally, suppose that $F(t)$ is unknown but that $G_1(t)$, $G_2(t)$ are discrete and specified. Then, if the expectation in question is dependent on these d.f.'s, we can show that $\text{Var}_{H_0}(S_n)$ is free of $F(t)$ using a methodology similar to that discussed on pp. 79-80.

5.3.1.2 Goodness of Fit of Bootstrap Distribution.

This subsection evaluates, via Monte Carlo simulations, the level of accuracy relative to the standard normal distribution–to which the bootstrap distribution of an SWL statistic approximates the true null distribution. At the time of writing this thesis, no other researcher had previously considered this problem. Before presenting the design and results of our study, we briefly review the bootstrap method as devised by Efron (1979).

Consider the sample

$$\mathbf{X} = (X_1, \ldots, X_n),$$

where the X_i are i.i.d. according to some unknown d.f. H, and where the X_i may be of more than one dimension. Let $\mathbf{x} = (x_1, \ldots, x_n)$ denote the realized values of $\mathbf{X}$. The problem we wish to solve then is the following: Given a specified random variable $R(\mathbf{X})$, estimate the sampling distribution of R on the basis of the observed data $\mathbf{x}$.

The bootstrap method consists of first constructing the sample probability distribution $\hat{H}$, putting mass $1/n$ at each point $x_1, \ldots, x_n$. We then draw the sample $\mathbf{X}^* = (X_1^*, \ldots, X_n^*)$ from $\hat{H}$, where each X_i^* independently takes value x_j with probability $1/n$, $j = 1, \ldots, n$. In other words, the values of $\mathbf{X}^*$ are selected with replacement from the set $\{x_1, x_2, \ldots, x_n\}$. Finally, we approximate the sampling distribution of $R(\mathbf{X})$ by the distribution of $R^* = R(\mathbf{X}^*)$. We refer to the latter distribution as the *bootstrap distribution* of R and to $\mathbf{X}^*$ as the *bootstrap sample.*

As far as our particular investigation is concerned, we wished to assess the adequacy of the bootstrap distribution (for $\rho = 0, \frac{1}{2}, 1$) of

$$U_n(\mathbf{X}) = \frac{\int_0^\infty \left[\hat{\bar{F}}_{KM}(t)\right]^\rho \left(dN_1^*(t) - \frac{R_1(t)}{R(t)}\, dN(t)\right)}{\sqrt{\int_0^\infty \left[\hat{\bar{F}}_{KM}(t)\right]^{2\rho} \frac{R_1(t)R_2(t)}{R^2(t)}\, dN(t)}},$$

where $\mathbf{X} = ((Y_1, \Delta_1), \ldots, (Y_n, \Delta_n))$, where the (Y_i, Δ_i) are i.i.d., and where $\hat{\bar{F}}_{KM}(t)$ is the Kaplan-Meier estimator of $\bar{F}(t)$. U_n is, of course, the Harrington-Fleming class of tests with conditional permutation variance. The failure time d.f. considered was $F(t) = 1 - e^{-t}$, while the censoring mechanism utilized was the random censorship model of section 5.2 with

$$\Pr(C_i < t) = G_1(t) = G_2(t) = 1 - e^{-(3t/7)}; \quad i = 1, 2, \ldots, n.$$

Hence, $\Pr(\Delta_i = 1) = \Pr(T_i < C_i) = 0.7$. The sample size configurations for $\rho = 0.0,\ 0.5$ were

(1) $n_1 = 12,\ n_2 = 8$;

(2) $n_1 = n_2 = 10$;

(3) $n_1 = 20,\ n_2 = 10$;

(4) $n_1 = n_2 = 15$;

(5) $n_1 = 30,\ n_2 = 10$.

For $\rho = 1.0$, we used just configuration (5).

We first generated the simple random samples, $(T_1, \ldots, T_n)$ and $(C_1, \ldots, C_n)$, following which we obtained

$$((Y_1, \Delta_1), \ldots, (Y_n, \Delta_n)),$$

where $Y_i = \min(T_i, C_i)$ $(i = 1, \ldots, n)$. We then generated independent realizations of the bootstrap sample

$$\mathbf{X}^* = ((Y_1^*, \Delta_1^*), \ldots, (Y_n^*, \Delta_n^*)),$$

$\mathbf{x}^{*1}, \mathbf{x}^{*2}, \ldots, \mathbf{x}^{*N}$, with $N = 1000$. (The first n_1 elements of $\mathbf{X}^*$ constituted sample 1, while the remaining elements formed sample 2.) The histogram of the corresponding values $U_n(\mathbf{x}^{*1}), \ldots, U_n(\mathbf{x}^{*N})$ was then taken as an approximation to the actual distribution of $U_n(\mathbf{X}^*)$. Specifically, we determined the sequence

$$P_{1.0},\ P_{2.5},\ P_{5.0},\ P_{10.0},\ P_{90.0},\ P_{95.0},\ P_{97.5}, \text{ and } P_{99.0},$$

where P_i denotes the i^{th} percentile point. We refer to these values as "bootstrap percentile points" and to the histogram constructing process as a "trial." For each case of SWL statistic with particular sample size configuration, ten trials were performed using ten independently generated $\mathbf{X}$'s, following which we calculated the mean and standard deviation of each bootstrap percentile point over the ten trials.

For each sample size configuration, we as well generated 10,000 independent realizations of

$$\mathbf{X} = \{(Y_1, \Delta_1), \ldots, (Y_n, \Delta_n)\},$$

$\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^{10,000}$. This process in turn yielded, for each ρ, a sequence of corresponding values

$$U_n(\mathbf{x}^1), U_n(\mathbf{x}^2), \ldots, U_n(\mathbf{x}^{10,000}),$$

the histogram of which provided an excellent approximation to the exact null distribution of U_n. The sequence of percentile points which we determined in this case coincides with that of the previous paragraph, and the actual values obtained are referred to as the "true percentile points".

Table: Assessment of goodness of fit of bootstrap distribution of U_n relative to the standard normal distribution.

Sample Sizes	Statistic	Percentile 1.0	2.5	5.0	10.0	90.0	95.0	97.5	99.0
	$\rho = 0.0$	-2.47793	-2.04401	-1.65663	-1.26169	1.21699	1.56943	1.88405	2.21083
		0.18983	0.08561	0.07001	0.03697	0.03337	0.05835	0.09013	0.10437
$n_1 = 12$		-2.51330	-2.12918	-1.76050	-1.33724	1.30154	1.67252	1.97261	2.31863
	$\rho = 0.5$	-2.47185	-2.02304	-1.67395	-1.27995	1.26081	1.60107	1.88135	2.20171
$n_2 = 8$		0.16002	0.09909	0.07669	0.04621	0.02960	0.05763	0.06196	0.07787
		-2.47967	-2.07685	-1.74841	-1.31105	1.28799	1.62786	1.92752	2.25491
	$\rho = 0.0$	-2.35952	-1.93063	-1.60147	-1.22997	1.26985	1.61556	1.92377	2.29039
		0.10178	0.06129	0.04563	0.04812	0.05699	0.04981	0.07599	0.15160
$n_1 = 10$		-2.50293	-2.02146	-1.69927	-1.31498	1.33515	1.71636	2.05604	2.44271
	$\rho = 0.5$	-2.34285	-1.94598	-1.62956	-1.26059	1.29717	1.64754	1.92090	2.26443
$n_2 = 10$		0.09564	0.06987	0.04308	0.04075	0.03584	0.04767	0.05543	0.12452
		-2.41630	-1.99161	-1.68366	-1.30461	1.30346	1.67091	2.00238	2.37359
	$\rho = 0.0$	-2.50277	-2.03118	-1.66367	-1.29707	1.19701	1.53264	1.79890	2.14562
		0.13065	0.11485	0.06017	0.08527	0.03916	0.05415	0.07858	0.08839
$n_1 = 20$		-2.65027	-2.15996	-1.78500	-1.38977	1.25210	1.60488	1.90105	2.21067
	$\rho = 0.5$	-2.42070	-2.03980	-1.70723	-1.29879	1.24071	1.56704	1.84351	2.16896
$n_2 = 10$		0.11358	0.10287	0.06896	0.05151	0.04117	0.06465	0.06367	0.11992
		-2.56650	-2.11259	-1.74648	-1.35019	1.26425	1.58698	1.86264	2.17378
	$\rho = 0.0$	-2.30042	-1.91322	-1.53374	-1.24387	1.25262	1.60324	1.91269	2.22943
		0.10601	0.10698	0.08482	0.06911	0.03704	0.05249	0.07090	0.09878
$n_1 = 15$		-2.35126	-2.00278	-1.69636	-1.31090	1.33568	1.69243	2.02264	2.41922
	$\rho = 0.5$	-2.31254	-1.93551	-1.61692	-1.25181	1.25431	1.63236	1.93457	2.25738
$n_2 = 15$		0.08968	0.08424	0.06244	0.06139	0.04570	0.05464	0.05296	0.08661
		-2.30402	-1.97767	-1.67813	-1.29308	1.33123	1.65777	1.99776	2.40626
	$\rho = 0.0$	-2.54331	-2.10301	-1.76155	-1.36167	1.15087	1.47056	1.74159	2.04886
		0.09702	0.06240	0.04888	0.04430	0.04699	0.06001	0.08824	0.07082
$n_1 = 30$		-2.71383	-2.26302	-1.86713	-1.43879	1.25242	1.60142	1.88517	2.18916
	$\rho = 0.5$	-2.49579	-2.10012	-1.75397	-1.34979	1.19769	1.51860	1.78180	2.06157
		0.07561	0.07121	0.05283	0.03929	0.04431	0.03970	0.06579	0.08775
$n_2 = 10$		-2.70015	-2.20357	-1.82237	-1.39988	1.24553	1.59625	1.86554	2.14904
	$\rho = 1.0$	-2.53811	-2.09399	-1.72872	-1.33871	1.21590	1.53126	1.77164	2.03430
		0.11376	0.06085	0.05789	0.05322	0.03670	0.03627	0.05498	0.07587
		-2.64248	-2.20581	-1.80499	-1.37793	1.25907	1.59535	1.84635	2.10863
Standard Normal Variate		-2.3263	-1.9600	-1.6449	-1.2816	1.2816	1.6449	1.9600	2.3263

The results of our study are presented in the table on page 117. Each box consists of three items. The first and second items are, respectively, the mean and standard deviation of the bootstrap percentile point in question over the ten trials, while the third item is the corresponding true percentile point.

The results of our study indicate that, in summary, over the ten trials, the bootstrap distribution of all three SWL statistics closely approximated their true null distributions. However, for all three tests, as the percentile increased from 90.0 to 99.0 or decreased from 10.0 to 1.0, the standard deviation of the associated bootstrap value tended to increase. Thus, while both ultra-extreme and moderately extreme bootstrap percentile points appear to be approximately unbiased for the corresponding true percentile points, the former estimators are less reliable than the latter ones.

Our investigation also revealed that when the sample sizes were equal, the normal and bootstrap distribution did equally well in approximating the true null distribution. When the samples were of unequal size, the bootstrap and normal upper-tail percentiles were similar with the exception of the log-rank test, where we observed the normal distribution to perform slightly better. On the other hand, for the case of unequal sample sizes with lower-tail percentiles, the bootstrap distribution almost always performed better than the normal distribution, regardless of the test involved.

This particular Monte Carlo study did not take into account such factors as intensity of censoring, equality status of censoring d.f.'s, and type of censoring mechanism present. It would be well worthwhile for some researcher to assess the goodness-of-fit of the bootstrap distribution, taking into consideration these parameters along with those of our study.

5.3.2 Small-Sample Power.

In this subsection, we present the results of several significant Monte Carlo studies which examined small-sample power properties of SWL statistics. The researchers behind these investigations include the references listed on p. 103 as well as Lee et al. (1975), Fleming et al. (1980), Fleming and Harrington (1981), Fleming et al. (1987), and Pepe and Fleming (1989). Two major objectives of these studies were:

(1) determine the effect of each of the parameters listed on p. 103 singularly as well as in combination--on power;

(2) compare the power of three specific tests against six important alternatives, under specific censoring conditions and sample size configurations.

As in the subsection concerned with goodness-of-fit of the standard normal distribution, we describe the design and major outcomes of the above investigations from a general perspective rather than analyzing each study separately.

For objectives (1) and (2), the WL statistics utilized were the same as those involved in the simulations of 5.3.1.1 (a). The variance estimators considered for objective (1) were identical to those utilized in the aforementioned simulations, while the variance estimator employed for objective (2) was the conditional permutation variance. The alternatives employed belong to the class of stochastic ordering alternatives ($\bar{F}_1 \leq \bar{F}_2$) and are as follows:

I. large early difference in survival curves (functions) with no crossing of hazard functions.

II. large early difference in survival curves with crossing of hazard functions.

III. a contiguous proportional hazards alternative.

IV. large middle difference in survival curves with crossing of hazard functions.

V. large late difference in survival functions with no crossing of hazards.

VI. a contiguous time-transformed location alternative for the logistic distribution.

With respect to objective (1), the censoring mechanisms considered were censoring schemes (a)-(g) of subsection 5.3.1.1 (a). For objective (2), censoring schemes (a), (b), (d), and (f) were separately utilized in combination with various configurations of equal sample sizes.

For objectives (1) and (2), each configuration of censoring mechanism, sample sizes, and alternative hypothesis was replicated a fixed number of times, say N, thus generating N i.i.d values of the SWL statistic in question. Power was thus approximated by the proportion of replications where the alternative hypothesis in question was declared to be true. For a given significance level (any one of 0.01, 0.02, ..., 0.1), the associated critical value was obtained from the standard normal distribution. Note that since all six alternatives are such that $\bar{F}_1(t) \leq \bar{F}_2(t)$, and since all weight functions in question are nonnegative, we reject H_0 for large *positive* values of the test statistic.

We now discuss the results of these Monte Carlo simulations and begin with the outcomes related to objective (1).

The results indicate that if the sample sizes are equal and the censoring mechanisms of the two samples are identical, all variance estimators for a given WL statistic yield similar powers against the alternative in question. Suppose, though, that the two censoring mechanisms are identical (censoring schemes (a), (b), (d), and (f)) but that the sample sizes are unequal. Then the results reveal that, when the larger sample has the longer lifetimes, an SWL statistic with conditional permutation variance generally has greater power than an SWL statistic with the same numerator but a permutation variance. On the other hand, when the larger sample has the shorter lifetimes, the converse is true. Suppose now that $n_1 = n_2$, but that one sample experiences censoring while the other does not (censoring schemes (c) and (e)). Then, we see the same power relationship between an SWL statistic with a conditional permutation variance and one with a permutation variance (both of which have the same numerator) as discussed above, treating the uncensored sample as the larger sample.

When the censoring patterns of the two samples are equal and $n_1 \neq n_2$, or when $n_1 = n_2$ and one sample is censored while the other is not, Prentice's test, with variance estimator $\mathcal{V}_0$, is generally more powerful than Prentice's test, with either of the two other variance estimators, if the larger or uncensored sample has the shorter lifetimes. On the other hand, if the larger or uncensored sample has the longer lifetimes, Prentice's test with variance estimator $\mathcal{V}_0$ generally yields the second best power amongst the three.

Consider now a scenario where the censoring mechanisms of the two samples are identical and where the censoring scheme, sample sizes, and censoring intensities are allowed to vary. Then, firstly, the power of any particular SWL statistic with specified variance estimator generally tends to be lower with censoring schemes (b), (d), and (f) than with (a), controlling of course for all other parameters. Secondly, if the size of both samples decreases and/or the censoring intensities increase, power decreases–controlling once again for all extraneous parameters. In cases where the censoring intensity is 90% and $n_1 = n_2 = 10$, the power is sometimes less than the desired error level and rarely above 20%.

We now present the outcomes relevant to objective (2). The log-rank and Prentice's test, with a consistent estimator of the null hypothesis asymptotic variance, are the asymptotically optimal SWL statistics against, respectively, alternatives III and VI (see pp. 60, 89). The Monte Carlo simulations reveal that, within the context of the three SWL statistics in question, such as well is the case in small samples for censoring schemes (b), (d), and (f). For alternative III under censoring scheme (a), however, the three tests perform equally well. No comparison of the tests was conducted for alternative VI under censoring scheme (a).

With respect to departure III under censoring scheme (b) or (d) or (f), the simulations indicate that Prentice's test is generally more powerful than Gehan's test, an observation which agrees with Leurgan's (1983) efficiency calculations. A comparison though between the log-rank and Gehan's test for alternative VI, under each of the above censoring schemes, was not considered.

Prior to discussing the results for alternatives I, II, IV, and V, we make some comments which will enable us to account for the observed outcomes. Consider some alternative hypothesis (not necessarily a stochastic ordering alternative) in which $\lambda_1(t) \gg \lambda_2(t)$ $\forall t$ in some interval $I = [t_0, t_2]$, and suppose $dN(t_1) = 1$ for some $t_1 \in I$. Let A denote the collection of events

$$\{dN(t_1) = 1, R_1(t_1) = r_1, R_2(t_1) = r_2\},$$

let $B = dN_1^*(t_1) - R_1(t_1)R^{-1}(t_1)\, dN(t_1)$, and let

$$C = \int_I \hat{Q}(t)(dN_1^*(t) - R_1(t)R^{-1}(t)\, dN(t)).$$

Now,

$$\Pr[dN_1^*(t_1) = a|A] = \begin{cases} p & \text{if } a = 1, \\ 1-p & \text{if } a = 0, \end{cases}$$

where $p = (r_1\lambda_1(t_1) + r_2\lambda_2(t_2))^{-1}\lambda_1(t_1)r_1$. Thus, conditional on A (assuming r_2 is *not* excessively greater than r_1), B is much more likely to be positive than it is negative. Now, recall that the class of WL statistics has the form

$$S_n = \int_0^\infty \hat{Q}(t)\left[dN_1^*(t) - \frac{R_1(t)}{R(t)}\, dN(t)\right].$$

Thus, to render S_n sensitive to the given difference in hazard functions on I, $\hat{Q}(t)$ should be much greater than zero on I. (We assume that H_0 is rejected for large positive values of the test statistic.) With $\hat{Q}(t)$ defined as such, and under the assumption that $R_1(t_0)$, $R_2(t_0)$ are sufficiently large and that $R_2(t_0)$ is not excessively greater than $R_1(t_0)$, C will tend to be much larger than zero. Similarly, if $\lambda_1(t) \ll \lambda_2(t)\ \forall t \in I$, $\hat{Q}(t)$ should be much less than zero on I.

Suppose, now, that H_A is such that $\lambda_1(t) = \lambda_2(t)\ \forall t \in I$, and that $dN(t_1) = 1$ for some $t_1 \in I$. Then, assuming $R_1(t_1) > 0$ and $R_2(t_1) > 0$,

$$\Pr[B > 0|A] = r_1(r_2 + r_1)^{-1}$$

is dependent strictly on the values of r_1 and r_2, and not on the magnitude nor on the sign of $F_1(t_1) - F_2(t_1)$. To enhance the power of S_n against H_A, therefore, S_n should be rendered completely insensitive to failures occurring in I; hence, $\hat{Q}(t)$ should equal zero $\forall t \in I$.

Finally, suppose $\lambda_1(t)$ is moderately larger than $\lambda_2(t)$ $\forall t$ in I, and that $dN(t_1) = 1$ for some $t_1 \in I$. Then, conditional on A (assuming $R_2(t_1)$ is not excessively greater than $R_1(t_1)$), B is slightly more likely to be positive than it is negative. It follows, therefore, that S_n should be mildly sensitive to failures occurring in I and, hence, that $\hat{Q}(t)$ should be moderately greater than zero on I. Analogously, if $\lambda_1(t)$ is moderately smaller than $\lambda_2(t)$ $\forall t$ in I, then $\hat{Q}(t)$ should be moderately less than zero on I.

We now describe the outcomes of the Monte Carlo simulations for alternatives I, II, IV, and V. For these particular alternatives, a comparison was made strictly between the log-rank and Prentice's test under censoring scheme (b).

Departure I is, more specifically, given by the alternative:

$$\begin{cases} \lambda_1(t) \gg \lambda_2(t) & \text{if } 0 \leq t \leq t_0, \\ \lambda_1(t) = \lambda_2(t) & \text{if } t > t_0, \end{cases}$$

for some $t_0 \in (0, \infty)$. For this particular alternative, both procedures have very good sensitivity. The Prentice test, however, does have somewhat better power. We can explain this observation as follows: The log-rank test places a large weight on failures occurring both before and after t_0. Hence, since a majority of the failures after t_0 will tend to originate from sample 2, the large positive value of

$$A = \int_0^{t_0} \hat{Q}(t)[dN_1^*(t) - R_1(t)R^{-1}(t)\,dN(t)]$$

will be nearly completely offset by the moderately large negative value of

$$B = \int_{t_0^+}^{\infty} \hat{Q}(t)[dN_1^*(t) - R_1(t)R^{-1}(t)\,dN(t)].$$

The Prentice test, on the other hand, is very sensitive to failures which occur in $(0, t_0]$ yet is weakly sensitive to failures occurring in $(t_0, +\infty)$, and so, in this case, A is slightly offset by B.

Departure II is, more specifically, given by:

$$\begin{cases} \lambda_1(t) \gg \lambda_2(t) & \text{if } 0 \le t \le t_0, \\ \lambda_2 \gg \lambda_1(t) & \text{if } t_0 < t \le t_1, \\ \lambda_2(t) = \lambda_1(t) & \text{if } t > t_1, \end{cases}$$

for some t_0, t_1 such that $t_0 < t_1$ and such that

$$\int_0^t \lambda_1(u)\, du > \int_0^t \lambda_2(u)\, du \quad \forall t \in [t_0, t_1].$$

In this scenario, Prentice's test performs better than the log-rank test; however, the power of both procedures here is reduced in comparison to the case of departure I. This reduction in power is due to the fact that, for both tests, the positive value of

$$A = \int_0^{t_0} \hat{Q}(t)[dN_1^*(t) - R_1(t)R^{-1}(t)\, dN(t)]$$

is somewhat offset by the negative value of

$$B = \int_{t_0^+}^{t_1} \hat{Q}(t)[dN_1^*(t) - R_1(t)R^{-1}(t)\, dN(t)].$$

The Prentice procedure, however, outperforms the log-rank test since A is more offset by B for the latter SWL statistic than for the former, and since the latter test places greater weight than the former one on failures occurring in $(t_1, +\infty)$.

Departure IV is, more specifically, given by:

$$\begin{cases} \lambda_1(t) = \lambda_2(t) & \text{if } 0 \le t \le t_0, \\ \lambda_1(t) \gg \lambda_2(t) & \text{if } t_0 < t \le t_1, \\ \lambda_2(t) \gg \lambda_1(t) & \text{if } t_1 < t \le t_2, \\ \lambda_1(t) = \lambda_2(t) & \text{if } t > t_2, \end{cases}$$

for some t_0, t_1, t_2 such that $t_0 < t_1 < t_2$ and such that

$$\int_0^t \lambda_1(u)\,du > \int_0^t \lambda_2(u)\,du \quad \forall t \in [t_1, t_2].$$

The two tests show the same relative sensitivities here to what they displayed for alternative II. The explanation for this outcome is analogous to the discussion of the previous paragraph.

Finally, departure V can be equivalently written as:

$$\begin{cases} \lambda_1(t) = \lambda_2(t) & \text{if } 0 \le t \le t_0, \\ \lambda_1(t) \gg \lambda_2(t) & \text{if } t_0 < t, \end{cases}$$

for some $t_0 \in (0, +\infty)$. Here, the log-rank procedure has very good power, while Prentice's test has unacceptably low sensitivity. This observation is not unexpected since

$$\int_0^{t_0} \hat{Q}(t)[dN_1^*(t) - R_1(t)R^{-1}(t)\,dN(t)],$$

for both tests, is likely to be close to zero (even though the log-rank test places greater weight than Prentice's procedure on failures occurring in $(0, t_0]$), yet

$$A = \int_{t_0^+}^{\infty} \hat{Q}(t)[dN_1^*(t) - R_1(t)R^{-1}(t)\,dN(t)],$$

for the log-rank procedure, is likely to be much greater than zero, while A for Prentice's procedure is likely to be only slightly greater than zero.

For many years, tests based on the log-rank and Prentice's statistic have been amongst the most frequently used censored data, nonparametric, two-sample procedures. Because of their method of formulation, however, these two tests are more sensitive to alternatives of ordered hazard functions than to alternatives of crossing hazards (both stochastic ordering and crossing of failure time d.f.'s). Consequently, there has been much research recently concerned with developing versatile nonparametric procedures that are sensitive to both the ordered hazards and crossings hazards departures. See, for example, Fleming et al. (1980), Fleming and Harrington (1981), Schumacher (1984), Breslow et al. (1984), Fleming et al. (1987), and Pepe and Fleming (1989). Indeed, these investigators show, via small-sample Monte Carlo simulations, that their suggested tests can outperform the two aforementioned two-sample tests under crossing hazards alternatives. Furthermore, these newly proposed tests compare favorably with the log-rank and Prentice's procedure under, respectively, departures III and VI.

REFERENCES

Altshuler, B. (1970), *Theory for the measurement of competing risks in animal experiments*, Math. Biosci. **6**, 1-11.

Andersen, P.K., Borgan, O., Gill, R.D. and Keiding N. (1982), *Linear non-parametric tests for comparison of counting processes with applications to censored survival data*, Int. Statist. Rev. **50**, 219-58.

Andersen, P.K. and Gill, R.D. (1982), *Cox's regression model for counting processes: a large sample study*, Ann. Statist. **10**, 1100–20.

Barnard, G.A. (1963), *Some aspects of the fiducial argument*, J. R. Statist. Soc. B **25**, 111 4.

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

Breslow, N.E. and Crowley, J. (1974), *A large sample study of the life table and product limit estimates under random censorship*, Ann. Statist. **2**, 437–53.

Breslow, N.E., Edler, L. and Berger, J. (1984), *A two-sample censored-data rank test for acceleration*, Biometrics **40**, 1049–62.

Brown, M.B. (1984), *On the choice of variance for the log rank test*, Biometrika **71**, 65 74.

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

Cox, D.R. (1975), *Partial Likelihood*, Biometrika **62**, 269–76.

Cox, D.R. and Oakes, D. (1984), *Analysis of Survival Data*, Chapman and Hall, London.

Crowley, J. (1974), *A note on some recent likelihoods leading to the log rank test*, Biometrika **61**, 533–8.

Cuzick, J. (1985), *Asymptotic properties of censored linear rank tests*, Ann. Statist. **13**, 133 41.

Efron, B. (1967), *The two sample problem with censored ata*, Proc. 5th Berkeley Symp. **4**, 831 53.

Efron, B. (1979), *Bootstrap Methods: Another Look at the Jackknife*, Ann. Statist. **7**, 1 26.

Fisher, R.A. and Yates, F. (1963), *Statistical Tables for Biological, Agricultural and Medical Research*, Longman, London.

Fleming, T.R. and Harrington, D.P. (1981), *A class of hypothesis tests for one and two samples of censored survival data*, Communications in Statistics, Part A Theory and Methods **10**, 763–94.

Fleming, T.R. and Harrington, D.P. (1991), *Counting Processes and Survival Analysis*, Wiley, New York.

Fleming, T.R., Harrington, D.P., and O'Sullivan, M. (1987), *Supremem versions of the log-rank and generalized Wilcoxon statistics*, Journal of the American Statistical Association **82**, 312 20.

Fleming, T.R., O'Fallon, J.R., O'Brien, P.D., and Harrington, D.P. (1980), *Modified Kolmogorov–Smirnov test procedures with application to arbitrarily right-censored data*, Biometrics **36**, 607–625.

Fraser, D.A.S. (1968), *The Structure of Inference*, Wiley, New York.

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

Gill, R.D. (1980), *Censoring and Stochastic Integrals. Mathematical Centre Tracts, no. 124*, Mathematische Centrum, Amsterdam.

Groggel, D.J., Schaefer, R.L., and Skillings, J.H. (1988), *The effect of withdrawal censoring on two-sample tests*, J. Statist. Comput. Simul. **28**, 275 302.

Gu, M.G., Lai, T.Z., and Lan, K.K.G. (1991), *Rank tests based on censored data and their sequential analogues*, American Journal of Mathematics and Management Science **11**, 147–176.

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

Harrington, D.P. and Fleming T.R. (1982), *A class of rank test procedures for censored survival data*, Biometrika **69**, 553–66.

Jennrich, R.I. (1983), *A note on the behavior of the log rank permutation test under unequal censoring*, Biometrika **70**, 133–7.

Johnson, R.A. and Mehrotra, K.G. (1972), *Locally most powerful rank test for the two-sample problem with censored data*, Ann. Math. Statist. **43**, 823 31.

Kalbfleisch, J.D. and Prentice, R.L. (1973), *Marginal likelihoods based on Cox's regression and life model*, Biometrika **60**, 267–78.

Kalbfleisch, J.D. and Prentice, R.L. (1980), *The Statistical Analysis of Failure Time Data*, Wiley, New York.

Kalbfleisch, J.D. and Sprott, D.A. (1970), *Application of likelihood methods to models involving as large number of parameters (with discussion)*, vol. 32.

Kaplan, E.L. and Meier, P. (1958), *Nonparametric estimation from incomplete observation.*, J. Am. Statist. Assoc. **53**, 457–81.

Lagakos, S.W. (1979), *General right censoring and its impact on the analysis of survival data*, Biometrics **35**, 139–156.

Latta, R.B. (1981), *A Monte Carlo study of some two-sample rank test with censored data*, Journal of the American Statistical Association **76**, 713 19.

Lee, E.T., Desu, M.M., and Gehan, E.A. (1975), *A Monte Carlo study of the power of some two-sample tests*, Biometrika **62**, 425 32.

Lehmann, E.L. (1953), *The power of rank tests*, Ann. Math. Statist. **24**, 23 43.

Leurgans, S. (1983), *The classes of censored data rank test: strengths and weaknesses under censoring*, Biometrika **70**, 651–58.

Leurgans, S. (1984), *Asymptotic behavior of two-sample rank tests in the presence of random censoring*, The Annals of Statistics **12**, 572 589.

Lustbader, E.D. (1980), *Time-dependent covariates in survival analysis*, Biometrika **67**, 697–8.

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

Mantel, N. (1967), *Ranking procedures for arbitrarily restricted observation*, Biometrics **23**, 65-78.

Mehrotra, K.C., Michalek, J.E., and Mihalko, D. (1982), *A relationship between two forms of linear rank procedures for censored data*, Biometrika **69**, 674–6.

Miller, R.C. (1981), *Survival Analysis*, Wiley, New York.

Oakes, D. (1981), *Survival times: aspects of partial likelihood (with discussion)*, Int. Statist. Rev. **49**, 199-233.

Pepe, M.S. and Fleming, T.R. (1981), *Weighted Kaplan–Meier statistics: a class of distance tests for censored survival data*, Biometrics **45**, 497–507.

Peto, R. (1972), *Rank tests of maximal power against Lehmann-type alternatives*, Biometrika **59**, 472–5.

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

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, 861–7.

Puri, M.L. and Sen, P.K. (1971), *Nonparametric Methods in Multivariate Analysis*, Wiley, New York.

Savage, I.R. (1956), *Contributions to the theory of rank order statists–the two-sample case*, Ann. Math. Statist. **27**, 590–615.

Schumacher, M. (1984), *Two-sample tests of Cramer–von Mises and Kolmogorov Smirnov type for randomly censored data*, International Statistical Review **52**, 263-81.

Tarone, R.E. and Ware, J. (1977), *On distribution–free tests for equality of survival distributions*, Biometrika **64**, 156-60.

Tsiatis, A.A. (1981a), *A large sample study of Cox's regression model*, Annals of Statistics **9**, 93-108.

Tsiatis, A.A. (1981b), *The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time*, Biometrika **68**, 311-15.

Tsiatis, A.A. (1982), *Repeated significance testing for a general class of statistics used in censored survival analysis*, J. Amer. Statist. Assoc. **77**, 855–861.

van der Waerden, B.L. (1953), *Ein neuer Test für das Problem der zwei Stichproben*, Math. Annalen **126**, 93–107.

Wilcoxon, F. (1945), *Individual comparisons by ranking methods*, Biometrics **1**, 80-3.