A PRELIMINARY TEST ESTIMATOR (PTE) FOR P(Y>X) CONDITIONAL ON A RANK TEST OF PROPORTIONAL HAZARDS IN THE UNCENSORED TWO-SAMPLE PROBLEM

by

David Hoberman

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1805T

September 1986

A PRELIMINARY TEST ESTIMATOR (PTE) FOR P(Y>X) CONDITIONAL ON A RANK TEST OF PROPORTIONAL HAZARDS IN THE UNCENSORED TWO-SAMPLE PROBLEM

by

David Hoberman

A Dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the School of Public Health.

Chapel Hill

1986

Approved by:

Adviser

Reader

Reader

ABSTRACT

DAVID HOBERMAN. A Preliminary Test Estimator (PTE) for P(Y>X) Conditional on a Rank Test of Proportional Hazards in the Uncensored Two-Sample Problem

The probability that a random variable in one group is greater than that in another group is a parameter of natural interest in survival analysis. Two choices for estimators are the U-statistic (Wilcoxon statistic), which is unbiased regardless of the relationship between the two distribution functions, and the maximum likelihood estimator derived from the probability of the joint rank vector. The latter has a simple form when proportional hazards holds and is a function of the unknown proportionality constant. Since the rank maximum likelihood estimator has less variance than the U-statistic under proportional hazards, we propose an adaptive rank test to detect departures from proportional hazards so that a choice of the two estimators can be made using the data itself.

The proposed test is a linear rank statistic whose null distribution is asymptotically normal. Various score functions are proposed and their consistency for selected alternatives to proportional hazards is examined. In addition we derive non-centrality parameters under selected local alternatives.

Finally, the properties of the preliminary test estimator are computed using numerical integration. As in previous literature, we use bias and mean square error as criteria for the performance of the PTE under local alternatives. We find that the PTE's performance depends mostly upon the correlations of the test statistic with the two estimators. The non-centrality parameter seems to have less influence.

ACKNOWLEDGEMENTS

I would first like to thank my dissertation advisor, P.K. Sen, for our long discussions, his guidance and extraordinary patience.

Also, I would like to thank the other members of my committee: Lawrence Kupper, C.E. Davis, Kant Bangdawala and Carl Shy.

On the less technical side, I want to express my appreciation to Ms. Jean Coble for her effective moral support. Also, I would like to thank Ms. Betty Owens for typing the manuscript under uncommon duress.

This research was supported by a grant from the National Institute for Environmental Health Sciences.

TABLE OF CONTENTS

CHAPTER I

An important problem in biostatistics is the comparison of two populations when the outcome of interest is time to event. Survival analysis, for example, often attempts to compare the effects of two different "treatments" by comparing the sample survivorship distributions of two independent groups. Usually, the investigator formulates the null hypothesis that the two distributions are identical versus the alternative hypothesis that one of the groups' "survival curves" is "above" the other, inidcating better survival experience in the former group. Since this view of the problem is equivalent to comparing the empirical distribution function of the two groups, one can alternatively pose the same null hypothesis and restate the alternative hypothesis to say that the time to event random variable in one population is stochastically larger than that in the other.

Finally, by forming the convolution of one distribution function with the survival distribution of the other, one can derive the parameter: the probability that the time to event in one population is larger than that in the other; i.e., $P(X < Y) = \int_0^1 G dF = \int_0^1 (1 - F) dG$ when X has cdf G and Y has c.d.f. F. Therefore, the estimation of $P(X < Y)$ can be an important statistical problem. However, as in many problems, the imposition of some mathematical relationship between the two populations wil suggest the use of one estimator over another. The structure in this paper will be proportional hazards (PH), so that $[1-G(Y)]^k = 1-F(Y)$, $k > 0$. One can easily show that $P(X < Y) = 1/(k+1)$, and so the problem reduces to that of estimating k, the unknown constant of proportionality under PH. When PH is not the case, it is clear that another estimator may be more appropriate, one that is optimal in some sense, but that does not depend upon any relation between the F and G.

These considerations lead to the notion of Preliminary Test Inference (PTI). In general, PTI is a procedure which uses a (preliminary) test of some hypothesis; and conditionally upon the result of that test, the investigator chooses among available procedures to answer the primary question of interest. In the present context, we may have several estimators for $P(X < Y)$ and must choose among them conditionally on test for PH. Since the procedure is inferential, a loss is incurred depending upon the Type I and Type II error of the test.

In order to make these ideas concrete, we introduce the notation of Ω-spaces. It will be shown in Chpater II that if F and G are known and PH holds, there is a monotone transformation $X \longrightarrow X^*$ where $X^* \sim \exp(1)$ and $Y \longrightarrow Y^*$, where $Y^* \sim \exp(k)$, where k is the PH constant of proportionality. Let

$$\Omega = \{F, G \mid F \text{ and } G \text{ arbitrary}\}$$

$$\Omega PH = \{F, G \mid [1-G]^k = 1-F ,\ k > 0\}$$

$$\Omega^* = \{F^*, G^* \mid G \sim \exp(\lambda),\ F^* \text{ arbitrary}\}$$

$$\Omega^* PH = \{F^*, G^* \mid G^* \sim \exp(k),\ F^* \exp(k\lambda) ,\ k > 0\}$$

We would like to take advantage of the transformation in order to reduce the problem from one which discriminates between PH and non-PH to one which discriminates between jointly independent exponentials and alternatively at least one distribution not being exponential. Therefore, we want the estimation and testing procedures to be invariant under the monotone transformation. Invariance is the optimal condition for estimation and testing in this problem because (1) $P(X^* < Y^*) = P(X < Y)$ and (2) F and G are never known. That is, if F or G were known, we could actually apply the known transformation to the data of the other group and reduce the original two-sample problem to the one-sample problem of testing for exponentiality of the transformed data (see Chapter II). Invariance makes knowledge of the actual transformation unnecessary. Summarizing, we can say that the estimators of $P(X < Y)$ will be invariant under monotone transformations and the testing procedure will also be invariant under a monotone transformation resulting in the induced map

$$[H_o : F,G \varepsilon \Omega_{PH}] \longrightarrow [H_o^*: F^*,G^* \varepsilon \Omega^*_{PH}]$$

The invariant test of choice will be a rank test for three reasons; (1) the rank vector of the joint sample of X's and Y's is invariant under all monotone transformations, (2) the rank vector is the maximal invariant statistic. That is, the rank vector is that statistic which extracts the maximum information about the distribution of X's and Y's, and yet is independent of the actual order statstics and the parameter(s) of the distribution. Consequently, if H_o is accepted, the estimator of choice would be a rank MLE computed under H_o, and (3) we want a test which makes no assumption about distribution functions of the two populations.

If H_o is rejected we use the modified form of the two-sample Mann-Whitney U-statistics

$$U = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (Y_j - X_i), \quad \text{where}$$

$$(Y_j - X_i) = \begin{cases} 1 & \text{if } X_j > X: \\ 0 & \text{otherwise} \end{cases}$$

is the kernal of $P(X < Y)$. Detailed justification for these two estimators if deferred to Chapters II and III.

Finally, since there is evidently no single UMP invariant test for PH, we may partition the space of alternative hypotheses into local and global alternatives. Thus the preliminary test may be optimal (locally MP) for local alternatives while consistent for a broad class of global alternatives.

LITERATURE REVIEW

A. $R = P(X < Y)$ when $X \sim \exp(\lambda)$, $Y \sim \exp(\mu)$

Generally speaking, properties of various estimators of R have been extensively studied only in the past ten to fifteen years. Since the estimation of R under PH is closely related to the problem of estimating R when both distributions are exponential, it is worth reviewing the most relevant literature.

Tong (1974), with a correction by Johnson (1975), derived a closed expression for the UMVUE of R when neither λ nor μ is known. Since it is very cumbersome to use in practice, one is led to consider the MLE, $\bar{Y}/(\bar{X} + \bar{Y})$ as a suitable alternative. Kelly et al., (1976) assumed μ known and computed, using numerical methods, the bias and mean square efficiency of the MLE relative to the UMVUE for different sample sizes and different values of R. They found that the UMVUE is superior to the MLE for $R \simeq 1$, but for $R < .5$, "the advantage is seen to be with the MLE even though it is biased." Even for $R \simeq 1$, "for large n, the MLE might be preferred since the computational difficulty probably outweighs the slightly greater efficiency." Sathe and Shah (1981) derived bounds for the mean and mean square error of the MLE and the UMVUE under the two conditions, known and unknown." In the case when μ is unknown, bounds for the first two moments of the MLE are obtained and also a lower bound based on the Bhattaharja bound is obtained for the variance of the UMVUE." The authors then computed the bias of the MLE and the upper bound for the ratio of the M.S.E. of the MLE to the variance of the UMVUE for equal sample sizes n, ranging from 5 to 100, and for λ / μ ranging from 0.01 to 0.9. For n=5, the largest exact absolute bias was 1.9×10^{-2}; for n = 10, the greatest absolute bound was 0.9×10^{-2}; for n = 25, the greated absolute bound was 0.4×10^{-2} and for n = 50, the greatest absolute bound was 0.2×10^{-2}. In the case of the ratio of the MSE of the MLE to the variance of the UMVUE, the ratio is less than or equal to one in every case except $\lambda/\mu = .01, 1.0$ and n = 5, 10. We conclude that based upon the work of Kelly et al. (1976) and Sathe and Shah (1981), the computational difficulties and complicated expressions associated with the UMVUE militate against its use in the light of the very good performance of the MLE over a relatively broad range of ratios of scale parameters and realistic sample sizes.

B. A U-Statistic as Estimator of $R = P(X < Y)$

Since there is no UMVUE of R for all specified F and G, we are led to consider the modified the Mann-Whitney U-statistic cited in the introduction when we reject the null hypothesis of PH. $\hat{R} = \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} \psi (Y_j - X_i)$ is unbiased (being a U-statistic) and has variance (m+n+1)/12mn. Lehmann (1950) showed that $\hat{R}$ is the UMVUE for arbitrary continuous F and G and later that $\sqrt{n}(\hat{R} - R)$ is asymptotically normal for any F, G under the assumption that $m = cn$, $n \to \infty$ and that R is not too near zero or one (Lehmann (1951)). Thus if either or both of the conditions are not satisfied, other estimates of $Var(\hat{R})$ should be used. Van Dantzig (1951) derived a sharp upper bound for $Var(\hat{R}) \leq R(1-R)/\min(m,n)$. Besides demonstrating the consistency of $\hat{R}$ for $P(X < Y)$ though Chebeshev's inequality, this upper bound has proved useful for constructing large-sample confidence intervals for $P(X < Y)$. (Govindarajulu (1968)). In the same paper, Govindarajulu further demonstrated the asymptotic normality for $\hat{R}$ when normalized by $r = \min(m,n)$ for any continuous F and G without the assumption that $m = cn$, $n \to \infty$ and derived a distribution free, unbiased and consistent estimator of $Var(\hat{R})$.

The first attempt to construct confidence intervals for $R = P(X < Y)$ which did not rely on large sample normality was that of Birnbaum (1956). For F and G unknown, using the convolution of the two empirical distribution functions (which equals $\hat{R}$) and a set inclusion argument, he obtained a lower bound for $Pr(R-R < \varepsilon)$. Noting that the inequality $R-\hat{R} < \text{Sup}\{F_m(s) - F(s)\} + \text{Sup}\{G(s) - G_n(s)\} = D_m^+ + D_n^+$ is crude, he acknowledged the large sample sizes required to achieve samll confidence intervals with high confidence coefficients. Thus for small sample sizes, the confidence intervals will be large.

Birnbaum and McCarty (1958) updated the procedure by obtaining the convolution of D_m^+ and D_n^+. They note that this procedure is still conservative. Govindarajulu (1968) used Van Dantzig's upper bound and the asymptotic normality of R to improve upon the confidence intervals of Birnbaum. Ury (1972) used Van Dantzig's upper bound and Chebeshev's inequality to improve upon Birnbaum intervals in the case where $m = n$.

C. R under PH

We have noted in the introduction that if PH holds, then $P(X < Y) = 1/(1+k)$ where k is the (unknown) constant of proportionality. There is little literature on the estimation of k. Abel (1982) has derived a "rather sharp upper bound for the variance of the Mann-Whitney estimator $\hat{R}$. However, we would anticipate a better estimator which used the actual probability of the joint rank vector. Pathak, Zimmer and Williams (1979) have derived a non-parametric estimator of k under PH. They first show that all the information on R is contained in ranks of one of the samples in the joint rank vector and also that no unbiased estimator of k exists. They then provide a closed form expression for k and a simple expression for its bias. However, the estimator does not use all the information in the joint rank vector if the sample sizes are unequal and the production of $\hat{k}$ requires roughtly 2^n calculations, where n is one of the sample sizes. Thus they resort to a randomization procedure which produces an estimate which converges strongly to k as the number of randomizations is increased. For further details on the estimation of an acceleration parameter, see McRae (1971) and Stech et al. (1974).

D. Tests for PH

For completeness, we recall that if F, say, were known, the problem of testing for proportional hazards would reduce to testing the exponentiality of the other sample's transformed empirical distribution. This is the one-sample problem and has a extensive background in reliability theory. Epstein (1960) published twelve tests for exponentiality, several of which were tailored to detect specific departures from exponentiality. Doksum and Yandell (1984) have written a review paper and bibliography on the one-sample case. They concentrate on tests based upon spacings, D_i, where $D_i = (n+1-i)\ (T_{(i)} - T_{(i-1)})$ and $T_1 < T_{(2)} \cdots < T_{(n)}$ are the order statistics from a sample of size n.

Schoenfeld (1980) constructed an omnibus goodness-of-fit tests using a sub-division of the time axis and then calculating observed vs expected number of events in each interval. Anderson (1982) proposed a test to see whether proportional hazards is true when the $(k+1)^{st}$ covariate is added to a model which already included k covariates in the exponential part of the Cox hazard function. By approximating the underlying hazard rate $\lambda(t)$ with a step function on pre-specified intervals of the time axis, Anderson constructs a statistic which is chi-square under the proportional hazards hypothesis. Nagelkerke et al (1984) proposed a test based on the autocovariance of the successive contributions to the derviative of the log likelihood.

E. Preliminary Test Inference

Often in statistical analysis, the investigator uses tests of significance in order to infer salient features of the statsitical structure of the data. Conditional upon these tests, he or she proceeds with either a testing or estimationprocedure, or both. This use of statisticalinference before proceeding to the answer to the primary question of interest is called Preliminary Test Inference and has been rigorously studied only since the 1940's. The first published formal investigation was that of Bancroft (1944) in which he studied two specific problems: (I) Testing for the homogeneity of two independent estimates of σ_1^2 and σ_2^2, and conditional on that test, estimating σ_1^2 where $\sigma_1^2 < \sigma_2^2$. (II) Given a regression model $E(Y) = b_1 X_1 + b_2 X_2$, use a test of significance for Ho: $b_2 = 0$ in order to choose between the two parameter model and $E(Y) = b_1 X_1$. In both cases, the bias of the preliminary test estimator is computed under three parameters: the sample sizes, a measure of the underlying departure from the preliminary test null hypothesis and different critical values for rejection of the null hypothesis. The distinguishing feature of these investigations was that exact results were obtainable by analytic methods. Their basic strategy was the following: find sufficient statistics for the test statistics and the quantity to be estimated. If need be, transform the problem so that some functions of the test statsitic and the quantity to be estimated are independent. Form the joint density function and then use a one-to-one transformation in order to derive the exact joint distribution of the test statistic and quantity to be estimated. In both problems studied by Bancroft, closed form solutions to the bias we derived in terms of the incomplete beta function.

Subsequent studies have included Mosteller (1948) in which he treated the problem of pooling means from two independent populations on the basis of unequal weights.

The regression problem has been studied extensively by Ashar (1970), Larson and Bancroft (1963a, 1963b), Chipman and Rao (1964), Bancroft (1964). Recently, pooling of data in contexts other than regression has been studied by Huntsberger (1955). Bancroft and Han (1980) have provided an overall view and bibliography of preliminary testing in ANOVA models. Of major concern in any preliminary test estimator scheme is the control of Type I error. The authors point out that "such investigations should include respective recommendations regarding the significance level of the preliminary test based on an acceptable criterion as regards the final inference." In their own investigation of a random effect ANOVA model, they found that conducting a test of homogeneity of variance at the $\alpha = .05$ level is unacceptable because it will admit pooling when not warranted and thus raise Type I error for the test of homogeneity of group means ("Testitesting"). Thus they recommend increasing the size of the preliminary test of $H_o: \sigma_1^2 = \sigma_2^2$. They also point out that use of mean square error as a criterion for the comparison of performance of estimators is not always feasible, and instead, they focue on a method for maximizing the relative efficiency of the preliminary test estimator relative to a competitor. Sahek and Sen provide an introduction to non-pametric preliminary test inference for a vector of parameter and simple unvariate and multivariate regression models. Summarizing, it can be stated the use of preliminary test estimation has been largely restricted to the pratical data analysis situation of pooling data, either in the serivce of estimating a population parameter, or model building in the content of the general linear model. In addition, the effect of a preliminary test of a general linear hypothesis in order to decide on the validity of restricted estimators has been studied recently by Brooks (1976), Bock, Yancy and Judge (1973), and Hill, Judge and Fomby (1978). The typical measures of loss under a specific estimator are bias and mean square error.

In Chapter II of this study, we concentrate upon the properties of the two estimators of $(P(Y>X)$ the rank MLE and the U-statistic. The rank MLE is shown to be the unique solution to the MLE equation and the asymptotic variance of $\sqrt{N}(\hat{k}_{NR}-k)$ is derived. Similarly the asymptotic variance of $\sqrt{N}\,(\hat{R}_{NW}-R)$, the relevant statistic based upon the Wilcoxon U-statistic is derived. Finally it is shown that when proportional hazards holds then the rank MLE has less variance than the U-statistic uniformly in k at least for $1< k < 3$.

Chapter III presents the theoretical work on the test statistic for proportional hazards under H_0. A rigorous statement of the null hypothesis is alternatives is stated. The basis of the derivation of the probability distribution of the test statistic under H_0 is the Chernoff-Savage theorem. The asymptotic normality of the test statistic is shown using the Chernoff-Savage theorem for allowable scores and by adapting a Chernoff-Savage approach to form and expansion of the rank MLE $\sqrt{N}(\hat{k}_{NR}-k)$. This expansion is then used to compute required covariances which are needed to derive the asymptotic variance of the final test statistic under H_o. The resulting statistic is an "adaptive" statistic in the sense that $\hat{k}_{NR}$ replaces k in the chosen score. It is argued that trying to construct a statistic along the lines of a Neyman C_α statistic is not really relevant and is inordinately complicated. In other words, the notion of an asymptotically locally most powerful test is not appropriate to this problem.

Chapter III also addresses the question of consistency of the statistic. Consistency properties obviously depend upon the estimate of k through the specified alternative. Essentially two classes of scores are proposed. 1) a class loosely based upon a locally most powerful score statistic and 2) the Savage score and Wilcoxon score. It is easier to demonstrate consistency for the latter scores than the former.

Finally, the properties of the preliminary test estimator for $R = P(Y>X)$ are derived. Among other things, this work requires the expansion of the U-statistic $\sqrt{N}(\hat{R}_{NW}-R)$ using the generalized U-statistic theorem and then expressions for its covariance with the test statistic.

Chapter IV presents theoretical results deriving the properties of the PTE under local alternatives to proportional hazards. The main task for this work is the derivation of the non-centrality parameter of the test statistic under the prescribed local alternative. The complicated form of the statistic due to the estimation of a nuisance parameter argues for the use of LeCam's Third Lemma rather than the classical (Taylor series) approach used for the computation of a Pittman efficiency.

Chapter V presents a discussion of the performance of the PTE using numerical analysis for the previously derived expressions. Results are presented for the cases under which H_o is true and when local alternatives exist in the direction of a linearly increasing hazard ratio and a hazard ratio which increases more slowly in the Makeham direction. All tests are one-sided with Type I error levels of .05, .10 and .20. Performance criteria are PTE bias and mean sequared error.

Chapter VI then presents topics for further research.

CHAPTER II

INTRODUCTION

Let us imagine two independent groups which are homogeneous with respect to all variables related to survival except for the "Treatment" of interest. In classical problems the question of interest is usually whether there is a "treatment difference" between the two groups, i.e., whether they have differential survival experience. It is reasonable to suggest at least two other questions relating to the groups' relative survivorship: a) given that we already know that one group will be different, we might be interested in some measure of the difference. b) Regardless of our prior knowledge, which estimators of a relevant parameter should one use? In both cases the parameter of interest in this study is $R = P(Y > X)$. In Chapter I, two choices were mentioned: the Wilcoxon estimator which is based solely on the rank vector itself, and the estimator derived from the probability of the rank vector under the proportioned hazards assumption. The following lemma gives a connection between the hazard ratio and the information contained in the rank vector.

Lemma: Under any monotone transformation of the random variables of two groups, the hazard ratio profile is the same over the real line. That is, even though the time <u>scale</u> is not invariant, the relative hazard ratio profile <u>is</u> invariant.

Proof: Let Y_i have cdf $F(x)$ and X have cdf $G(x)$. Let $H(x_j) = x_j^*$ where H is continuous, monotone and at least once differentible. Similarly, let $T(Y_i) = Y_i^*$.

Then $F(x) = Pr(Y \leq x) = Pr(H(Y) \leq H(x)) = Pr(Y^* \leq H(x)) = F^*H(x)$

thus $f(x) = f^*(H(x))\ H'(x)$

Similarly $G(x) = G^*H(x)$

$$g(x) = g^*(H(x))\ H'(x)$$

Thus,

$$(1)\quad \left[\frac{f(x)}{1-F(x)}\right] / \left[\frac{g(x)}{1-G(x)}\right] = \left[\frac{f^*H(x)H'(x)}{1-F^*H(x)}\right] / \left[\frac{g^*(H(x)\ H'(x)}{1-G^*H(x)}\right]$$

$$= \left[\frac{f^*(x^*)}{1-F^*(x^*)}\right] / \left[\frac{g^*(x^*)}{1-G^*(x^*)}\right]$$

where $x^* = H(x)$. QED.

So far we shown that the rank vector provides information which leads to a choice of two rank estimators of P(Y>X) and a relationship which implies that if the hazard ratio is constant over time in the original random variable space X,Y, then the hazard ratio is constant in the transformed space of X^*, Y^*. What if the foregoing condition were true? Then if a rank estimator were derived under the constant hazard ratio assumption, it would likely be "better" in some way than an estimator derived under more general conditions. On the other hand, if proposed hazards were not true, the latter estimator might be better.

Since the state of nature is unknown, we are forced to make a statistical test to see whether the proportional hazards assumption is a reasonable one. If we accept the null hypothesis, then we choose the estimator based upon the probability of the ranks, the RMLE $\hat{R}_{NR}$. If we reject, then we choose the Wilcoxon estimator, $\hat{R}_{NW}$ We do this anticipating that, on the average, it is better to use a statistical test to choose one or the other, rather than using one exclusively and thus risking loosing information that can increase MSE either through increase in variance of an unbiased estimator or an increase in bias of a biased estimator.

In this Chapter we make use of the lemma above to derive the probability of the rank vector under proportional hazards. Then we study the asumptotic properties of $\sqrt{N}(\hat{R}_{NR}-R)$ and $\sqrt{N}(\hat{R}_{NW}-R)$ and determine how much information is gained by using $\hat{R}_{NR}$ if it is known that proportional hazards is true. We then discuss what can go wrong if $\hat{R}_{NR}$ is used when proportionality does not hold.

RANK LIKELIHOOD UNDER PROPORTIONAL HAZARDS

Referring to equation (1) we see that if $g^*(x^*)/1-G^*(x^*)$ were equal to one, then possible hazard ratios in the original space can be "indexed" by various F^*'s in the transformed spaces. It follows immediately from the lemma that if the hazard ratio in the original space is k, then if $x_j \longrightarrow x_j^* \sim \exp(1)$, then $Y_i \longrightarrow Y_i^* \sim \exp(k)$. In fact, the transformation $X_j^* = -\log [1-G(x_j)]$ since $G(x_j) = U_j \sim U(0,1)$ so

$$\Pr(X_j^* < x^*) = \Pr [-\log U_j < x^*] = \Pr [U_j < e^{-x^*}]$$
$$= e^{-x^*}.$$

Now let $Y_i \longrightarrow Y_i^*$ $i = 1,\ldots,n_1$ $F(y) = 1-e^{-ky}$

$X_j \longrightarrow X_j^*$ $j = 1,\ldots,n_2$ $G(x) = 1-e^{-x}$

Further let $z_{n\ell} = \begin{cases} \ell & \text{if the } \ell^{th} \text{ order statistic is from the Y-sample} \\ 0 & \text{otherwise.} \end{cases}$

Also let $z_{n\ell}^* = z_{n\ell} + \ldots + z_{nn}$ $1 < \ell < N$ where $N = n_1 + n_2$. Thus $z_{n\ell}^*$ is the number of observations equal to or greater that the ℓ^{th} order statstics of the combined sample of N observations. If $\underline{R}$ is the rank vector and z_n is the z- vector just defined, then they contain exactly the same information and thus we can write:

$$P(\underset{\sim}{R}) = P(z_n) = k^{n_1} \int \dots \int \prod_{\ell=1}^{N} \exp\{-\{t_\ell(1-z_{n\ell}) + kz_{n\ell}]\} \, dT_1 \dots dt_N$$

$$= \frac{k^{n_1}}{\prod_{\ell=1}^{N} \{(N-\ell+1) + (k-1)z^*_{N\ell}\}}$$

RANK MLE UNDER H_o

$$\log P(\underset{\sim}{R}) = n_1 \log k - \sum_{\ell=1}^{N} \log \{N-\ell+1) + (k-1)\, z_{n\ell}{*})$$

$$(1) \quad \frac{\partial \log P(\underset{\sim}{R})}{\partial k} = \frac{n_1}{k} - \sum_{\ell=1} \frac{z_{N\ell}{*}}{(N-\ell+1 \; +(k-1)z_{N\ell}{*}} .$$

Thus $\hat{k}$ is the zero of (1).

In order to show that the zero of (1) us unique, we rewrite (1) as

$$(2) \quad k \,\{\frac{\partial}{\{\partial k} (\log P(\underset{\sim}{R})) = \sum_{\ell=1} [z_{N\ell} - \frac{Kz_{N\ell}{*}}{N-\ell+1-z_{N\ell}{*}) + kz_{N\ell}{*}}]$$

Thus the zero of the RHS of (2) is the zero of the RHS of (1). Since $Kb/(a+b)=k/((a/b)+k)$ is increasing in k, ($z_{N\ell}{*} < N - \ell + 1 \;\forall \ell$), then $k \frac{\partial}{\partial k} \log P(\underset{\sim}{R})$ is decreasing in k. As $k \longrightarrow 0$, then the RHS of (2) converges to $n_2 > 0$ and when $k \longrightarrow \infty$, the RHS of (2) converges to $n_2 - N < 0$. Then $k \frac{\partial \log P(\underset{\sim}{R})}{\partial k}$ crosses the line $y = 0$ at most once and so $\hat{k}_{NR}$ is unique.

In the sequel, we will be concerned with the distribution of $(\hat{k}_{NR}-k^*)$ where k^* is a constant. The consistency of $\hat{k}_{NR}$ for k^* when $k^* = k$ (i.,e., H_o is true) and the asymptotic normality of $(\hat{k}_{NR}-k)$ has been demonstrated by Tsiatis (1981) among others. Both properties will be derived in a more useful way for this investigation in Chapter III.

Letting k_o be the null hypothesis value of k,

$$\sqrt{N}(\hat{k}_{NR}-k_o) = \frac{\frac{1}{\sqrt{N}} \dfrac{\partial \log P(\underset{\sim}{R},k)}{\partial k}}{\frac{1}{N} \dfrac{\partial^2 \log P(\underset{\sim}{R},k)}{\partial k^2}\Bigg|_{k=k'}}$$

where k' is between k_o and $\hat{k}_{NR}$ we can calculate the asymptotic variance of $\sqrt{N}(\hat{k}_{NR}-k_o)$ as follows. Let $k=e^{\beta}$. Then $\log P(\underset{\sim}{R}, \beta) =$

$$N_1\beta - \sum_{\ell=1}^{N} [N-\ell+1] + (e^{\beta}-1)\, z_{N\ell}^*$$

$$\frac{\partial \log P(\underset{\sim}{R},\beta)}{\partial \beta} = n_1 - \sum_{\ell=1}^{N} \frac{z_{N\ell}^* e^{\beta}}{(N-\ell+1)+(e^{\beta}-1)z_{N\ell}^*}$$

$$\frac{\partial^2 \log P(\underset{\sim}{R},\beta}{\partial \beta^2} = - \sum_{\ell=1}^{N} \frac{[(N-\ell+1) + (e^{\beta}-1)z_{N\ell}^*]z_{N\ell}^* e^{\beta} - (z_{N\ell}^* e^{\beta})^2}{[(N-\ell+1) + (e^{\beta}-1)z_{N\ell}^*]^2}$$

$$= - \sum_{\ell=1}^{N} \frac{z_{N\ell}^* e^{\beta}}{(N-\ell+1)+(e^{\beta}-1)z_{N\ell}^*} \left[1 - \frac{z_{N\ell}^* e^{\beta}}{(N-\ell+1)+(e^{\beta}-1)z_{N\ell}^*}\right]$$

LEMMA: $$\underset{N\to\infty}{\text{p lim}} - \frac{1}{N} \frac{\partial^2 \log P(\underset{\sim}{R},B)}{\partial \beta^2} \equiv I_R(B) =$$

$$\int_o^{\infty} \frac{k}{\dfrac{\bar{H}(x)}{\lambda\bar{F}(x)} + k-1} \left[1 - \frac{k}{\dfrac{\bar{H}(x)}{\lambda\bar{F}(x)} + k-1}\right] dH(x)$$

where $H(x) = (1-\lambda)\, G(x) + \lambda F(x)$; $k = e^{\beta}$. See Proof.

PROOF:

$$\frac{1}{N}\sum_{\ell=1}^{N} \frac{e^{\beta}}{\frac{N-\ell+1}{Z_{N\ell}^{*}} + e^{\beta}-1}\left[1 - \frac{e^{\beta}}{\frac{N-\ell+1}{Z_{N\ell}^{*}} + e^{\beta}-1}\right]$$

$$= \int_0^{\infty} \frac{k}{\frac{\bar{H}_N(x)}{\lambda \bar{F}_{n_1}(x)} + e^{\beta}-1}\left[1 - \frac{k}{\frac{\bar{H}_N(x)}{\lambda \bar{F}_{n_1}(x)} + e^{\beta}-1}\right] dH_N(x)$$

$$= \int_0^{\infty} \frac{k}{\frac{\bar{H}_N(x)}{\lambda \bar{F}_{n_1}(x)} + e^{\beta}-1}\left[1 - \frac{k}{\frac{\bar{H}_N(x)}{\lambda \bar{F}_{n_1}(x)} + e^{\beta}-1}\right] d[H_N(x) - H(x)]$$

$$+ \int_0^{\infty} \frac{k}{\frac{\bar{H}_N(x)}{\lambda \bar{F}_{n_1}(x)} + e^{\beta}-1}\left[1 - \frac{k}{\frac{\bar{H}_N(x)}{\lambda \bar{F}_{n_1}(x)} + e^{\beta}-1}\right] dH(x).$$

where $H_N(x) = (1-\lambda)\, G_{n2}(x) + \lambda F_{n1}(x)$. $\lambda = \lim_{N\to\infty} n_1/N$, $0 < \lambda < 1$

Noting that $\frac{\bar{H}_N(x)}{\bar{F}_{N1}(x)} = \frac{\bar{H}_N(x)}{\bar{H}(x)} \cdot \frac{\bar{H}(x)}{\bar{F}(x)} \cdot \frac{\bar{F}(x)}{\bar{F}_{n1}(x)}$.

then by the Glivenko Cantelli Lemma,

$$\sup_{x,\ N\to\infty} \left| H_N(x)\, [H(x)/H_N(x) - 1] \right| \xrightarrow{a.s.} 0 .$$

and

$\sup_{\substack{x \\ N\to\infty}} | F_{n1}(x) \; [F_{n1}(x)/F(x) - 1] \xrightarrow{a.s.} 0$. Thus as

$$N \to \infty \quad \frac{\bar{H}_N(x)}{\bar{F}_{n1}(x)} \xrightarrow{a.s.} \frac{\bar{H}(x)}{\bar{F}(x)} \quad \forall x. \quad \text{QED.}$$

The result follows.

Corollary:

$$\text{If } \bar{F} = \bar{G}^k, \text{ then } I_R(k) = \frac{k}{\lambda(1-\lambda) \int_0^1 \frac{dx}{\lambda k+(1-\lambda)x^{1-k}}}, \; k=e^{\beta}.$$

$$I_R(\beta) = \int_0^1 \frac{k}{\frac{1-\lambda}{\lambda}\bar{G}^{1-k}+k} \left[1 - \frac{k}{\frac{1-\lambda}{\lambda}\bar{G}^{1-k}+k}\right] d[(1-\lambda)\bar{G}+\lambda\bar{G}^k]$$

$$= \int_0^1 \frac{k\lambda\,(1-\lambda)\,\bar{G}^{1-k}\,(\lambda k\bar{G}^{k-1}d\bar{G} + (1-\lambda)d\bar{G})}{[(1-\lambda)\,\bar{G}^{1-k} + k\lambda]^2}$$

$$= k\lambda(1-\lambda) \int_0^1 \frac{d\bar{G}}{k\,\lambda+(1-\lambda)\,\bar{G}^{1-k}} = k\lambda(1-\lambda) \int_0^1 \frac{dx}{k\,\lambda+(1-\lambda)x^{1-k}}$$

Thus the asymptotic variance of $\sqrt{N}(\hat{\beta}_{NR}-\beta)$ is

$$\left[k\lambda(1-\lambda) \int_0^1 \frac{dx}{\lambda k+(1-\lambda)x^{1-k}}\right]^{-1}.$$

Since $\sqrt{N}(\hat{\beta}_{NR}-\beta) \sim N(0, I_R(\beta)^{-1})$, then since $k = g(\beta) = e^{\beta}$, where g is monotone increasing in β, $\sqrt{N}(\hat{k}_{NR}-k) \sim N[0,g'(\beta)]^2 I_R(\beta)^{-1}]$. Then $\sqrt{N}(\hat{k}_{NR}-k) \sim N(0, \sigma_R^2)$ where

$$\sigma_R^2 = \frac{k}{\lambda(1-\lambda) \int_0^1 \frac{}{\lambda k+(1-\lambda)x^{1-k}}} = \frac{1}{I_R(k)}$$

$$= \frac{2k}{\int_0^1 \frac{dx}{k+x^{1-k}}} \quad \text{where } \lambda = \frac{1}{2}.$$

Since $P(Y>X) = \int_0^\infty [1-F(t)]dG(t) = \int_0^\infty e^{-(k+1)t}dt = \frac{1}{1+k}$,

then $\hat{R}_{NR}-R_o = \frac{1}{\hat{k}_{NR}+1} - \frac{1}{k+1} = \frac{1}{k+\frac{T}{\sqrt{N}}+1} - \frac{1}{k+1}$

$$= -\frac{T}{\sqrt{N}} \cdot \frac{1}{(k+1)^2} + \frac{T^2}{N} \cdot \frac{2}{(k+1)^3} + 0_p(\tfrac{1}{N}), \quad \text{where}$$

$\hat{k} = k_o + \frac{T}{\sqrt{N}}$ and $T \sim 0_p(1)$. Thus $N(\hat{R}_{NR}-R\ ) =$

$$-\frac{1}{(k_o+1)^2} \sqrt{N}(\hat{k}_{NR}-k) + o_p(1).$$ Finally, then,

$$\sqrt{N}(\hat{R}_{NR}-R) \sim N(0,\sigma_R^2), \quad \text{where } \sigma_R^2 = \frac{2k}{(k+1)^4} \left[\int_0^1 \frac{dx}{k+x^{1-k}}\right]^{-1}.$$

U-STATISTIC AS ESTIMATOR OF P(Y > x)

The U-statistic $\hat{R}_{NW} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(Y_i > X_j)$ is clearly an and unbiased estimator for R, where $I(T) = \begin{cases} 1 & \text{if } T > 0 \\ 0 & \text{otherwise} \end{cases}$.

Letting $\phi(Y_j, X_i) \equiv I(Y_j - X_i)$ and noting that $E\phi^2(Y_j, X_i) < \infty$ then it follows from the two-sample U-statistics theorem that

$$\sqrt{N}(\hat{R}_{NW}-R) \sim N(0, V_2),$$

where
$$V_2 = \frac{E_{10}}{1-\lambda} + \frac{E_{01}}{\lambda}$$

where $\lambda = \lim_{N\to\infty} \frac{n_1}{N} \quad 0 < \lambda < 1.$

and
$$E_{1,0} = E[I(Y_1-X_1)\, I(Y_2-X_1)] - R^2$$
$$= Pr(X_1 < Y_1, X_1 < Y_2) - R^2$$
$$= \int_0^\infty [1-F(x)]^2\, dG(x) - \left[\int_0^\infty [1-F(x)]dG(x)\right]^2$$

$$E_{0,1} = E[I(Y_1-X_1)\, I(Y_1-X_2)] - R^2$$
$$= Pr(X_1 < Y_1, X_2 < Y_1) - R^2$$
$$= \int_0^\infty G^2(y)dF(y) - \left[\int_0^\infty [1-F(x)]dG(x)\right]^2$$

Under H_o , $1-F(y) = [1-G(y)]^k$

Thus $y = F^{-1}\{1-[1-G(y)]^k\}$. Also, $1-G(x) = [1-F(x)]^{1/k}$. Letting $s = F(x)$, then $GF^{-1}(s) = G(x) = 1-(1-s)^{1/k}$

$$\int_0^\infty [1-F(x)]^2 dG(x) = \int_0^1 [1-FG^{-1}(x)]^2 dt = \int_0^1 (1-t)^{2k}\, dt = \frac{1}{2k+1}$$

$$\int_0^\infty G^2(y)\,dF(y) = \int_0^1 [GF^{-1}(s)]^2 ds = \frac{2}{(k+1)\ (k+2)}$$

$$\sigma_W^2 = \lim_{N\to\infty} N\,\mathrm{Var}[\hat{R}_{NW}-R] = \frac{\frac{1}{2k+1} - \frac{1}{(k+1)^2}}{1-\lambda} + \frac{\frac{2}{(k+1)(k+2)} - \frac{1}{(k+1)^2}}{\lambda}$$

$$= \frac{2k}{(k+1)^2}\left[\frac{k^2+4k+1}{(2k+1)(k+2)}\right]$$ when the sample sizes are equal, i.e. $\lambda = \frac{1}{2}$.

Notice that $\lambda(1-\lambda)\,\sigma_W^2$ is still a function of λ , i.e., $\sigma_W^{\ 2}$ is not a symmetric function of λ. In many classical problems $\lambda(1-\lambda)\sigma_W^{\ 2}$ is independent of λ indicating in some cases that $\lambda = \frac{1}{2}$ produces the least variance. However, that is not the case here except in the case where k = 1, i.e., the classical case of interchangability. When k = 1, $\sigma_W^{\ 2} = \frac{1}{12\lambda(1-\lambda)}$, so $\min_\lambda \sigma_W^{\ 2} = \frac{1}{3}$ when $\lambda = \frac{1}{2}$.

For any other value of k, it may not be optimal to choose equal sample sizes. However, since k is not known in advance, it may be difficult to justify different sample sizes. The same comment applies to $\sigma_R^{\ 2}$.

THE ESTIMATORS UNDER NON-PROPORTIONAL HAZARDS

When H_o is not true, the U-statistic has the advantage of being unbiased for all alternatives. Thus $\sqrt{N}\,(\hat{R}_{NW}-R) \sim N(0,*)$ where $*$ depends upon the specific alternative. However, $\sqrt{N}(\hat{R}_{NR}-R) \to \infty$ as $N \to \infty$ since $\hat{k}_{NR} \to k^* \neq k$ as $N \to \infty$ when H_o is not true. k^* is the solution to the MLE equation:

$$\frac{1}{k^*} = \int_0^\infty \frac{dH(x)}{\dfrac{(1-\lambda)e^{-x}}{\lambda\bar{F}(x)} + k^*}$$

Thus, when H_o is not true, $\sqrt{N}\,(\hat{R}_{NR}- R)$ does not even have an asymptotically normal distribution. This fact indentifies the importance of finding a test of H_o which will be consistent against a broad class of alternatives.

In summary, the lower variance of $\sqrt{N}(\hat{R}_{NR}-R)$ under H_o and the unbiasedness of $\sqrt{N}(\hat{R}_{NW}-R)$ under H_A motivates the use of a preliminary test estimator of $\hat{R}$ For a fixed alternative, the consistency of the test will determine that we will always use $\hat{R}_{NW}$ and so our estimator will be unbiased with a variance which has been estimated by others. The burden of this paper will be to investigate the behavior of the preliminary test estimator when H_o is true (Chapter III) and under a sequence of local alternatives, H_{NA} (Chapter IV), since the power properties of the test and the resulting expectation and mean squared error of the PTE are not meaninful if the power is arbitrarily close to one as $N \to \infty$.

COMPARISON OF RANK MLE WITH U-STATISTIC

For a preliminary test estimator to be a reasonable procedure, we should expect $\sigma_R^2 \leq \sigma_W^2$ since we are deriving the estimator with asymptotic variance σ_R^2 under the asumption of proportional hazards. In fact $\frac{\sigma_W^2}{\sigma_R^2}$ lies between 1.32 and 1.33 for all values of k between 1.0 and 3.0.

We have thus produced two possible estimators of P(Y>X). One is unbiased and is always asymptotically normally distributed no matter what the distributions of the two populations. The other is the rank MLE computed under the assumption that the two populations are related by proportional hazards. It is consistent for P(Y>X) when proportional hazards holds, but in general will not be if proportional hazards does not hold. In addition we have confirmed the usefulness of such a preliminary test procedure by the lower variance of the rank MLE which proportional hazards hold.

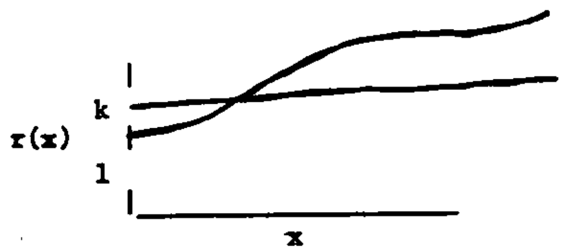
CHAPTER III

In this chapter we begin by stating the null and alternative hypotheses in rigorous ways which allow the state of nature to be parameterized by families of probability distributions. Then we develop the asympotic null distribution of a linear rank test for departures from proportional hazards. A wide class of fixed alternatives for which the stratum is consistent is derived; and finally we investigate the performance for the preliminary test estimator when H_o is true.

Null and Alternative Hypotheses

H_o is states as follows: the ratio of the hazard rates of two independent groups is equal to an unknown constant k for all $T > 0$. There are two classes of alternatives for this study:

H_A : A fixed alternative to H_o is a hazard ratio profile such that one group's hazard is always greater than the other's and the ratio is "essentially" monotone increasing. "Essentially" means that at the beginning of the study, the hazard ratio is less than some constant k, at the end of the study the ratio is greater or equal to k and that after the beginning of the study, the hazard ratio equals k only once. Referring back to the equation in Chapter II indicating the invariance of the hazard ratio under a monotone transformation, we choose one of the transformed random variables to be exp (1) and the other to be a distribution with "essentially" increasing hazard rate. In order to impose a compatability condition on the now "alternative distribution" to exp(k). we require that the alternative F* have finite first and second moments where $E(x) = 1/k$. See illustration.



Thus the problem of detecting non-constant hazard has been transformed to a problem of detecting departure from a mixture of exp(1) and exp(k) distributions in the distribution one group having a distribution with "essentially" increasing hazard. Thus

$$H_o: \ r(x) \equiv k \qquad x \geq 0$$

$$H_a: \ r(x) = k + \psi(x), \text{ where } \psi(0) < k,$$

$$\psi(x_o) = k \ (x_o > 0 \text{ exists and is unique})$$

$$\text{and} \quad \int_o^\infty e^{-kx - \int_o^x \psi(t)dt} dx = 1/k$$

H_{AN}: This is a root-n sequence of local alternatives to proportional hazards. Just as for fixed alternatives,

$\int_o^\infty \bar{F}(x)dx = 1/k$; however $\bar{F}_N(x) = f(x; \kappa_N, \theta_N)$ where

$$\kappa_n = k + \Delta/\sqrt{N}, \ \theta_n = \theta_o + \delta/\sqrt{N}$$

As $N \to \infty$ $F(x, \kappa_N, \theta_N) \longrightarrow 1 - e^{-kx} \ \forall x.$

However, there is a mathematical technicality which must be addressed concerning this sequence of alternatives. Recall the transformation from the original spare:

$[f(x)/1-F(x)]/[g(x)/1-G(x)] = f^*(x^*)/1-F^*(x^*)$ where

$g^*(x^*)/1-G^*(x^*) = 1$. Note that if the LHS =0, thus $f(0) = 0 \Rightarrow f^*(0) = 0$. But under a sequence of alternatives we would like $f^*(x^*)/1-F^*(x^*) \equiv r(x)$ to be equal to k, uniformly in x. Thus the transformation induces a singularity at the origin. To avoid this problem, we modify the statement of the local alternative slightly so that for any sequence $\{x_n\}$ where lim $x_N = 0$, then $\lim_{N\to\infty} r(x_N) \to k$. Finally, we note that since we index the sequence by a two-dimensional parameter space generating a family of distributions, and that there is a restriction

$\int_0^\infty \bar{F}(x)dx = \frac{1}{k}$ then the sequence is along a curve embedded in the unrestricted parameter space. For the remainder of this thesis, we drop the (*) notation for convenience so that all derivations are done in the transformed space.

The Inappropriateness of the Neyman C_α test

Apart from consistency, a desirable feature of any test is optimal power under a sequence of local alternatives. The presence of a nuisance paramter in any candidate for the statistic suggests basing the test on a principle developed by Neyman (1959). In his seminal paper, Neyman showed that if interest centered on testing against the null value of a parameter when a nuisance parameter is present, it is possible to construct an asymptotically locally most powerful test by computing the efficient score statistic based on the parameter of interest, and at the same time, substituting a "weakly" consistent or "strongly" consistent estimate of the nuisance parameter into the resulting score statistic. θ_n is "weakly" consistent if

$\sqrt{N}(\hat{\theta}_n - \theta) \sim O_p(1)$ and $\hat{\theta}_n$ is "strongly" consistent if

$\sqrt{N}(\hat{\theta}_n - \theta) \sim o_p(1)$. In this former case, other terms must be added to the basic score statistic to provide the optimal properties. It is this general framework which suggests that locally optimal rank test can be found.

For all we would have to do is take the derived adaptive test statistics and produce rank scores by taking expectations given the rank vector. That is, we would treat the problem as "parametric", derive the appropriate "parameter" c statistic and take expected values given the rank vector. Unfortunately, this procedure will not work in the present context. Recall that in the case of a simple rank score when the null hypothesis is i.i.d. random variables, the optimality of the efficient score requires the score function to equal the derivative of the log likelihood. Only then will the Cauchy-Schwarz inequality become an equality and yield a maximum for the non-centrality parameter. We may then conclude that the rank statistic is asymptotically efficient. However, if the random variables are <u>not</u> i.i.d. under H_o, we shall see in the next section that the statistic cannot be expressed as a simple sum of "efficient scores" which are equal to the derivative of the log-likelihood under a specified sequence of local alternatives. Consequently, there is no equivalence of the Cauchy-Schwarz inequality in the rank statistic case to the parametric cases as there is when there are i.i.d. random variables under H_o.

Moreover, even if it were possible to construct an "optimal test" using the Neyman theory, there would be at least two draw backs: 1) the expected value of a function of an order statistic depends upon k in a very complicated way and 2) computation of the variance of the statistic would be inordinately difficult due to both the extra term and the estimation of k by $\hat{k}$.

In addition, it can be argued that efficiency of the statistic is not really relevant here. After all, the bona fide null hypothesis is proportional hazards in the untransformed space, not two independent exponential populations in the transformed space. Thus there is no test in the untransformed space with which to compare this test. Even if we make the comparison in the transformed space, we run into problems; for there is no such thing as a two-sample parametric test for detecting departure from two exponentials based upon the joint likelihood, where one of the populations is exp(1) under both null and alternative hypotheses. Lastly, there is no indication, a priori, that optimality (in some sense) of the test statistic is related to desirable performance of the PTE, the main focus of this study.

THE CHERNOFF-SAVAGE THEOREM (1958)

Let $Y_1, \ldots Y_n$. be the ordered observations of a random sample from a population with continuous cumulative distribution function $F(x)$. Let $x_1, \ldots x_{n2}$ be the ordered observation of a random sample from a population with continuous cumulative distribution function $G(x)$. Let $N = n_1 + n_2$, $\lambda = n_1/N$ $0 < \lambda < 1$. Define $H_N(x) = (1-\lambda)\, G_{n2}(x) + \lambda F_{n1}(x)$.

Define $T_N = J_N\, [H_N(x)]\, dF_{n_1}(x)$

$= 1/n_1 \sum_{i=1} Eni\ Zni$ where

the Eni are scores and $Zni = \begin{cases} 1 \text{ if obs. is from Y-sample} \\ 0 \text{ otherwise} \end{cases}$

Theorem: If

(1) $J(H) = \lim J(H)$ exists for $0 < H < 1$ and is not constant.

(2) $\int_{I_N} [J_N(H_N) - J(H_N)]\, dF_{n1}(x) = o_p(N^{-1/2})$

(3) $J_N(1) = o(\sqrt{N})$

(4) $J^{(i)}(H) = | d^iJ/dH^i | < K\, [H(1-H)]^{-i - 1/2 + \delta}$

for $i = 0,1,2,$ and for some $\delta > 0$,

then for fixed F, G, λ,

$$\lim P\, (T_n - \mu_T / \sigma_T < t) = \int_{\infty}^{t} 1/\sqrt{2\pi} e^{-x^2/2}\, dx$$

where $\mu_T = \int_{-\infty}^{\infty} J(H(x))\, dF(x)$

and
$$N\sigma_T^2 = 2(1-\lambda) \{\iint_{-\infty < x < y < \infty} G(x)[1-G(y)J'(H(x))\, J'(H(y))\, dF(x)\, dF(y)$$
$$+ (1-\lambda)/\lambda \iint F(x)[1-F(y)]J'(H(x)J'(H(y))\, dG(x)\, dG(y)\}$$

providing $\sigma_N \neq 0$.

The utility of the theorem is in providing the asymptotic normality of a linear rank statistic based upon a mixture of distributions. As developed in Chapter I, under H_o, the combined sample consists of n_1 random variables from a population with cdF $F=1-(1-G)^k$ and n_2 random variables with cdF G. Thus, in this investigation $F(x) = F = F(x,k)$ where k is unknown. Consequently $J[H(x)] = J[H(x,k)]$.

The condition on the second deviation in condition (4) has since been shown to be unnecessary. Stronger results need only $i = 0,1$. The older approach has been taken here for too reasons: 1) the 2nd deviatives are simple and 2) this formulation is useful for the Taylor expansion of $\hat{k}_{NR} - k$.

Adjustment in T_N for Estimation k

Since k is estimated from the data, we expand $T_n(\hat{k}_{NR})$ in a Taylor series about the true value k. Since $\sqrt{N}(\hat{k}_{NR}-k) = O_p(1)$ under H_o, we can write $k_N = k + t/\sqrt{N}$. Thus $T_N(k+t/\sqrt{N}) = T_N(k) + t B + o_p(N^{-1/2})$ for a suitable constant B. We can also express the above as

$$\sqrt{N}\,[T_N(k_N) - T_N(k) - (\hat{k}_{NR}-k)B] \xrightarrow{P} 0 \quad \text{as } N \longrightarrow \infty .$$

In general, Chernoff and Savage show that

$$T_N = \int_{-\infty}^{+\infty} J(H(x)dF(x) + (1-\lambda)\ \{1/n_1 \sum_{i=1}^{n_1}[B(Y_i) - E\,B(Y_i)]$$

$$- 1/n_2 \sum_{j=1}^{n_2}[B(X_j) - EB(X_j)]\}$$

$$+ O_p(N^{-1/2})$$

where $B(Y_i) = \int_{Y_o}^{Y_i} J'[E(y)]dG(y)$

$B(X_j) = \int_{X_o}^{X_i} J'[H(x)]dF(x)$

thus

$$T_N \xrightarrow{P} \int_{-\infty}^{+\infty} J(H(x))\,dF(x) \quad \text{as} \quad N \longrightarrow \infty$$

Consequently

$$B = \lim_{N\to\infty} \frac{\int_{-\infty}^{\infty} J[H(x,k+t/\sqrt{N})]\,dF(x) - \int_{-\infty}^{\infty} J[H(x,k)]\,dF(x)}{t/\sqrt{N}}$$

$$= \lim_{N\to\infty} \frac{\int_{-\infty}^{+\infty} \{J[H(x,k+t/\sqrt{N})] - J[H(x,k)]\}\,dF(x)}{t/\sqrt{N}}$$

If

$$\frac{\partial J[H(x,k)]}{\partial k} < M(x) \quad \text{where} \quad \int_{-\infty}^{+\infty} M(x)\,dF(x) < \infty,$$

then $B = \int_{-\infty}^{+\infty} \frac{\partial J[H(x,k)]}{\partial k}\,dF(x)$.

In fact, let k^* be a running variable in the expression for $EJ(H(Y_i,k^*)) = 0$ where $Y_i \sim \exp(k^*)$ then under regularity conditions (see appendix),

$$d/dk^* \int_0^{\infty} J(H(y,k^*)]\, k^* e^{-k^* y}\,dy =$$

$$\int_0^{\infty} \frac{\partial\{J\{H(y,k^*)]k^* e^{-k^*y}\}}{\partial k^*}\,dy$$

$$= -\int_0^{\infty} yJ[H(y,k^*)]k^* e^{-k^* y} + \int_0^{\infty} \frac{\partial J[H(y,k^*)]}{\partial k^*}\, k^* e^{-k^* x} = 0$$

Setting $k^* = k$, then we find that

$$B = \int_0^{\infty} yJ[H(y,k)]\, ke^{-k\,y} = \mathrm{Cov}\,[Y_i, J[H(Y_i,k)]]$$

Theorem: Let $t = \sqrt{N}(\hat{R}_{NR}-R) > 0$, so that t is $O_p(1)$ on a compact interval of the real line. Then, for $C > 0$,

$$\Pr(\max_{t<C} |\sqrt{N}\,[T_N(\hat{k}_{NR} - T_N(k)] - tB| > \varepsilon) \longrightarrow 0$$

as $N \longrightarrow \infty$. This statement is important because it says that when we substitute $\hat{k}$ for k in $T_N(k)$, then the adjustment is good uniformly on a compart (bounded and closed) interval, not just pointwise. The theorem is proved by adapting a method due to Jurescova (1969) which she used in a different context.

First, partition the interval $[-c,c]$ into r divisions, such that $-c$ $\Delta_o < \Delta_1 < \ldots < \Delta_r = c$. From the above, we know that we can chose r large enough so that for N large enough,

$$Pr\{|\sqrt{N}(T_{N\Delta} - T_{No}) - \Delta_i B| > 1/4\,\varepsilon\} < \eta/r \text{ where}$$

$\eta, \varepsilon > 0$, $i = 0,\ldots,r$. Simultaneously, N is large enough so that $(\Delta_i - \Delta_{i-1})B < 1/2\,\varepsilon$. Say $N > N_o$.

Now choose any $\Delta\varepsilon[-c,c]$, $\Delta_{i_o-1} < \Delta < \Delta_{i_o}$

Assume $B > 0$ so that $T_{N\Delta}$ is non-decreasing in Δ.

The case $B < 0$ can be handled analogously.

$\sqrt{N}T_{N\Delta} - \sqrt{N}T_{No} - \Delta B \geq 0$, then $|\sqrt{N}T_{N\Delta} - \sqrt{N}T_{No} - \Delta B| \leq$

$$\sqrt{N}T_{N\Delta i_o} - \sqrt{N}T_{No} - \Delta B \leq \sqrt{N}T_{N\Delta i_o} - \sqrt{N}T_{No} - B\Delta_{i_o-1}$$

$$= \sqrt{N}T_{N\Delta i_o} - \sqrt{N}T_{No} - B\Delta_{i_o} + B\Delta_{i_o} - B\Delta_{i_o-1}$$

$$\leq |\sqrt{N}T_{N\Delta i_o} - \sqrt{N}T_{No} - B\Delta_{i_o}| + |B(\Delta_{i_o} - \Delta_{i_o-1})|$$

$$\leq |\sqrt{N}T_{Ni_o} - \sqrt{N}T_{No} - B\Delta_{i_o}| + |B(\Delta_{i_o} - \Delta_{i_o-1})| + |\sqrt{N}T_{N\Delta i_o} - \sqrt{N}T_{No} - \Delta_{i_o-1}|$$

Thus
$$\max_{|\Delta| \leq c} |\sqrt{N}T_{N\Delta} - \sqrt{N}T_{N_0} - B\Delta| \leq$$

$$2 \max_{o \leq i \leq r} |\sqrt{N}T_{N\Delta_i} - \sqrt{N}T_{No} - B\Delta_i| + (1/2)\varepsilon$$

Thus $\Pr\{ \max_{|\Delta| < c} |\sqrt{N} T_{N\Delta} - T_{No} - B\Delta| > \varepsilon\} \leq$

$$\Pr\{2 \max_{o \leq i \leq r} |\sqrt{N} T_{N\Delta_i} - \sqrt{N} T_{N\Delta_o} - B\Delta_i| + (1/2)\varepsilon > \varepsilon\}$$

$$= \Pr\{\max_{o \leq i \leq r} |\sqrt{N}\, T_{N\Delta_i} - \sqrt{N} T_{No} - B\Delta i| > \varepsilon/4\} < \eta$$

by construction QED.

Asymptotic Distribution of $\sqrt{N}\,[T_N(\hat{k}_{NR}) - \mu]$

By the Chernoff-Savage theorem, $\sqrt{N}\,[T_N(k)-\mu]$ converge distribution to a normal r.v. with mean zero and finite variance, where $\mu=\mu(k)= \int_{-\infty}^{+\infty} J[H(x,k)]\, dF(x,k)$. We must now investigate the effect using the adoptive statistic $T_N(\hat{k}_{NR})$. Since $\mu(\hat{k}_{NR}) \to \mu(k)$ in probability (since $\hat{k}$ is consistent for k), then we can write

$$\sqrt{N}[T_N(\hat{k}_{NR})-\mu] = \sqrt{N}[T_N(k) - \mu] + \sqrt{N}(\hat{k}_{NR}-k)B$$

with probability approaching 1 as $N \to \infty$. Since the RHS is the sum of two asympotically normally distributed random variables, $\sqrt{N}\,[T_N(\hat{k})- \mu]$ is asympotically normally distributed with mean zero and some variance. We can now write that

$$N \operatorname{Var} T_N(\hat{k}_{NR})-\mu] = N \operatorname{Var}[T_N(k)-\mu] + 2BN\operatorname{Cov}[T_N(k)-\mu,\ \hat{k}_{NR}-k]$$

$$+ NB^2\operatorname{Var}(\hat{k}_{NR}-k), \text{where } N \operatorname{Var}(\hat{k}_{NR}-k)= [I_R(k)]^{-1}$$

The first term is given by the Chernoff-Savage theorem. The third term is given by $\lim_{N\to\infty} \frac{N}{I_R(k)} = \frac{2K}{\int_0^1 \frac{dx}{k+x^{1-k}}}$ derived in Chapter II.

In order to derive the second term, it us necessary to express $\hat{k}_{NR}-k$ in a form using a linear combination of random variables, just as in the Chernoff-Savage theorem. We do this by noting, as in Chapter II, team the Taylor series expansion for $\hat{k}_{NR}-k$ is

$$\hat{k}_{NR}-k = \frac{\frac{\partial \log P(\underset{\sim}{R},k)}{\partial k}}{\underset{-}{<} \frac{\partial^2 \log P(\underset{\sim}{R},k)}{\partial k^2}} + o_p(N^{-\frac{1}{2}})$$

where $\frac{\partial \log P(\underset{\sim}{R},k)}{\partial k} = \lambda/k - \int_0^\infty \frac{1}{\frac{1-\lambda}{\lambda}\frac{\bar{G}_{n2}(x)}{\bar{F}_{n1}(x)} + k} dH_N(x)$

Let $M(F_{n1}, G_{n2})$ be the integrand of the second term. Then we expand $M(F_{n1}, G_{n2})$ about $M(F,G)$ in a Taylor series where $\bar{F}(x) = [\bar{G}(x)]^k$.

1. $M(F_{n1}, G_{n2}) = M(F,G) + (F_{n1}-F)\partial M/\partial F_{n1}(F,G) + (G_{n2}-G)\partial M/G_{n2}(F,G)$

$$+ (1/2)(F_{n1}-F)^2\, \partial^2 M/\partial F_{n1}^2\, (V_a)$$

$$+ (F_{n1}-F)(G_{n2}-G)\, \partial^2/\partial F_{n1}\partial G_{n2}\, V_a)$$

$$+ 1/2\, (G_{n2}-G)^2\, \partial^2 M/\partial G_{n2}\, (V_a)$$

where $V_a = [a(F_{n1}, G_{n2}) + (1-a)F, G] \quad 0 < a < 1$

2. $dH_N(x) = d[H_N(x) - H(x) + H(x)]$

Let $M_1(F) = \partial M / \partial F_{n1}(F, G)$, $M_2(G) = \partial M / \partial G_{n2}(F, G)$

then $M(F_{n1}, G_{n2})\, dH_N = M(F, G)\, d\, [H_N - H) + M(F, G) dH$

$$+ (F_{n1} - F)\, M_1(F) + (G_{n2} - G)\, M_2(G)\, d[H_N - H] +$$

$$(F_{n1} - F)\, M_1(F)\, dH[H_N - H] + (G_{n2} - G)\, M_2(G)(F, G) dH +$$

$$1/2\, (F_{n1} - F)^2\, \partial^2 M / \partial F_{n1}^2\, (V_a)\, dH_N$$

$$+ (F_{n1} - F)(G_{n2} - G)\, \partial^2 M / \partial F_{n1} \partial G_{n2}\, (V_a) dH_N + (1/2)(G_{n2} - G)^2\, \partial^2 M / \partial G_{n2}^2 (V_a) dH_N.$$

$$= M(F, G) dH + (F_{n1} - F)\, M_1(F) dH + (G_{n2} - G)\, M_2(G) dH$$

$$+ M(F, G) d\, (H_N - H) + \sum_{i=1}^{5} R_{Ni}$$

$$R_{1N} = \int_0^1 (F_{n1} - F)\, M_1(F) d[H_N - H]$$

$$R_{2N} = \int_0^1 (G_{n2} - G)\, M_2(G) d[H_N - H]$$

$$R_{3N} = \int_0^1 (1/2)(F_{n1} - F)^2\, \partial^2 M / \partial F_{n1}^2\, (V_a) dH_N$$

$$R_{4N} = \int_0^1 (F_{n1}-F)\ (G_{n1}-G)\ \partial^2 / \partial F_{n1} \partial G_{n2}\ (V_a) dH_N$$

$$R_{5N} = \int_0^1 (1/2)(G_{n2}-G)^2\ \partial^2 M / G_{n2}{}^2\ (V_a) dH_N$$

Since $1-F = (1-G)^k$, then $M(F,G) = M_F(F) = M_G(G)$

$$\equiv \frac{\lambda}{(1-\lambda)\ \bar{F}^{\frac{1}{k}-1} + k\lambda} \equiv \frac{\lambda}{(1-\lambda)\ \bar{G}^{1-k} + k\lambda}$$

$M(F,G)dH = \lambda M_F(F) + (1-\lambda) M_G(G)dG$

$$(F_{n1}-F)\ M_1(F)dH = \lambda(F_{n1}-F)\ M_1(F)dF + (1-\lambda)(F_{n1}-F)\ M_1(F)dG$$

$$(G_{n2}-G)\ M_2(G)dH = \lambda(G_{n2}-G)\ M_2(G)dF + (1-\lambda)\ (G_{n2}-G)\ M_2(G)dG$$

$$M(F,G)\ d[H_N-H] = \lambda M_G(G)\ d[F_{n1}-F] + (1-\lambda)\ M_G(G)\ d[G_{n2}-G]$$

Integrating the last expression by parts we find that

$$\int^{\infty} M(F_{n1}(x), G_{n2}(x))\ dH_N(x) = \int^{\infty} M_F[F(x)]\ dH(x)$$

$$+ \lambda \int_o^{\infty} [F_{n1}(x)-F(x)]\ M_1[F(x)]\ dF(x) + (1-\lambda) \int_o^{\infty} [F_{n1}(x)-F(x)]\ M_1[F(x)] dG(x)$$

$$+ \lambda \int_o^{\infty} [G_{n2}(x)-G(x)]\ M_2[G(x)]\ dF(x) + (1-\lambda) \int_o^{\infty} [G_{n2}(x)-G(x)]\ M_2[G(x)] dG(x)$$

$$- \lambda \int_o^{\infty} [F_{n1}(x)-F(x)]\ dM_G[G(x)] - (1-\lambda) \int_o^{\infty} [G_{n2}(x)-G(x)]\ dM_G[G(x)]$$

$$+ \sum_{i=1}^{5} R_{Ni}.$$

Evaluating $\partial M/\partial G_{n2}$ at (F,G), noting that $\bar{F} = \bar{G}^k$ then

$$M_2(G) = \frac{\frac{1-\lambda}{\lambda}(1-G)^{-k}}{[\frac{1-\lambda}{\lambda}(1-G)^{1-k} + k]^2}$$

Similarly, $$M_1(F) = \frac{-\frac{1-\lambda}{\lambda}(1-F)^{1/k-2}}{[\frac{1-\lambda}{\lambda}(1-F)^{1/k-1} + k]^2}$$

Noting that $dM_G(G) = (1-k)\ M_2(G)dG$ and

$$dM_F(F) = (1/k - 1)\ M_1(F)dF = (k-1)/k\ M_1(F)dF,$$

we have

$$\int_0^\infty M(F_{n1}(x),\ G_{n2}(x))dH_N(x) =$$

$$\lambda\ (k/k-1)\int_0^\infty [F_{n1}(x)-F(x)]\ dM_F[F(x)] + (1-\lambda)(k/k-1)\int_0^\infty [F_{n1}(x)-F(x)]M_F(F(x)]dG(x)$$

$$+\lambda/1-k\int_0^\infty [G_{n2}(x)-G(x)]\ M_G'\ [G(x)] + (1-\lambda)/(1-k)\int_0^\infty G_{n2}(x)-G(x)]\ dM_G[G(x)]$$

$$-\ \lambda\int_0^\infty [F_{n1}(y) - F(y)]dM_G[G(x)] - (1-\lambda)\int_0^\infty [G_n n_2(x) - G(x)]dM_G[G(x)]$$

Collecting terms we have on the RHS

$$\lambda/(k-1)\int_0^\infty [F_{n1}(x)-F(x)]\ dM_F[F(x)] = \frac{k(1-\lambda)}{k-1}\int_0^\infty [F_{n1}(x)-F(x)]\ M_F'\ [F(x)]dG(x)$$

$$+\ \lambda/(1-k)\ [G_{n2}(x)-G(x)]\ M_G'\ [G(x)]\ dF(x) + \frac{(1-\lambda)k}{(1-k)}\int_0^\infty [G_{n2}(x)-G(x)]\ dM_G[G(x)]$$

$$+\sum_{i=1}^{5} R_{Ni}$$

Integrating by parts and collecting terms, we have finally that

$$\int_0^\infty M[F_{n1}(x),\ G_{n2}(x)]\ dH_N(x) = \int_0^\infty M_F[F(x)]\ dH(x)$$

$$+\ \lambda/(k-1)\ \{1/n_2 \sum_{j=1}^{n_2} [D_2(X_i) - ED_2(X_i)] + 1/n_1 \sum_{i=1}^{n_1} (M_F[F(Y_i)] - EM_F[F(Y_i)]\}$$

$$+\ \ k(1-\lambda)/(k-1)\ \{1/n_2 \sum_{j=1}^{n_2} (M_G[G(X_i)] - EM_G[G(X_i)]) - 1/n_1 \sum_{i=1}^{n_1} [D_1(Y_i) - ED_1(Y_1)\ ]$$

$$+ \sum_{i=1}^{5} R_{Ni}$$

where $D_1(Y_i) = \int_{Y_o}^{Y_i} M_F'[F(x)]\ dG(y)$

$D_2(X_i) = \int_{X_o}^{X_i} M_G'\ [G(x)]\ dF(x)$

Now $\int_0^\infty M_F[F(x)]\ dH(x) = \int_0^\infty M_G[G(x)]\ dH(x)$

$$= \int_0^1 M_G(G)\,dH = \lambda \int_0^1 \bar{G}^{k-1} \frac{[1-\lambda) + k\lambda\ \bar{G}^{k-1}}{(1-\lambda) + \lambda k \bar{G}^{k-1}}\ d\bar{G}$$

$$= \lambda/k \quad \text{since} \quad d\bar{H} = [1-\lambda) + \lambda \bar{G}^{k-1}]\,d\bar{G}$$

and $\frac{\bar{G}}{\bar{F}} = \bar{G}^{1-k}$. We will show that $\sum_{i=1}^{5} R_{Ni} = o(N^{-1/2})$. (See Appendix A).

Thus $\hat{k}_{NR}-k$ can be expressed as a linear combination of a mixture of independent random variables with zero mean and finite variance. The consistency of $\hat{k}_{NR}$ for k is immediate. In addition, the asymptotic normality of $\sqrt{N}\ (\hat{k}_{NR}-k)$ follows from the central limit theorem since each

D-function and M-function has finite second moment. This follows since the asymptotic variance of $\sqrt{N}\,(\hat{k}_{NR}-k)$ was shown earlier to be bounded function of $k < \infty$.

Finally, asymptotic joint normality of $\sqrt{N}(\hat{k}_{NR}-k)$ and $\sqrt{N}(T_N(k) - \mu)$ follows by expressing the pair as a vector, each of whose components is the sum of independent random variables with finite second moment. The result follows from the multivariate central limit theorem.

THEOREM: $\hat{k}_{NR} \longrightarrow \dfrac{\bar{x}_N}{\bar{y}_N}$ where $X \sim \exp(1)$, $Y \sim \exp(k)$ as $k \longrightarrow 1$. In other words, the rank maximum likelihood estimator converges in probability to the efficient estimator of k as the two groups approach homogeneity.

PROOF: Using a Taylor series expansion, we may write

$$\frac{\bar{x}_N}{\bar{y}_N} = k-k(1-\bar{x}_N) + k^2(1/k- \bar{y}_N) + o_p(N^{-1/2}) \quad \text{or} \quad \bar{x}_N/\bar{y}_N - k$$

$$= k(\bar{x}_N-k\bar{y}_N).$$

Using the binomial theorem, we can write $D_1(Y_i)$, $D_2(X_i)$ as infinite series and integrate term by term. We do not have to be concerned with constant terms since they will cancel when

$ED_1(Y_i)$, $ED_2(X_i)$, $E_2M_F[F(Y_i)]$ and $EM_G[G(X_i)]$ are subtracted off the stochastic terms. Also, when we expand our expression about $k = 1$, we need only account for those terms with $k-1$ as a coefficient since that term will be divided by $k-1$ and the rest of the stochastic terms will be of small order $(k-1)$.

Letting $U_i = 1-G(x_i)$, then substitution yields

$$D_1(Y_i) = (1-1/k) \int_{u_o}^{u_i} \frac{p\, u^{1-2k}}{(p\, u^{1-k}+k)^2}\, du \,, \quad \text{where} \quad p = \frac{1-\lambda}{\lambda}$$

$$D_2(x_j) = k(k-1) \int_{u_0}^{u_1} \frac{p\, u^{-1}}{(p\, u^{1-k}+k)^2}\, du$$

X-sample: $k(k-1)\; 1/p \int_{u_o}^{u_i} \left[\sum_{j=o}^{\infty} \binom{n}{j} u^{\,j(1-k)-1} \left(\frac{k}{p}\right)^{-2-j} \right] du$

converges when $u^{1-k} < \frac{k}{p}$.

Since each term is bounded, we can reverse the order of integration and summation:

$$D_2(x_i) = \frac{k(k-1)}{p} \sum_{j=o}^{\infty} \int_{u_o}^{u_i} \binom{-2}{j} u^{j(1-k)-1} \left(\frac{k}{p}\right)^{-2-j} du$$

$$= \frac{k(k-1)}{p} \int_{u_o}^{u_i} \left[\left(\frac{k}{p}\right)^{-2} u^{-1} + \sum_{j=1}^{\infty} \binom{-2}{j} u^{\,j(1-k)-1} \left(\frac{k}{p}\right)^{-2-j} \right] du$$

$$= \frac{k(k-1)}{p} \left[\left(\frac{k}{p}\right) \log u_i + \sum_{j=1}^{\infty} \binom{-2}{j} \frac{u_i^{\,j(1-k)}}{j(1-k)} \left(\frac{k}{p}\right)^{-2-j} \right] + \text{constant.}$$

Expanding $u_i^{\,j(1-k)}$ about $k=1$, we get

$$\frac{k(k-1)}{p} \left[\left(\frac{k}{p}\right)^{-2} \log u_i - \sum_{j=1}^{\infty} \binom{-2}{j} \frac{(k-1)u_i^{\,j(1-k)} \log u_i}{1-k} \left(\frac{k}{p}\right)^{-2-j} \right|_{k=1} +$$

$$+ o(k-1) + \text{constant}$$

$$= \frac{k(k-1)}{p} \left[\left(\frac{k}{p}\right)^{-2} \log u_i + \sum_{j=1}^{\infty} \binom{-2}{j} \left(\frac{k}{p}\right) \text{Log}\, u_i + o\,(k-1) + \text{constant} \right]$$

Then the contribution of $D(x_j)$ to the overall statistic is

$$1/p\ [(1/p)^{-2} + (1+1/p)^{-2} - (1/p)^{-2}]\ \log u_i$$

$$= -\lambda^2(1-\lambda)\ X_i$$

Expanding $M_G[G(x_i)]$ about $k=1$, we find that its contribution is

$$\frac{(1-\lambda)\ p}{(p+1)^2}\quad \text{Log } u_i = (1-\lambda)^2\lambda\, x_i$$

Thus the total contribution of the x-sample is $-\lambda\,(1-\lambda)x_i$. But since $\hat{k}_{NR}-k$ is a function of $-D_2(x_i)$, then the contribution of the x-sample to

$\hat{k}_{NR}-k$ is $N/I_R(1)\ \lambda(1-\lambda)\ \bar{\bar{x}} = \bar{\bar{x}}$ when $u^{1-k} < k/p$.

When $u^{1-k} > \frac{k}{p}$, then

$$D_2(x_j) = \frac{k(k-1)}{p}\int_{u_o}^{u_i}\frac{u^{-1}d}{u^{1-k}+\frac{k}{p})^2} = \frac{k(k-1)}{p}\ \frac{u^{-1}\ du}{u^{2(1-k)}[1+\frac{k/p}{u^{1-k}}]^2}$$

$$= \frac{k(k-1)}{p}\int_{u_o}^{u_i}\frac{u^{2(k-1)-1}du}{[1+\frac{k/p}{u^{1-k}}]^2}$$

$$= \frac{k(k-1)}{p} \int_{u_o}^{u_i} u^{2(k-1)-1} \sum_{j=o}^{\infty} \binom{-2}{j} \left[\frac{k/p}{u^{1-k}}\right]^{-2-j} du$$

$$= \frac{k(k-1)}{p} \int_{u_o}^{u_i} \sum_{j=1}^{\infty} u^{j(k-1)-1} \binom{-2}{j} \left(\frac{k}{p}\right)^{-2-j} du$$

and the result follows from the previous case.

Y-sample: $u^{1-k} < \frac{k}{p}$

$$D_1(Y_i) = \frac{k-1}{kp} \int_{u_o}^{u_i} \sum_{j=1}^{\infty} \binom{-2}{j} \left(\frac{k}{p}\right)^{-2-j} u^{j(1-k)+1-2k}]\, du$$

$$= \frac{k-1}{kp} \int_{u_o}^{u_i} \left[\left(\frac{k}{p}\right)^{-2} u^{1-2k} + \sum_{j=1}^{\infty} \binom{-2}{j} \left(\frac{k}{p}\right)^{-2-j} u^{j(1-k)+1-2k}\right] du$$

$$= \frac{k-1}{kp} \left[\left(\frac{k}{p}\right)^{-2} \frac{u_i^{2(1-k)}}{2(1-k)} + \sum_{j=1}^{\infty} \binom{-2}{j} \left(\frac{k}{p}\right)^{-2-j} u_i^{(j+2)(1-k)} + \text{constant}\right]$$

$$= \frac{k-1}{kp} \left[\left(\frac{k}{p}\right)^{-2} \text{Log } u_i + \sum_{j=1}^{\infty} \binom{-2}{j} \left(\frac{k}{p}\right)^{-2-j} \text{Log } u_i + o\,(k-1) + \text{constant}\right]$$

Comparing this to the comparable expression for the contribution of the x-sample, we find that the Y-sample contributes $-\lambda(1-\lambda)Y_i$ to $D_1(Y_i)$ and thus $-\bar{Y}$ to $\hat{k}_{NR}$. Similarly, when $u^{1-k} >$, the result is the same. Thus

$$\lim_{k \to 1} \hat{k}_{NR} - k = \bar{x} - \bar{y} \text{ in probability.}$$

$$\sigma_{TR} = N \text{ Cov } [T_N(\hat{k}_{NR}) - \mu , \hat{k}_{NR} - k] =$$

$$N/(k-1). \ 1/I_R(k) \text{ Cov } [k(1-\lambda) \overline{D_1(Y_i) - ED_1(Y_i)} + \lambda \overline{M_F[F(Y_i)] - EM_F[F(Y_i)]}$$

$$- \lambda \overline{D_2(X_j) - ED_2(X_j)} - k(1-\lambda) \overline{M_G(G(X_i)] - EM_G[G(X_j)]} ,$$

$$(1-\lambda) \overline{(B_1(Y_i) - EB_1(Y_i)} - \overline{B_2(X_j) - EB_2(X_j))}]$$

where $1/I_R(k)$ = Fisher information at k from the rank likelihood. Letting $\lambda = 1/2$ for this investigation then σ_{TR} =

$$(1/2(k-1)) \ . \ (1/(I_R(k) \ \{\text{Cov } (D_2(X_j), B_2(X_j)) + \text{Cov } [M_F[F(Y_i)], \ B_1(Y_i)]$$

$$+ k \text{ Cov } [D_1(Y_i), B_1(Y_i)] + k\text{Cov } [M_G[G(X_j)], \ B_2(X_j)]$$

Each covariance term can be written in compact form using expected values of stochastic integrals. It is necessary to evaluate expressions of the form

$$E \int_0^\infty \int_0^\infty [G_1(s)-G(s)][G_1(t) - G(t)] \ Z_1(s) \ Z_2(t) \ dF(s)dF(t)$$

where $G_1(s)$ is the empirical distribution function for a sample of one from a poulation with cumulative distribution function G. Then the above expectation can be taken inside the integral when Fubing's theorem holds, (see Appendix).

$$E \int_0^\infty \int_0^\infty [G_1(s)-G(s)][G_1((t)-G(t)] \ Z_1(s) \ Z_2(t)dF(s)dF(t)$$

$$= \quad \{E[G_1(s)G_1(t)] - G(s)G(t)] \ Z_1(s) \ Z_2(t)dF(s)dF(t)$$

Now $E\,[G_1(s)G_2(t)] = E\,[I(X_j<s)\ I(x_j< t)]$

$$= Pr(x_j<\min(s,t)] = \min\,[G(s),\ G(t)]$$

Consequently, the required expectation equals

$$\int_0^\infty\int_0^\infty \{\min[G(s),G(t)] - G(s)G(t)\}\ Z_1(s)\ Z_2(t)dF(s)dF(t)$$

$$= \int\int_{-\infty<s<\ <\infty} G(s)[1-G(t)]\ Z_1(s)\ Z_2(t)dF(s)dF(t)$$

$$+ \int\int_{-\infty<t<s<\infty} G(t)[1-G(s)]\ Z_1(s)\ Z_2(t)dF(s)dF(t)$$

By symmetry of t and s we can write the second term as

$$\int\int_{-\infty<s<\ <\infty} G(s)[1-G(t)]\ Z_1(t)\ Z_2(s)dF(t)dF(s)$$

Finally we have

$$E \int^\infty\int^\infty [G_1(s)-G(s)][G_1(t)-G(t)]\ Z_1(s)\ Z_2(t)dF(s)dF(t)$$

$$= \int\int_{-\infty<s<t<\infty} G(s)[1-G(t)]\ \{Z_1(s)\ Z_2(t) + Z_1(t)\ Z_2(s)\}dF(s)dF\ (t).$$

Similarly:

$$E \int_0^\infty\int_0^\infty [F_1(s)-F(s)][F_1(t)-F(t)]\ Z_1(s)\ Z_2(t)\quad dG(s)dG(t) =$$

$$\int\int_{-\infty<s<t<\infty} F(s)[1-F(t)]\{Z_1(s)\ Z_2(t) + Z_1(t)Z_2(s)\}dG(s)dG(t).$$

Now $B_1(Y_i) = \int_{Y_o}^{Y_i} J'[H(s)]\,dG(s)$

Integrating by parts we have that

$$B_1(Y_i) - E B_1(Y_i) = \int_0^\infty B_1(y)\,d[F_1(y)-F(y)] = -\int_0^\infty [F_1(y)-F(y)]J'(H(y)dGy$$

Similarly

$$B_2(X_j) - E B_2(X_j) = -\int_0^\infty [G_1(s)-G(s)]\ J'[H(s)]dF(s)$$

$$D_1(Y_i) - E D_1(Y_i) = -\int_0^\infty [F_1(s)-F(s)]\ M_F'[F(s)]dG(s)$$

$$D_2(X_j) - E D_2(X_j) = -\int_0^\infty [G_1(s)-G(s)]\ M_G'[G(s)]dF(s)$$

$$M_F[F(Y_i)] - EM_F\{F(Y_i)] = -\int_0^\infty [F_1(s)-F(s)]\ M_F'[F(s)]dF(s)$$

$$M_G[G(X_j)] - EM_G[G(X_j)] = -\int_0^\infty [G_1(s)-G(s)]\ M_G'[G(s)]dG(s)$$

$\underline{\mathrm{Cov}\ [D_2(X_j), B_2(X_j)]} = \sigma_{D2B2}$

$$\iint_{-\infty<s<t<\infty} G(s)[1-G(t)]\ \{J'[H(s)]M_G'[G(t)] + J'[H(t)]M_G'[G(s)]\}dF(s)dF(t)$$

$\mathrm{Cov}\ \{M_F[F(Y_i)],\ B_1(Y_i)\} = \sigma_{FB1}$

$$\iint_{-\infty<s<t<\infty} F(s)[1-F(t)]\ J'[H(s)]M_F'[F(t)]dG(s)dF(t)$$

$$+ \iint_{-\infty<s<t<\infty} F(s)[1-F(t)]\ J'[H(t))]M_F'[F(s)]dG(t)dF(s)$$

But since $1-F(t) = [1-G(t)]^k$, $dF(t) = k[1-G(t)]^{k-1}dG(t)$

and so $\text{Cov}\ \{M_F[F(Y_i)],\ B_1(Y_i)] =$

$$k \iint_{-\infty< s < t < \infty} F(s)[1-F(t)]\ \{J'[H(s)]M_F'[F(t)][1-G(t)]^{k-1} + J'[H(t)]M_F'[F(s)][1-G(s)]^{k-1}\}dG(s)dG(t)$$

Similarly, $\text{Cov}\ [M_G[G(X_j)],\ B_2(X_j)] = \sigma_{GB2}$

$$k \iint_{-\infty< s < t < \infty} G(s)[1-G(t)]\ \{J'[H(s)]M_G'[G(t)]\ [1-G(s)]^{k-1} + J'[H(t)]M_G'[G(s)]\ [1-G(t)]^{k-1}\}dG(s)dG(t)$$

$\text{Cov}\ [D_1(Y_i),\ B_1(Y_i)] = \sigma_{D1B1}$

$$\iint_{-\infty< s < t < \infty} F(s)[1-F(t)]\ \{J'[H(s)]M_F'[F(t)] + J'[H(t)]M_F'[F(s)]\}dG(s)dG(t)$$

All terms for calculation of NVar $[T_N(\hat{k}_{NR})-\mu]$ have now been derived.

Thereom: Let $Y_i \sim \exp(k)$ then if Y_i is uncorrelated with the score function, the $B = 0$. Consequently NVar $T_{NR}(\hat{k}) = N\ \text{Var}(T_{NR}(k)\ \forall\ k > 0$.

PROOF: The result follows immediately from the result derived earlier, i.e., $B = \text{Cov}\ (Y_i,\ J(H(Y_i,k))$.

CONSISTENCY

In order to demonstrate the consistency of a test, we must show that the centering constant of the statistic under a fixed alternative is different from zero. In classical nonparametric theory there are useful theorems which provide sufficient conditions for consistency. However, in this problem, the centering constant will depend upon an estimate of the nuisance parameter k. Therefore, we do not expect to find simple sufficient conditions which make the proposed rank statistic consistent for all fixed alternatives to proportional hazards. Moreover, that goal is not really requireed for the purposes at hand. The ultimate goal of the pr... is to provide an estimate of the paramater $P(Y>x)$ which is not informative if the survival curves cross. The goal then becomes to identify a subclass of alternatives for which $\int_0^\infty J(H(x,k^*))\, dF(x) = 0$ where k^* is the limit in probability of the estimate $\hat{k}_{NR}$ of k and $\int_0^\infty \bar{F}(x) = 1/k$.

Recall that the transformed null hypothesis is that the two populations consist of exponential (1) and exponential (k) random variables. Under a fixed alternative, one population is exp(1) and the other is arbitrary with $\int_0^\infty \bar{F}(x)dx = 1/k$. The following lemma shows that for a broad class of alternativs, $k^* > k$.

LEMMA: Let $G(x) = 1-e^{-x}$. If $r(x) = \dfrac{f(x)}{\bar{F}(x)} > 1\ \forall x$, $r(o) < k$ and $r(x) = k$ only once, then $k^* > k$ when $\int_0^\infty \bar{F}(x)\, dx = 1/k$.

PROOF: The rank maximum likelihood equation is

$$\frac{1}{k^*} - \int_0^\infty \frac{\bar{F}_{n2}(x)}{[(1-\lambda)\bar{G}_{n2}(x) + \lambda k^* \bar{F}_{n1}(x)]}\, dH_N(x) = 0$$

We have already shown in this chapter that asympotically, this equation becomes

$$(1)\quad \frac{1}{k*} - \int_0^\infty \frac{\bar{F}(x)}{(1-\lambda)\bar{G}(x) + k^*\bar{F}(x)}\, dH(x) = o$$

Let $h(x) = (1-\lambda)\, e^{-x} + \lambda f(x)$

and $h^*(x) = (1-\lambda)\, e^{-x} + \lambda f^*(x)$

where $f^*(x) = k\bar{F}(x)$ so that $\int_0^\infty f^*(x) = 1$.

Since the LHS of (1) $= \left.\frac{\partial \log P(\underset{\sim}{R})}{\partial k}\right|_{k=k^*}$ under the assumption of proportional hazards and is non-increasing in k^*, then

$$k^* > k \Leftrightarrow \text{LHS} > o.$$

Rewriting (1), we have that $k^* > k \Leftrightarrow$

$$\frac{1}{k}\left[1 - \int_0^\infty \frac{k\,\bar{F}(x)h(x)}{h^*(x)}\right] > 0 \Leftrightarrow \frac{f^*(x)h(x)}{h^*(x)}\, dx < 1]$$

Now

$$\int_0^\infty \frac{f*(x)h(x)}{h^*(x)}\, dx = \int_0^\infty f^* dx + \int_0^\infty [h(x) - h^*(x)]\, \frac{f^*(x)}{h^*(x)}\, dx =$$

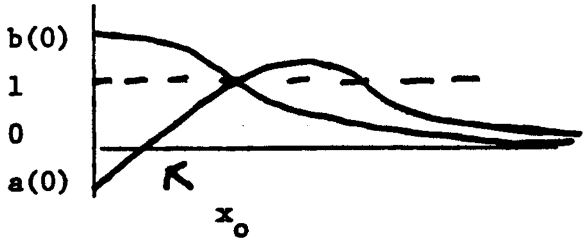
$$= 1 + \int_0^\infty [h(x) - h^*(x)]\, \frac{f^*(x)}{h^*(x)}\, dx$$

$$= 1 + \int_0^\infty a(x)\, b(x)\, dx$$

where $a(x) = h(x) - h^*(x) = \lambda\, (r(x)-k)\bar{F}(x)$, where $r(x) = f(x)/\bar{F}(x)$

$$b(x) = f^*(x)/h^*(x) = \frac{k}{(1-\lambda)\,\frac{\bar{G}(x)}{\bar{F}(x)} + \lambda k}$$

Note that since by the restriction of F, $r(0) < k$ and so $a(0) < 0$. As x increases, $a(x)$ increases to a maximum greater than zero and then decreases to zero in order that $\int_0^\infty a(x)dx = 0$. See Figure.



Let $S(x) = e^{-x}/\bar{F}(x)$. Then $S(x)$ is increasing in x

$$\Longleftrightarrow e^{-x}/\bar{F}(x)\ [r(x)-1] > 0$$

$\Longleftrightarrow r(x) > 1\ \forall x$. Thus if $r(x) > 1\ \forall x$, then $b(x)$ is monotone decreasing in x.

Define $x_o \ni a(x_o) = 0$. Thus $b(x) > 0\ \forall_x$ and $a(x) \leqslant 0$ for $x \leqslant x_o$.

Thus $\int_0^\infty a(x)b(x)dx < b(x_o)\ [\int_0^{x_o} a(x)dx + \int_{x_o}^\infty a(x)dx] = 0$

Thus $\int_0^\infty \frac{f^*(x)h(x)}{h^*(x)}dx < 1$ and so $k^* > k$. QED.

It is clear that the most identifiable class of distribution functions which satisfy the sufficency conditions are IFR distributions with $r(x) > 1$.

It is still possible to say something about the additional class F where $r(x) < 1$ and $r(x) = k$ only one. In this case, $b(x)$ is <u>not</u> monotone since there is a solution to the equation $r(x) = 1$. Let x_o be as before and define $x_1 \ni b(x_1) = b(0)$. Then for that subclass of distribution F such that $x_1 < x_o$, then as before $k^* > k$.

EXAMPLE: Weibull alternative.

$$F(x) = \exp[-(\kappa x)^{1+\theta}] \quad \theta > 0,\ \kappa > 0.$$

$$r(x) = (1+\theta)\kappa^{1+\theta} x^{\theta}. \quad \text{Then} \quad x_o^{\theta} = \kappa[(1+\theta)\kappa^{1+\theta}]^{-1}$$

$$b(x_1) = b(0) \iff (\kappa x_1)^{1+\theta} = x_1 \iff x_1^{\theta} = k^{-(1+\theta)}$$

Thus $x_1 < x_o \iff \kappa < 1\ \theta$. Thus for those Weibull distributions such that $\int_0^\infty \bar{F}(x)dx = 1/k$ and $\kappa < 1+\theta$, $k^* > k$.

THEOREM: Let $a(x,k^*) \equiv J(H(x,k^*))$ and WLOG , $Ea(X,k) = 0$ where the expectations is w.r.t. an exponential (k) cdf. If $\frac{\partial a(x,k^*)}{\partial x}$ is decreasing in x uniformly in k^*, then $T_N(\hat{k}_{NR})$ is consistent for all IFR distributions such that $k^* > k$. Proof:

$$\mu_{TN} = Ea(X,k^*) = \int_0^\infty a(x,k^*)dF(x) =$$

$$\int_0^\infty a(x,k^*)\ d\,[F(x) - e^{-k^* x}] = \int_0^\infty (\bar{F}(x) - e^{-k^* x}) \frac{\partial a}{\partial x}(x,k^*)dx$$

Since $F(x)$ is IFR and $E(x) = 1/k$, then by a well-known theorem in reliability theory, $\bar{F}(x) > e^{-kx}$ for $x < 1/k$ and

$\bar{F}(x) < e^{-kx}$ for some $x_1 > \frac{1}{k}$

Since $k^* > k$, then $\bar{F}(x) \geq e^{-k^* x}$ for $x \leq x_o$, say. Then if $\frac{\partial a(x,k^*)}{\partial x}$ is monotone decreasing in x, then

$$\mu_{TN} > \frac{\partial a}{\partial x}(x_o,k^*) \int_o^\infty [F(x)-e^{-k^* x}]dx$$

$$= \frac{\partial a}{\partial x}(x_o,k^*)\,[1/k - 1/k^*]$$

$$> 0 \iff k^* > k$$

QED.

Finally, the foregoing results on sufficient conditions for consistency show that the statistics on the pseudo-efficient scores and log x are consistent when the one groups hazard is always above or below the other's and the hazard ratio is increasing since each $J(H(x,k))$ has non-increasing first derivative. Also the logrank statistic is consistent since it is a monotone function of k when $k > 1$. Consistency follows for every distribution for which $k^* > k$ $k^* < k$ since

$$\mu_{TN}(k^*) = -\int_o^\infty \log[1-\lambda)(1-v)^{1/k^*} + \lambda(1-v)]dv = \mu_{TN}(k) \text{ only when }$$
$k^* = k$.

Scores for the Linear Rank Test

In the beginning of this Chapter, four conditions were given for the asymptotic normality of the Chernoff-Savage representation theorem. Hajek and Sidak (1967) show that if $a_N(i)=J(\frac{i}{N+1})$ then the first three conditions are satisfied.

Chernoff and Savage showed that they are also simplified $a_N(i) = E\,J(u_N^i)$, where u_N^i is the ith order statistic in a sample of size N from the uniform distribution. We show now that the former method is preferable when one attempts to construct "optimal" scores in this problem.

In Chapter I, we showed that there exists a monotone transformation of the original data which maps the random variable from one group to exponential random variables with scale parameter 1 and the random variables from the other group to exponential random variables with scale parameter k when H_o is true.

Although this particular transformation is arbitrary, it is convenient to think of an alternative to proportional hazards beings a departure from a mixture of (unobserved) exponentials. If $x_j \to x_j^*$ where $G(x^*) = 1-e^{-x^*}$, then for some alternative $Y_i \to Y_i^*$ where $F(y^*) = F(y^*, \kappa, \theta)$ where κ, θ are scale and shape parameters respectively. We would like these same scores to result in a statistics which is consistent against a broad class of fixed alternatives to the exponential with constant hazard k.

The locally most powerful rank test for a general alternative has been presented by Hajek and Sidak (1967):

Theorem: Let the family of densities $d(x,\Theta)$ $\Theta \epsilon J$ satisfy the following three conditions. Then the test with the critical region

$$\sum_{i=1}^{N} c_i a_n(R_i, d) \geq K$$

is the locally most powerful rank thest for H_o against $\{q_\Delta, \Delta > 0\}$ at the respective level, where $a_N(R_i,d) =$

$$E \left\{ \frac{d'(X_N^i, 0)}{d(X_N^i, 0)} \right\}$$

with X_N^i denoting the i-th order statistic from a sample of size N from the distribution with density $d(x,0)$.

(i) $d(x,\Theta)$ is absolutely continuous in Θ for almost every x.

(ii) the limit $d'(x,0) = \lim_{\Theta \to 0} 1/\Theta \, [d(x,\Theta) - d(x,0)]$ exists for almost every x.

(iii) $\lim_{\Theta \to 0} \int_{-\infty}^{+\infty} |d'(x,\Theta)| dx = \int_{-\infty}^{+\infty} |d'(x,0)| \, dx < \infty$

The production of the locally efficient score relies on the expansion:

$$Q_\Delta (\underset{\sim}{R} = \underset{\sim}{r}) = \int_{\underset{\sim}{R} = \underset{\sim}{r}} \cdots \int q_\Delta \, dx_1 \ldots dx_N \, .$$

$$\Delta \sum_{\ell=1}^{N} c_\ell \int_{\underset{\sim}{R} = \underset{\sim}{r}} \cdots \int \left[\frac{d'(x_\ell, 0)}{d(x_\ell, 0)} \prod_{i=1}^{N} d(x_i, 0) \right] dx_1 \ldots dx_N \, .$$

The score follow directly.

However, in the present problem, such a precedure leads to the following result.

$$\frac{Q_\Delta(\underset{\sim}{R} = \underset{\sim}{r})}{Q_0(\underset{\sim}{R} = \underset{\sim}{r})} = 1 + g(k)\Delta \sum_{\ell=1}^{N} c_\ell \int_{\underset{\sim}{R}=\underset{\sim}{r}} \cdots \int \frac{d_2'(y_\ell, 0)}{d_2(y_\ell, 0)} \prod_{i=1}^{n_1} d_2(y_i, 0) \times$$

$$\prod_{j=1}^{n_2} d_1(x_j)\, dy_1 \ldots dy_{n_1}\, dx_1 \ldots dx_{n_2}$$

There is no longer a simple score which is the expected value of a simple random variable. We conclude that under these circumstances, this classical approach is not feasible and some substitute must be found. For the purposes of this investigation, we use score functions of two kinds:

1) The log-rank function $J(H(x,k) = -\text{Log}\,[1-H(x,k)]$

2) "Pseudo-efficient" functions. These are the score functions based upon the log- likelihood under a local alternative, ignoring the function $g(k)$. Note that the function will depend on k in general, but usually in a simple manner.

3) Log(x), where x is a running variable for a random variable with exp(k) distribution. (See Appendix B for the justification of these scores for Chernoff-Savage theoren and conmputation of B, the adjustment coefficient.)

As for the computation of the score itself, we recall from the beginning of this section that the score can be regarded as the expected value of a function of an ith order statistic; however, a more convenient formulation is $a_N(i) = J(i/N+1)$.
The latter formulation is preferable in our case since, under the null hypothesis, we are not dealing with order statistics derived from an i.i.d. sample. We will see shortly that this method leads to scores which must be derived numerically and which depend upon our estimate of k under H_o.

Psuedo-Efficient Score for a Two-Parameter Sequence of Local Alternatives.

In Chapter I, we saw that a sequence of local alternatives H_N to proportional hazards is really a two dimensional sequence when the alternative cdf is parameterized by scale and shape parameters. Let Ω be the two-dimensional space of these parameters. Let one group be the x-sample, where x_j^*, $j = 1,\ldots n_2$ are exponential (1) and the Y-sample Y_i^*, $i = 1,\ldots,n_1$ are from a continuous density

$$f(x^*,\kappa,\theta) \quad \text{and} \quad F(x^*,k,\theta_o) = ke^{-kx^*}$$

The likelihood

$$LR(x_j^*,y_j^*,\kappa,k,\theta) = \frac{\prod_{j=1}^{n_2} e^{-x_j^*} \prod_{i=1}^{n_1} f(Y_i^*,\kappa,\theta)}{\prod_{j=1}^{n_2} e^{-x_j^*} \prod_{i=1}^{n_1} ke^{-kY_i^*}}$$

Since the x_j^*'s are exponential (1) under both null and alternative hypotheses, then the score can be thought of as being a function of the ranks of n_1 unobserved exponential (k) random variables amongst the total of N observations. Dropping the (*) for notational convenience, we find that

$$\log LR = \sum_{i=1}^{n_1} [\log F(y_i\kappa,\theta) - \log F(y_i,k,\theta_o0]$$

$$= \sum_{i=1}^{n_1} [(\kappa-k,\theta - \theta_o] \cdot \nabla \quad \text{Log } F(y_i \; \kappa,\theta) \Big|_{\substack{\kappa = k \\ \theta = \theta_o}}$$

$$+ \quad [[\kappa-k,\theta-\theta_o) \cdot \nabla]^2 \; \text{Log } F(y_i,k',\theta')\cdot$$

where

$$(k',\theta') = \varepsilon\,(k,\theta_o) + (1-\varepsilon)(\;,\theta) \quad \text{for some} \quad 0 < \varepsilon < 1.$$

Thus $\int_0^\infty \bar{F}(x)dx = \frac{e^\Theta}{\kappa} \int_0^1 u^\Theta e^{-\Theta u} du = \frac{1}{k}$

$<=> \quad \kappa = g(\Theta)k \quad$ where $\quad g(\Theta) = e^\Theta \int_0^1 u^\Theta e^{-\Theta u} du$

Let $\quad v = \Theta u \quad du = \frac{1}{\Theta} dv$

Then $\quad g(\Theta) = \frac{e^\Theta}{\Theta^{\Theta+1}} \int_0^\Theta d^\Theta e^{-v} dv$

Letting $(\kappa, \Theta) = (\kappa(t), \Theta(t)) = (\kappa(t), t) = \underset{\sim}{\alpha}(t)$,

then $\underset{\sim}{\alpha}'(t) = [kg'(t), 1]$ so $\underset{\sim}{\alpha}'(0) = [kg'(0), 1]$

We first show that $g'(0) = \lim_{t \to o} \frac{g(t)-g(0)}{t}$ exists.

1) $g(0) = \lim_{t \to o} g(t)$ exists.

$$\int_0^t x^t e^{-x} dx = \int_0^t x^t (1-x+ \frac{x^2}{2!} \ldots)dx =$$

$$\int_0^t x^t - x^{t+1} + \frac{x^{t+2}}{2!} - \ldots)dx =$$

$$\frac{t^{t+1}}{t+1} - \frac{t^{t+2}}{t+2} + \frac{t^{t+3}}{2!(t+3)} - \ldots = \sum_{i=1}^{\infty} (-1)^{i-1} \frac{t^{t+i}}{(i-1)!(t+i)}$$

thus $\quad g(t) = \frac{e^t}{t^{t+1}} \sum_{i=1}^{\infty} (-1)^{i-1} \frac{t^{t+i}}{(t+i)(i-1)!}$

$$= e^t \sum_{i=1}^{\infty} (-1)^{i-1} \frac{t^{i-1}}{(t+1)(i-1)!} = e^t \left[\frac{1}{t+1} + \text{Polynomial in } t\right]$$

$$\to 1 \text{ as } t \to o$$

thus
$$\lim_{t\to o} \frac{g(t)-g(0)}{t} = \lim_{t\to o} \frac{e^t \left[\frac{1}{t+1} + \sum_{i=2}^{\infty} (-1)^{i-1} \frac{t}{(t+i)(i-1)!} - 1\right]}{t}$$

$$\lim_{t\to o} \frac{e^t \left[-\frac{t}{t+1} + \sum_{i=2}^{\infty} (-1)^{i-1} \frac{t^{i-1}}{(t+i)(i-1)!}\right]}{t}$$

$$\lim_{t\to o} e^t \left[-\frac{1}{t+1} + \sum_{i=2}^{\infty} (-1)^{i-1} \frac{t^{i-2}}{(t+i)(i-1)!}\right]$$

$$\lim_{t\to o} \left[-\frac{1}{t+1} - \frac{1}{t+2} + \text{polynomial in } t\right]$$

$$\to -1 - \frac{1}{2} = -\frac{3}{2} \quad \text{as } T \to o.$$

For the Makeham distribution we find that

$$\nabla \operatorname{Log} f(\kappa,\theta) \Big|_{\substack{\kappa = k \\ \theta = 0}} = \left[\frac{1}{k} - x,\ 2(1-e^{-kx}) - kx\right]$$

thus the pseudo-efficient score is

$$-\frac{1}{2} k \left(\frac{1}{k} - x\right) + 2(1-e^{-kx}) - kx = 2(1-e^{-kx}) + \frac{kx}{2} - \frac{3}{2}$$

LINEAR HAARD:

$$\int_0^\infty \bar{F}(x)dx = \int_0^\infty e^{-\kappa x - \theta \frac{x^2}{2}} dx = \int_0^\infty e^{\frac{-(x+\frac{\kappa}{\theta})^2}{\sqrt{\frac{1}{\theta}}^{\,2}} + \frac{\kappa^2}{2\theta}} dx$$

Letting
$$z = \frac{x+\frac{\kappa}{\theta}}{\frac{1}{\sqrt{\theta}}} \Rightarrow dx = \frac{1}{\sqrt{\theta}} dx$$

thus
$$\int_0^\infty \bar{F}(x)dx = e^{\frac{\kappa^2}{2\theta}} \cdot \frac{1}{\sqrt{\theta}} \cdot \sqrt{2\pi} \int_{\frac{\kappa}{\sqrt{\theta}}}^{\infty} \frac{e^{\frac{-z^2}{2}}}{\sqrt{2\pi}} dz$$

$$= e^{\frac{\kappa^2}{2\theta}} \cdot \frac{1}{\sqrt{\theta}} \sqrt{2\pi} \left[1 - \Phi\left(\frac{\kappa}{\sqrt{\theta}}\right)\right]$$

where $\Phi(t)$ is the cdf of a standard normal random variable.

Let $\mu = \frac{\kappa}{\sqrt{\theta}}$. then we must have

$$\frac{1}{k} = e^{\frac{\mu^2}{2}} \cdot \frac{\mu}{\kappa} \sqrt{2\pi} \; [1-\Phi(\mu)]$$

or

$$\kappa = k\mu\sqrt{2\pi} \; e^{\frac{\mu^2}{2}} [1-\Phi(\mu)]$$

It is well-known that $\sqrt{2\pi} \; e^{\frac{\mu^2}{2}} [1-\Phi(\mu)] = R(\mu)$ where $R(\mu)$ is Mill's ratio where $R(\mu) = \frac{1}{\mu} - \frac{1}{\mu^3} + 0\left(\frac{1}{\mu^4}\right)$.

Thus

$$\mu R(\mu) = 1 - \frac{1}{\mu^2} + 0\left(\frac{1}{\mu^3}\right) = 1 - \frac{\theta}{\kappa^2} + \underset{\theta \to 0}{0}\left(\frac{\theta^{\frac{3}{2}}}{\kappa^3}\right)$$

In an arbitrarily smally neighborhood of $\theta = o$,

$$\frac{\kappa}{k} = 1 - \frac{\theta}{\kappa^2} \Rightarrow \theta = \kappa^2 - \frac{\kappa^3}{k}$$

$$\theta'(\kappa) = 2\kappa - \frac{3\kappa^2}{k}$$

$$\underset{\sim}{\alpha}(t) = [\kappa(t), \theta(t)] = [t, \theta(t)]$$

$\underset{\sim}{\alpha}(T) = [1, \theta'(t)]$. Since $\theta'(T) > 0$ for any small neighborhood $|\kappa - k| < \varepsilon \;\; \varepsilon > o$ then $\theta(\kappa)$ in κ and so the inverse function $\kappa = \theta^{-1}(s)$ exists in a neighborhood $0 < s < \varepsilon'$ ~

Perparameterizing, we can say that

$$\underset{\sim}{\alpha}(t) = \underset{\sim}{\alpha}^*(u) = [\kappa(u), \theta(u)] = [\kappa(u), u]$$

where

$$\kappa'(u)\Big|_{u=0} = [u'(\kappa)]^{-1}\Big|_{\kappa=k} = -\frac{1}{k} \; .$$

$$\nabla \operatorname{Log} f(\kappa,\theta)\Big|_{\substack{\kappa=k \\ \theta=0}} = \left[\frac{1}{k} - x, \frac{x}{k} - \frac{x^2}{2}\right]$$

Thus the pseudo-efficient score for a local sequence of linear hazards is

$$-\frac{x^2}{2}+\frac{2x}{k}-\frac{1}{k^2}\ .$$

A simple calculation show that

$$\int_0^\infty x\,\left(-\frac{x^2}{2}+\frac{2x}{k}-\frac{1}{k^2}\right)ke^{-kx}dx = 0,$$

so that for the statistic using this score, $B = 0$ and the adjustment is not needed.

It must be stressed that a "pseudo" efficient score is exactly that is several respects:

1) We ignore a coefficient which is a complicated function of k.

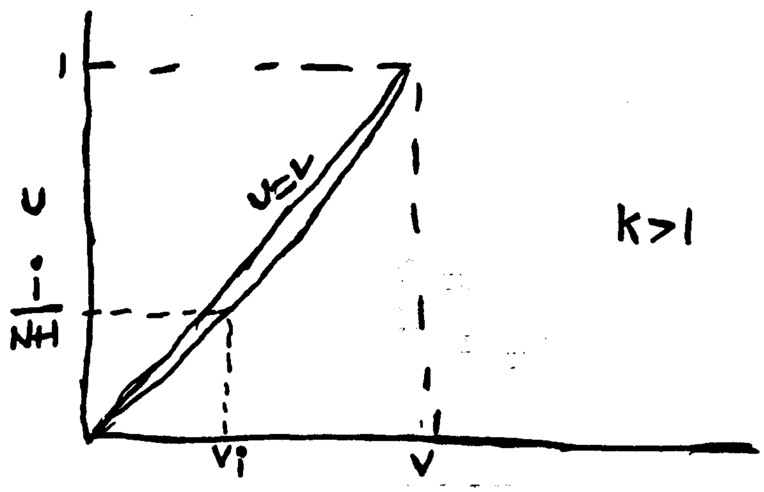
2) Even for the truly "efficient" score, the concept of Pittman efficiency does not apply to this problem. Pittman efficiency is meaningful when there is a parametric test with which to compare the competing test. Since the original null hypothesis is that of proportional hazards, the null hypothesis is actually an equivalence relation between distributions functions where $F \sim G \Leftrightarrow 1-F = (1-G)^k$, $k > 0$ and k can be different for difference pairs (F,G) in the equivalence relation.

3) Moreover, even for the truly efficient score, the comparison of the rank test to a parametric test to detect the departure of one of the populations from an exponential distribution would not make sense for the simple reason that there can be no parametric test which uses the joint likelihood of the sample. That is, the parametric version reduces to two independent one-sample problems.

COMPUTATION OF THE SCORES

Typically, rank scores depend upon the rank, only. That is, the score function is a simple function of $u = H(x)$: $J(u) = -\log(1-u)$, the "log-rank" score, or $J(u) = u$ for the Wilcoxon score. In this investigation, however, derivation of the pseudo-efficient scores reveals that scores depend upon the nuisance parameter k and Y_i^*, an unobserved random variable with the exponential (k) distribution. In order to make sense of the scores, then, we let $v = 1-e^{-kx}$ so that under H_o, $u = (1-\lambda)[1-(1-v)^{\frac{1}{k}}]+\lambda v$ or $1-u = (1-\lambda)(1-v)^{\frac{1}{k}} + \lambda(1-v)$.

Set $\frac{N-i+1}{N+1} = (1-\lambda)(1-v_i)^{\frac{1}{k}} + \lambda(1-v_i)$ for the ith ordered observation and solve for v_i. then $x_i = 1/k \, \mathrm{Log}(1-v_i)$ where the rank MLE for k is substituted for k. Note that we have suppressed the dependence of v_i on k. See graph below.



Obviously, numerical methods are required to compute the scores so that a score such as $-\log(\frac{N-i+1}{N+1})$ is preferred if its performance is comprable to these requiring numerical computation. Setting $u = \frac{i}{N+1}$ instead of $\frac{i}{N}$ is allowed since $\frac{i}{n} \equiv \frac{i_N}{N}$. As $N \to \infty$, $\frac{i_N}{N} \to p$ $0 < p < 1$, so that $\frac{i_N}{N+1} = \frac{i_N}{N} \cdot \frac{N}{N+1} \to p$ as $N \to \infty$.

PRELIMINARY TEST ESTIMATOR UNDER H_o

We are now interested in the bias and mean square error of the PTE when H_o is true. Recall that $\sqrt{N}(\hat{R}_{NR}-R)$, $\sqrt{N}(\hat{R}_{NW}-R)$ and $\sqrt{N}(T_N(\hat{k}_N) - \mu_N)$ are jointly normal by the multivariate central limit theorem. Thus let

$$z_{11} = \sqrt{N}(\hat{R}_{NR}-R), \quad z_{12} = \sqrt{N}(\hat{R}_{NW}-R) \qquad p_1 = \text{corr}\left(\frac{z_{11}}{\sigma_R}, z_2\right)$$

$$z_2 = \frac{\sqrt{N}(T_N(\hat{k}_N - \mu_N)}{\sigma_T} \qquad p_2 = \text{corr}\left(\frac{z_{12}}{\sigma_W}, z_2\right)$$

Let C_α be the critical value of the one-sided test in which we reject H_o if $\Pr(z_2 \geq C_\alpha | H_o) < \alpha$. Then the expected value of the PTE is defined as

$$E(PTE) = \int_{-\infty}^{C_\alpha} \int_{-\infty}^{\infty} z_{11} f_1(z_{11},z_2)dz_{11}\, dz_2 + \int_{C_\alpha}^{\infty} \int_{-\infty}^{+\infty} z_{12} f_2(z_{12},z_2)dz_{12}dz_2$$

where

$$f_{11}: (z_{11},z_2) \sim N\begin{bmatrix} 0 & \sigma_R^2 & \sigma_{RT} \\ 0 & ;\sigma_{RT} & 1 \end{bmatrix}$$

$$f_{22}: (z_{12},z_2) \sim N\begin{bmatrix} 0 & \sigma_W^2 & \sigma_{WT} \\ 0 & ;\sigma_{WT} & 1 \end{bmatrix}$$

We can immediately write down the asymptotic covariances

$$\sigma_{RT} = \frac{N \text{ cov}[T_N(\hat{k}_{NR}), \hat{R}_{NR}-R]}{\sigma_T} = \frac{N \text{ cov}[T_{NR}(k)\hat{R}_{NR} - R] + B\, N \text{ cov}[\hat{k}_{NR}-k, \hat{R}_{NR}-R]}{\sigma_T}$$

where $\hat{R}_{NR}-R = -\frac{1}{(k+1)^2}(\hat{k}_{NR}-k)$ using the standard Taylor's expansion

$$N \operatorname{cov}[\, T_N(\hat{k}_{NR}), \hat{R}_{NW}-R] .$$

We can split to the contributions to the covariance into two parts:

a. $N \operatorname{cov}[T_N(k), \hat{R}_{NW}-R)]$

b. $BN \operatorname{cov}[\hat{k}_{NR}-k, \hat{R}_{NW}-R]$.

a) We first express $\sqrt{N}(\hat{R}_{NW}-R)$ as a sum of independent stochastic integrals. Using the expansion for the generalized U-statistic for

$$\hat{R}_{NW} = \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\phi(X_j, Y_i) \qquad \text{where}$$

$$\phi(X_j, Y_i) = \begin{cases} 1 & \text{if } Y_i > X_j \\ 0 & \text{otherwise} \end{cases},$$

we have

$$\phi(Y_i, X_j) - R = \phi_{10}(X_j) - R \quad + \phi_{01}(Y_i) - R$$

$$+ \quad \phi(Y_i X_j) - \phi_{10}(X_j) - \phi_{01}(Y_i) + R$$

where

$$\phi_{10}(x) = E\ [(Y_i, X_j) | X_j = x] = 1 - F(x)$$

$$\phi_{01}(y) = E\ [(Y_i, X_j) | Y_i = y] = G(y)$$

we recall that $F(x) = 1-e^{-kx}$, $G(y) = 1-e^{-y}$.

Write $\sqrt{N}(\hat{R}_{NW}-R) = \frac{N}{n_2}\sum_{j=1}^{n_2}\phi_{10}(X_j) - R \quad + \frac{N}{n_1}\sum_{i=1}^{n_1}\phi_{01}(Y_i) - R$

$$+ \frac{\sqrt{N}}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left[\phi(Y_i, X_j) - \phi_{10}(X_j) - \phi_{01}(Y_i) + R\right]$$

where the last term is $o_p(\frac{1}{N})$.

$$T_N(k) = \text{constant} + (1-\lambda)\left[\frac{1}{n_1}\sum_{i=1}^{n_1} B(Y_i) - E\,B(Y_i) \;-\; \frac{1}{n_2}\sum_{j=1}^{n_2} B(X_j) - B(X_j) \right]$$

where

$$B_1(Y_i) = -\int_0^\infty [\,F_1(y)-F(y)\,]\,J'(H(y))\,dG(y)$$

$$B_2(x_j) = -\int_0^\infty [\,G_1(x)-G(x)\,]\,J'\,(H(x))\,dF(x)$$

Now for $\phi_{10}(X_j)$ we write $C_2(X_j) = 1-F(X_j) = 1-\int_0^{X_j} dF(x)$, so that

$C_2(X_j) - E\,C_2(X_j) = \int_0^\infty [\,G_1(x)-G(x)\,]\,dF(x)$ using integration by parts. Also let

$$C_1(Y_i) = G(Y_i) = \int_0^{Y_i} dG(y)$$

Thus $C_1(Y_i) - E\,C_1(Y_i) = -\int_0^\infty [\,F_1(y)-F(y)\,]\,dG(y)$

thus $$\sqrt{N}\,(\hat{R}_{NW}-R) = \frac{\sqrt{N}}{n_2}\sum_{j=1}^{n_2} C_2(X_j) - E\,C_2(X_j) + \frac{\sqrt{N}}{n_1}\sum_{i=1}^{n_1}[\,C_1(Y_i)-EC_1(Y_i)\,] + o_p(\tfrac{1}{N})$$

$$\lim_{N\to\infty} N\,\mathrm{cov}[\,T_N(k), \hat{R}_{NW}-R = (1-\lambda)\left[-\frac{1}{1-\lambda}\,\mathrm{cov}\;C_2(X_j), B_2(X_j) + \frac{1}{\lambda}\,\mathrm{cov}[\,C_1(Y_i), B_1(Y_i)\,]\right]$$

$$= -\,\mathrm{cov}[\,C_2(X_j), B_2(X_j)\,] + \frac{1-\lambda}{\lambda}\,\mathrm{cov}\;C_1(Y_i), B_1(Y_i)\,.$$

$$\mathrm{Cov}\;[\,C_2(X_j), B_2(X_j)\,] = -\int_0^\infty\!\!\int_0^\infty [\min(G(s),G(t)) - G(s)G(t)]\;J'\,H(s)\;(dF(s))dF(t)$$

$$\mathrm{Cov}\;[\,C_1(Y_i), B_1(Y_i)\,] = \int_0^\infty\!\!\int_0^\infty [(\min F(s),F(t)) - F(s)F(t)\,]\,J'\,[H(s)]\;dG(s)dG(t)$$

Finally,

$$N\,\mathrm{Cov}\,[\,T_N(k), \hat{R}_{NW}-R] = \iint_{-\infty<s<t<\infty} G(s)(1-G(t))\{J'[H(s)] + J'[H(t)]\}\,dF(s)dF(t)$$

$$+ \iint_{-\infty<s<t<\infty} F(s)\,(1-F(t))\,\{J'[H(s)] + J'[H(t)]\}\,dG(s)dG(t).$$

b) Using the expression for $\hat{k}_{NR}-k$ in terms of D- functions and M-functions, we can write

$$N \text{ cov } [\hat{k}_{NR}-k, \hat{R}_{NW}-R]$$

$$\frac{1}{k-1} \sigma_R^2 [k \cdot \frac{1-\lambda}{\lambda} \text{ Cov } (D_1(Y_i), C_1(Y_i)) + \text{cov } (M_F(F(Y_i)), C_1(Y_i))$$

$$- \frac{\lambda}{1-\lambda} \text{ cov } (D_2(X_j), C_2(X_j)) - k\left[\text{cov } (M_G\, G(X_j)\ ,\ C_2(X_j))\right.$$

thus

$$\sigma_{D1C1} = \text{cov } (D_1(Y_i), C_1(Y_i)) = \int_0^\infty\int_0^\infty \left[\min(F(s)), F(t) - F(s)F(t)\right] M_F'\left[F(s)\right] dG(s)dG(t)$$

$$\sigma_{FC_1} = \text{cov } \left[M_F\, F(Y_i), C_1(Y_i)\right] = \int_0^\infty\int_0^\infty\left[\min\, (F(s)), F(t) - F(s)F(t)\right] M'_F\left[F(s)\right] dF(s)dG(t)$$

$$\sigma_{D_2C_2} = \text{cov } \ D_2(x_j), C_2(x_j) \ = - \int_0^\infty\int_0^\infty\left[\min(G(s), G(t)) - G(s)G(t)\right] M'_G\left[G(s)\right] dF(s)dF(t)$$

$$\sigma_{GC_2} = \text{cov } M_G(G(x_j)),\ C_2(x_j) \ = - \int_0^\infty\int_0^\infty\left[\min(G(s), G(t)) - G(s)G(t)\right] M_G'\left[G(s)\right] dG(s)dF(t)$$

Thus when $\lambda = \frac{1}{2}$,

$$N \text{ cov}\left[\hat{k}_{NR}-k, \hat{R}_{NW}-R\right] = \frac{\sigma_R^2}{k-1} \left(k\sigma_{D1C1} + \sigma_{FC1} - \sigma_{D2C2} - k\sigma_{GC2}\right)$$

ASYMPTOTIC BIAS OF PTE UNDER H_o

After accounting for the two correlations, we can now rewrite E (PTE) as:

$$\int_{-\infty}^{C_\alpha} \left[\int_{-\infty}^{+\infty} z_{11} f_{11}(z_{11}|z_2)dz_{11}\right] f_2(z_2)dz_2 + \int_{C_\alpha}^{\infty} \left[\int_{-\infty}^{+\infty} z_{12} f_{12}(z_{12}|z_2)dz_{12}\right] f_2(z_2)dz_2$$

$$= \int_{-\infty}^{C_\alpha} E(z_{11}|z_2) f_2(z_2)dz_2 + \int_{C_\alpha}^{\infty} E(z_{12}(z_2)\ f_2(z_2)dz_2$$

$E(z_2) = E(z_{12}) = 0$ under H_o, then $E(PTE) =$

$$\int_{-\infty}^{C_\alpha} \left[E(z_{11}|z_2) - E\ (z_{12}|z_2)\right] f_2(z_2)dz_2$$

$$= \int_{-\infty}^{C_\alpha} (\sigma_R p_1 z_2 - \sigma_W p_2 z_2)\ f_2(z_2)dz_2$$

$$= (p_1\sigma_R - p_2\sigma_W) \int_{-\infty}^{C_\alpha} z_2 f_2(z_2)dz_2$$

$$= \frac{(p_2\sigma_W - p_1\sigma_R)}{\sqrt{2\pi}} \exp - \left[\frac{C_\alpha^2}{2}\right]$$

Since the centering constants of both z_{11}, and z_{12} are zero under H_o, the PTE will be positively or negatively biased if

$$\frac{p_2}{p_1} > \frac{\sigma_R}{\sigma_w} \text{ or } \frac{p_2}{p_1} < \frac{\sigma_R}{\sigma_w}$$

respectively. But this is true if and only if

$$\sigma_{WT} > \sigma_{RT} \text{ or } \sigma_{WT} < \sigma_{RT}, \text{ respectively.}$$

ASYMPTOTIC VARIANCE OF PTE UNDER H_o

$$\text{Var}(\text{PTE}) = \int_{\infty}^{C_\alpha}\int_{-\infty}^{+\infty} z_{11}^2 f_{11}(z_{11},z_2)dz_{11}dz_2 + \int_{C_\alpha}^{\infty}\int_{-\infty}^{+\infty} z_{12}^2 f_{12}(z_{12},z_2)dz_{11}dz_2$$

$$= \int_{-\infty}^{C_\alpha} \left[E(z_{11}^2|z_2) - E(z_{12}^2|z_2)\right] f_2(z_2)dz_2 + \sigma_w^2$$

since $E(z_{12}^2) = \sigma_w^2$ under H_o.

Now $z_{11}|z_2 \sim N\left[\sigma_R p_1 z_2,\ \sigma_R^2(1-p_1^2)\right]$

$z_{12}|z_2 \sim N\left[\sigma_W p_2 z_2,\ \sigma_W^2(1-p_2^2)\right]$

thus

$$\text{Var}(\text{PTF}_2) = \int_{-\infty}^{C_\alpha} \{\sigma_R^2 (p_1^2 z_2^2 + 1-p_1^2) - \sigma_w^2(p_2^2 z_2^2+1-p_2^2)\}$$

$$\times\ f_2(z_2)dz_2 + \sigma_w^2$$

$$= (\sigma_R^2 p_1^2-\sigma_W^2 p_2^2) \int_{-\infty}^{C_\alpha} z_2^2 f_2(z_2)dz_2 + \left[\sigma_R^2(1-p_1^2) - \sigma_W^2(1-p_2^2)\right] \Phi(C_\alpha) + \sigma_w^2$$

$$= (\sigma_R^2 p_1^2-\sigma_w^2 p_2^2) \quad -\left[\frac{C_\alpha \exp}{\sqrt{2\pi}} \frac{(-C_\alpha^2)}{} + \Phi(C_\alpha)\right]$$

$$- (\sigma_R^2 p_1^2 - \sigma_W^2 p_2^2) \Phi (C_\alpha) + (\sigma_R^2 - \sigma_W^2) \Phi (C_\alpha) + \sigma_W^2$$

$$= C_\alpha \frac{(\sigma_W^2 p_2^2 - \sigma_R^2 p_1^2)}{\sqrt{2\pi}} \exp (- \frac{C_\alpha^2}{2}) + (\sigma_R^2 - \sigma^2) \Phi (C_\alpha) + \sigma^2 .$$

In Chapter III we have seen that the problem demands a restriction on the alternative hypothesis for problems to be well defined. The one chosen in this study, that the (fictitious) random variable associated with the ratio of the two hazard functions have the same expectation under the null and alternative hypothesis, is somewhat arbitrary, but has the heuristic advantage of signifying some sort of correspondence of the null and alternative through some sort of measure of central tendency. One would not, for instance, anticipate radicaly different results if the medians were set equal to each other rather than the means.

We have also seen that reasonable, sufficient conditions on the profile of the alternative and the nature of the scores can be found so that consistency is assured under the restriction pointed out in the previous paragraph. Further, the preferred test (locally most powerful) is not appropriate for several reasons and so we must make our choice of scores for the linear rank statistic in a somewhat arbitrary way. Certainly simplicity is a virture due to the potentially complicated means of obtaining the scores numerically.

Lastly, the necessary quantities needed for finding the asymptotic joint distribution of the test statistic and the two estimators are made possible by an expansion of the MLE equation of $\hat{k}$. Simple formulas for the PTE bias and variance under the null hypothesis are available due to the asymptotic joint normality of the required statistics.

CHAPTER IV

INTRODUCTION

In previous chapters we have examined the performance of the PTE when H_o is true and for a fixed alternative. Since the outcome in the latter case is trivial in the sense that one would use the Wilcoxon estimator with probability approaching one, we are led to alternatives between H_o and a fixed alternative, close to H_o. The sequences of local alternatives will be equivalent to one group being exp(1) under both H_o and H_{NA} and the other group having a distribution which converges in law to exp(k). As pointed out in Chapter III this sequence can be parameterized by a 2-dimensional parameter space (κ_N, θ_N) restricted along a curve in that space.

In this chapter, we derive the properties of the two estimators and the test statistic under local alternatives. Then we assess the performance of the PTE under these circumstances.

THE RANK MLE UNDER H_{NA}

In Chapter II, it was easy to compute the rank likelihood of the data under the assumption of proportional hazards. Following the strategy adopted here to represent fixed alternative to proportional hazards, we would attempt to evaluate an N-dimensional integral where F(x) would not be exp(k). Clearly, this is an intractible task. Fortunately, this is not necessary when examining behavior of $\sqrt{N}\ (\hat{R}_{NR}-R)$ under local alternatives. There are two ways to go in this situation:

I) We recognize that log likelihood of the ranks derived under H_o will still coverge to some number different from zero for a fixed alternative.

Thus e^{-kx} is replaced by $\bar{F}(x,\kappa,\theta)$. One can then make a Taylor expansion of the constant term around the point (k,θ_o) to compute the centering constant for $\sqrt{N}(\hat{k}_{NR}-k)$. Then, arguing that the remainder terms are still $o_p(1)$ by contiguity of $f(x,\kappa_N,\theta_N)$ to ke^{-kx}, the asymptotic normality of $\sqrt{N}(\hat{k}_{NR}-k)$ is shown under a local alternative. This strategy is somewhat clumsy and is not particularly appealing because the method does not fit into a more general theory for deriving non-centrality parameters. In fact, one can see that this method is extremely cumbersome when it comes to calculating the non-centrality parameters for $\sqrt{N}(T_N(\hat{k}_{NR})-\mu_N)$. There is a much more unified context for this problem by a lemma due to LeCam (1960): Let S_N be a statistic.

LEMMA: If the pair $(S_N, \text{Log}L_N)$ is under P_N asymptotically normal $(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\sigma_{12})$ with $\mu_2 = -\frac{1}{2}\sigma_2^2$, then S_N is under Q_N asymptotically normal $(\mu_1+\sigma_{12},\sigma_1^2)$, where P_N, θ_N are probability measures induced by densities p_N and q_N.

Note that this lemma obviates the need for Taylor expansion of a constant term that is itself part of a Taylor expansion.

NON-CENTRALITY PARAMETER FOR $\sqrt{N}(\hat{R}_{NR}-R)$ UNDER H_{NA}

In Chapter III we expressed the log likelihood for a local alternative in terms of the directional derivative (stochastic) plus a term which converges to a constant in probability. It is clear from that expression that $\text{Log}L \sim N(-\frac{1}{2}\sigma_2^2, \sigma_2^2)$ since the constant is simply the Fisher Information evaluated at a point along a curve, and the directional derivative is a sum of i.i.d, random variables:

$$\text{Let} \quad \text{Log } L = \frac{\delta}{\sqrt{N}}\sum_{i=1}^{n_1} L(Y_i) = \frac{\delta n_1 \overline{L(Y_i)}}{\sqrt{N}}$$

We can express the stochastic part of $\sqrt{N}(\hat{k}_{NR}-k)$ which has non-zero covariance with LogL as

$$\frac{\sqrt{N}}{(k-1)I_R(k)} \quad k(1-\lambda)\,\overline{D_1\ (Y_i)} \ + \lambda M_F \left[F(Y_i)\right] \ + o_p(1)$$

Thus the non-centrality parameter c_{11} of $\sqrt{N}(\hat{R}_{NR}-R)$ =

$$-\frac{\delta}{(k+1)^2(k-1)I_R(k)} \quad \left[k(1-\lambda)\right] \text{cov}\ (L_i, D_{1i}) + \gamma\ \text{cov}\ (L_i,\ M_{Fi})$$

Since $(\hat{R}_{NR}-R) = -\frac{1}{(k+1)^2}\ (\hat{k}_{NR}-k)$,

$$L_i - EL_i = \int_0^\infty L(y)\ d\left[F_1(y) - F(y)\right] = -\int_0^\infty \left[F_1(y) - F(y)\right] L'(y)dy$$

$$D_{1i} - ED_{1i} = \int_0^\infty D_1(y)d\left[F_1(y) - F(y)\right] = -\int_0^\infty \left[F_1(y) - F(y)\right] M_F'\left[F(y)\right]dG(y)$$

$$M_{Fi} - EM_{Fi} = \int_0^\infty M_F\left[F(y)\right]\ d\left[F_1(y) - F(y)\right] = -\int_0^\infty \left[F_1(y)-F(y)\right] M_F'\left[F(y)\right] dF(y)$$

Thus

$$_{11} = -\frac{\delta}{(k+1)^2(k-1)I_R(k)} \Big\{k(1-\lambda) \iint_{-\infty<s<t<\infty} F(s)\left[1-F(t)\right]\Big\{L'(s)M'\left[F(t)\right] G'(t)$$

$$+ L'(t)M'\left[F(s)\right] G'(s)\Big\}\ dsdt$$

$$+ \lambda \iint_{-\infty<s<t<\infty} F(s)\ \left[1-F(t)\right]\Big\{L'(s)M'\left[F(t)\right] F'(t) +$$

$$L'(t)M'\left[F(s)\right] F'(s)\Big\}\ dsdt\Big\}$$

NON-CENTRALITY PARAMETER FOR $\frac{\sqrt{N}T_N(\hat{k}_{NR}-\mu)}{\sigma_T}$ UNDER HA.

Expressing the relevant stochastic part of $\frac{\sqrt{N}\ (T_N(\hat{k}_{NR}-\mu)}{\sigma_T}$:

$$\frac{\sqrt{N}(1-\lambda)\ \overline{B_1(Y_i)}}{\sigma_T} + \frac{B\ \ N}{\sigma_T(k-1)I_R(k)} \left[k(1-\lambda)\ \overline{D_1(Y_i)} + \lambda\ \overline{M_F\ F(Y_i)}\right] + o_p(1)\ ,$$

we find $C_2 = \frac{\delta}{\sigma_T}\left[(1-\lambda)\ \mathrm{cov}\ (L_i, B_{1i}) + \frac{B}{(k-1)I_R(k)}\right] x$

$$\left[k(1-\lambda)\ \mathrm{cov}\ (L_i, D_{1i}) + \lambda\ \mathrm{cov}\left[L_i, MF_i\right]\right]$$

where $\mathrm{cov}\ (L_i, B_{1i}) = \iint\limits_{-\infty<s<t<\infty} F(s)\left[1-F(t)\right]\left\{L'(s)J'(H(t))G'(t)\right.$

$$\left. + L'(t)J'(H(s))\ G'(s)\right\}\ dsdt$$

NON-CENTRALITY PARAMETER FOR $\sqrt{N}(\hat{R}_{NW} - R)$

The relevant stochastic part of $\sqrt{N}(\hat{R}_N - R)$ is

$$\frac{\sqrt{N}}{n_1}\sum_{i=1}^{n_1} C_1(Y_i) - EC_1(Y_i) \qquad \text{where}$$

$$C_1(Y_i) - EC_1(Y_i) = -\int_0^\infty \left[F_1(y) - F(y)\right] dG(y)$$

thus $C_{12} = \delta\ \mathrm{cov}\ Li,\ C_{1i} =$

$$\delta \iint\limits_{-\infty<s<t<\infty} F(s)\left[1-F(t)\right]\left[L'(s)\ G'(t) + L'(t)G'(s)\right]\ dsdt$$

SUMMARY

As in Chapter III, we note that the vector of the three relevant random variables consists of components which are the sum of independent random variables with finite second moment. Thus, by the multivariate central limit thereon, the vector is asymptotically jointly normal and we may use LeCan's third Lemma. By contiguity, the variances of each random variable remain the same (asymptotically) as they were under H_o.

Thus the only differences in the behavior of these three asymptotic statistics from H_o to H_{NA} are the centering constants (shifts). Consequently the covariances of each estimator with the test statistic are asymptotically the same under H_{NA} as they were under H_o. We are now able to express the bias and mean square error of the PTE under H_{NA} in terms similar to those used under H_o.

ASYMPTOTIC BIAS OF PTE UNDER H_{NA}

Let z_{11}, z_{12}, z_2 be defined as under H_o. However, under H_{NA},

$$f_{11}: (z_{11}, z_2) \sim N \begin{bmatrix} c_{11} & \sigma_R^2 & \sigma_{RT} \\ c_2 & ; \sigma_{RT} & 1 \end{bmatrix}$$

$$f_{12}: (z_{12}, z_2) \sim N \begin{bmatrix} c_{12} & \sigma_W^2 & \sigma_{WT} \\ & ; & \\ c_2 & \sigma_{WT} & 1 \end{bmatrix}$$

Again, $E(\text{ PTE}) = \int_{\infty}^{c_\alpha} E(z_{11}|z_2)\ f_2(z_2)dz_2 + \int_{c_\alpha}^{\infty} E\ (z_{12}|z_2)f_2(z_2)dz_2$

Since $E(z_{12}) = C_{12}$, then

$$E(\text{ PTE}) = \int_{\infty}^{C_\alpha} [\ E(z_{11}|z_2) - E(z_{12}|z_2)]\ f_2(z_2)dz_2 + C_{12}.$$

$$z_{11}|z_2 \sim N\ [C_{11} + \sigma_R\ p_1(z_2-C_2)\ ;\ \sigma_R^2(1-p_1^{\ 2})]$$

$$z_{12}|z_2 \sim N\ [C_{12} + \sigma_W p_2\ (z_2-C_2),\ \sigma_W^{\ 2}(1-p_2^{\ 2})]$$

$$E(\text{PTE }) = \int_{-\infty}^{C_\alpha}(C_{11}-C_{12}) + (\sigma_R p_1 - \sigma_W p_2)(z_2 - C_2)\ \ f_2(z_2)dz_2 + C_{12}$$

$$= (C_{11}-C_{12})\ \Phi\ (C_\alpha-C_2) + \frac{(\sigma_W p_2-\sigma_R p_1)}{2\pi}\ \exp\left[-\ \frac{(C_\alpha-C_2)^2}{2}\right] + C_{12}\ .$$

Thus $\text{BIAS}(\text{PTE}) = C_{11}-C_{12}) \; \Phi \; (C_\alpha - C_2) + \frac{\sigma_w p_2 - \sigma_R p_1}{\sqrt{2\pi}} \exp\left[-\frac{(C_2-C_2)^2}{2}\right] + C_{12}$

$$- \lim_{\substack{N\to\infty \\ (K_N,\theta_M)\,\varepsilon\,\underset{\sim}{\alpha}(t)}} \sqrt{N} \left\{ \int_{-\infty}^{+\infty} \left[1-F(x,K_N,\theta_N)\right] dG(x) - \frac{1}{1+K} \right\}$$

where the last term : the shift in the parameter $\sqrt{N}\left[P_N(Y>X) - \frac{1}{1+K}\right]$ for a sequence of local alternatives to proportional hazards.

VARIANCE OF PTE UNDER H_{NA}

We first calculate Var(PTE).

Let $z_{11}^* = z_{11} - C_{11}$, $z_{12}^* = z_{12} - C_{12}$, so $E(z_{11}^*) = E(z_{12}^*) = 0$

Then $\text{Var}(\text{PTE}) = \int_{-\infty}^{C_\alpha} z_{11}^{*2} f_{11}^* (z_{11}^*, z_2) dz_{11}^* dz_2 + \int_{C_\alpha}^{\infty} z_{12}^{*2} f_{12}^* (z_{12}^*, z_2) dz_{12}^* dz_2$

Now $z_{11}^* | z_2 \sim N\left[\sigma_R p_1 (z_2 - C_2),\; \sigma_R^2 (1-p_1^2)\right]$

$z_{12}^* | z_2 \sim N\left[\sigma_W p_2 (z_2 - C_2),\; \sigma_W^2 (1-p_2^2)\right]$

thus $\text{Var}(\text{PTE}) = \int_{-\infty}^{C_\alpha} E(z_{11}^{*2} | z_2) f_2(z_2) dz_2 + \int_{C_\alpha}^{\infty} E(z_{12}^{*2} | z_2) f_2(z_2) dz_2$

$$= \int_{-\infty}^{C_\alpha} \left[E(z_{11}^{*2}|z_2) - E(z_{12}^{*2}|z_2)\right] f_2(z_2) dz_2 + \sigma_w^2$$

Since $E(z_{12}^{*2}) = \sigma_W^2$

$\text{Var}(\text{PTE}) = \int_{-\infty}^{C_\alpha} \{\sigma_R^2 \left[(p_1^2 (z_2 - C_2)^2 + 1 - p_1^2\right] - \sigma_W^2 \left[p_2^2 (z_2 - C_2)^2 + \right.$

$\left. + 1 - p_2^2\right] \} f_2(z_2) dz_2 + \sigma_W^2 =$

$$(\sigma_W^2 p_2^2 - \sigma_R^2 p_1^2) \frac{(C_\alpha - C_2)}{\sqrt{2\pi}} \exp\left[-\frac{(C_\alpha - C_2)^2}{2}\right] + (\sigma_R^2 - \sigma_W^2)\, \Phi\, (C_\alpha - C_2) + \sigma_w^2.$$

In this chapter we have seen that by slight modifications of the PTE bias and variance formulas developed in the last chapters we can produce formulas for those quantities under specified local alternatives. Previous chapters have developed expressions for the stochastic behavior of the test statistic and the two estimators. Using LeCam's third Lemma, it is then possible to calculate the required non-centrality parameters for the three relevant statistics.

CHAPTER V

Numerical integration of relevant expressions was performed by IMSL programs DCADRE and DBLIN. The parameters which measure the performance of the PTF_2 are defined as follows:

BIAS: The bias of the PTE for the parameter $\sqrt{N}\left(R_{NA} - \frac{1}{1+k}\right)$.

BIASR: The comparable bias of the rank estimator.

BIASW: The comparable bias of the Wilcoxon estimator.

RTMSR: The ratio of the mean square error of the PTE to σ_R^2.

RTMSW: The ratio of the mean square error of the PTE to σ_W^2.

For the test statistic, itself, we calculate the standardized non-centrality parameter C_2.

Table 1 gives the combination of scores and local alternatives which were used in this study:

TABLE 1

SCORE	LOCAL ALTERNATIVE
Log x	Linear Hazard Makeham
Pseudo-efficient linear hazard	Linear Hazard
Pseudo-efficient Makeham	Makeham
Savage Score (Log-Rank)	Linear Hazard Makeham
Wilcoxon Score	Linear Hazard Makeham

The values of C_2 used were $C_2 = .84, 1.282, 1.645$ corresponding to one-sided tests of level .20, .10, .05 respectively.

THE PTE UNDER LOCAL ALTERNATIVES

It is clear from Tables 2 through 9 that the biases and ratios of mean square error to variance vary little over the range of $k = 1.1$ to $k = 1.9$. There is more variability in the non-centrality parameters. Not surprisingly, the absolute value of the non-centrality parameter decreases as k increases, since, in that case, the relationship of the groups moves further away from the state of homogeneity in which $\hat{k}_{NR}$ is asymptotically efficient for k.

In addition, for each k, there are various patterns in the computed parameters as α decreases. For the Makeham local alternative, the absolute bias increases with decreasing α for all four scores. However, RTMSR and RTMSW increase with decreasing α for the log x and pseudo-efficient scores; whereas they decrease for the Savage and Wilcoxon score. The similarity between the results for the Savage and Wilcoxon scores follow from the fact that $-\log(1-v) \simeq v$ for small v. In both cases, however, the RTMSR ratios approach a number very close to 1. For, as α decreases to zero in the limit, the variance of the PTE will approach σ_R^2 since we will accept H_o with probability approaching one. Thus, RTMSR $\rightarrow 1 + \frac{C_{11}^2}{\sigma_R^2} \simeq 1$ since C_{11}^2 is small compared to σ_R^2. In the case of the parameter RTMSW, RTMSW increases with decreasing α when the log x and pseudo-efficient score are used, and decreases with decreasing α when the Savage and Wilcoxon scores are used. In this case,

$$\text{RTMSW} \rightarrow \frac{\sigma_R^2 + C_{11}^2}{\sigma_W^2} < 1 \text{ as } \alpha \rightarrow 0 \text{ since } \frac{\sigma_R^2}{\sigma_W^2}$$

is considerably less than 1 for the given range of k.

The behavior of the PTE under a Linear Hazard local alternative is somewhat different. Except for one instance for the log x score, the absolute bias increases with decreasing α for the Log x, Savage and Wilcoxon scores. The case of the pseudo-efficient score stands out because it is the only case among all those in this investigation for which the PTF_2 bias becomes zero for some $\alpha = \alpha(k)$ for $1.1 \leq K \leq 1.9$, where $.05 \leq \alpha \leq .20$. (See Table 7).

Finally, we note that in virtually all cases, the absolute value of the PTE bias is less than that of the rank estimator under each local alternative.

NON-CENTRALITY PARAMETERS

The non-centrality parameter C_2 results from two additive components: the covariance of $\sqrt{N}\, T_N(k)$ and $\sqrt{N}(\hat{k}_{NR}-k)$ with logL. The effect of estimating the nuisance parameter k can be considerable. One might expect that the pseudo-efficient score for the Makeham local alternative would produce a positive C_2. However, the covariance of $\sqrt{N}(\hat{k}_{NR}-k)$ and logL is negative; so negative in fact that the positive contribution of $\sqrt{N}T_N(k)$ is swamped by the negative contribution of $\sqrt{N}(\hat{k}_{NR}-k)$. This extreme case shows that the effect of substituting an estimate of a nuisance parameter can be very unpredictable since C_2 is so dependent upon the log likelihood under the local alternative.

For instance, Tables 4 and 5 show that for the Makeham alternative, the Savage and Wilcoxon scores have considerably more local power than the other two scores. However, the picture is exactly reversed for the Linear Hazard alternative. (See Tables 8 and 9). We can understand this phenomenon better by looking at what happens when $k=1$ under a Makeham alternative. In that case, $\sqrt{N}\, T_N(k)$ is directly proportional to $\sqrt{N}\, \bar{Y}$

and so is $\sqrt{N}\,(\hat{k}_{NR}-k)$. At the same time, LogL is an increasing function of Y_i. Thus, both components of the statistic contributed "large" positive amounts to the overall covariance. However, in the Linear Hazard case, LogL is a parabolic function of Y_i and so positive covariances are cancelled by negative covariances. Nevertheless, comparison of Table 4 with Table 8 shows a remarkably similar pattern of PTE behavior under both local alternatives in the face of quite different non-centrality parameters.

Next we look at the source of bias and PTE ratio variation among different scores for each of the two local alternatives.

PTE BIAS AND MSE RATIOS

In general, we would expect the MSE of the PTE to fall between σ_R^2 and σ_W^2. However, there are several occasions in which the MSE of the PTE is less than the variance of the rank estimator. The comparison of the pseudo-efficient and Savage score for the Makeham alternative is instructive (Tables 3 and 4). Notice that the biases are virtually identical so that the differences in the RTMSR arises from the fact that the variance of the PTE is greater when the Savage score is used. The explanation is evident when the equation for the variance of the PTE is examined. We wish to find the circumstances under which

$$(\sigma_W^2 p_2^2 - \sigma_R^2 p_1^2)\, z\, \phi(z) + (\sigma_R^2 - \sigma_w^2)\, \Phi(z) + \sigma_w^2 < \sigma_R^2. \text{ Letting}$$

$f = \dfrac{\sigma_W^2}{\sigma_R^2} > 1 \;\forall\; k$, and $z = C_\alpha - C_2$ then we can say that $\sigma_{PTE}^2 < \sigma_R^2$ if and only if:

$(fp_2^2 - p_1^2)\, z\, \phi(z) < (1-f)\, [1-\Phi(z)]$ or $(p_1^2 - fp_2^2)\, z\, \phi(z) > (f-1)\, [1-\Phi(z)]$

or

$$\frac{p_1^2 - fp_2^2}{f-1} > \frac{1-\Phi(z)}{z\, \phi(z)} ,$$

If the LHS < 0, then $\sigma^2_{PTE} > \sigma_R^2$ when $z > 0$. This is in fact the case for the Savage score. Thus when $p_1^2 \simeq p_2^2$ then $\sigma_{PTE}^2 > \sigma_R^2$. However, if $p_1^2 >> p_2^2$, then it is possible for $\sigma^2_{PTE} < \sigma_R^2$. We can then rewrite the inequality as

$$\frac{f-1}{p_1^2 - fp_2^2} < \frac{z\, \phi(z)}{1-\Phi(z)} .$$

The RHS is increasing in z since the hazard function for the standard normal distribution is increasing in z without bound. Thus, if $z >> 0$ and $p_1^2 >> p_2^2$, then $\sigma^2_{PTE} < \sigma_R^2$. This is what happens with the pseudo-efficient score. Since $C_2 < 0$, then $z > C_\alpha$ and LHS < RHS.

The foregoing is a good illustration of the fact that local optimability of the score function is not necessarily related to the performance of a preliminary test statistic. In fact, one of the two crucial quantities, $p_1^2 - fp_2^2$, has absolutely nothing to do with the particular local alternative. Finally, we note that the foregoing is equivalent to the statement that a sufficient condition for RTMSR > 1 is that $\sigma_{RT} < \sigma_{WT}$.

PTE BIAS

Since the bias of the rank MLE estimator is always negative under the two specified local alternatives, and the bias of the Wilcoxon estimator is always zero, we should expect the bias of the PTE to be less negative than the former but still negative. On occasion, however, the PTE bias is positive on the domain $1.1 \leq k \leq 1.9$ for some $\alpha(k)$, $.05 \leq \alpha \leq .20$.

This is clearly the case for the pseudo-efficient linear hazard score. (See Table 7). As $\alpha \to 0$, the PTE bias must approach that of the rank MSE and so since the bias is clearly a continuous function of α, we know that the bias is zero for some α in the range used in this study.

The general (but obviously not sufficient) condition for this situation to occur is that $p_1 < 0$ and $p_2 > 0$. In order to see this, we note that the PTE bias can be written:

$$\text{PTE(BIAS)} = (C_{11}-C_{12})\ \Phi(z) + (\sigma_W p_2 - \sigma_R p_1)\ \phi(z)$$

since the Wilcoxon statistic is unbiased for $P(Y>x)$. Since $C_{11} < C_{12}$ in all cases, the second term must be "as positive as possible" to make the total positive. This is what occurs in the case just cited. It does not occur for any other case.

The role of the non-centrality parameter is also important. For the PTE bias to be positive, we must have:

$$\int_{-\infty}^{C_\alpha} E(z_{11}|z_2) + \int_{C_\alpha}^{\infty} E(z_{11}|z_2) > C_{12} > 0$$

or

$$\int_{-\infty}^{C_\alpha} [C_{11} + p_1(z_2-C_2)]\ f_2(z_2)dz_2 + \int_{C_\alpha}^{\infty} [C_{12} + p_2(z_2-C_2)]\ f_2(z_2)dz_2 > C_{12} > 0.$$

In this example $|p_1| < < p_2$. Referring to Table 7, we see that as k increases, C_2 decreases, and the PTE bias increases. Thus the zero PRF_2 bias will occur with a test of size larger than .84. The explanation is that since $|p_1| << p_2$, the first term does not contribute much of a negative component, whereas the decrease in C_2 allows the second term to contribute a sizable positive component.

In summary, it is clear that the performance of the PTE both in terms of bias and MSE is a complicated function of the covariances of the test statistic and the estimators of $P(Y > X)$. When viewed globally, the size of the test and the non-centrality parameter have secondary influence.

When proportional hazards is true, Tables 10-14 show that the bias of the PTE is always positive and is very small when either the Savage or Wilcoxon scores are used. However, the variance of the PTE tends to be larger for those scores than for the others.

This chapter have demonstrated that the non-centrality parameter is very much dependent upon the local alternative since it depends heavily upon the covariance of the log likelihood and $\sqrt{N}\,(\hat{k}_{NR}-k)$. However, the behavior of the PTE appears similar despite the alternative. This is a somewhat encouraging development since we would like to use a statistic which will have predictable performance over a range of circumstances. Another way of putting this is that local power is a property of a test statistic and may be maximized under given circumstances. However, this study has suggested that considerations for optimal performance of a PTE in asymptotic situations where a nuisance parameter comes into play, may render local power irrelevant or a factor of diminished importance.

We end by stating that this problem has often veered into territory off the beaten path of traditional rank test theory. Once the null hypothesis does not require homogeneity (iid r.v.'s), consistency and local optimality results no longer apply. Therefore this problem does not appear to fit into a nice self-contained mathematical-statistical structure out of which "nice" results fall.

CHAPTER VI

The fact that the set of possible joint rank vectors does not have equiprobably elements under H_o, except when $k = 1$, produces considerable practical difficulties when the issue of censorship is addressed. For instances, under Type II censoring, it is not difficult to show that for the linear rank statistic be a martingal (an thus permit the use of the martingale central limit theorem), the constant score assigned to the censored observations must be a function of the probability that the $(r+1)$ st observation comes from a specified group if the r^{th} observation is the last one used. Although this probability can be expressed as a function of k is this problem, the expression is useless because it is so complex. If censoring is to be studied, some approximation methods for usable censored scores may be the way to go.

In addition, the choice of the pair of estimators is not obvious when censoring is present. With random censoring, does one simply use the censored version of the Cox estimator? What modification of the U-statistic is required?

Another topic for further research is the exploration of optimal scores for specified local alternatives. This is an extremely difficult problem due to the estimation of k and the fact that classical optimality results require the null hypothesis to be homogeneity of two groups. Even if one casts the problem in terms of a "restricted" one, i.e., makes it a constrained maximization problem, the difficulties are great because this problem is not one amenable to Lagrange multipliers. It is a problem in infinite dimensional control theory.

BIBLIOGRAPHY

Abel, (1982). A note on the Mann-Whitney statistic for Lehmann alternative. BIOMETRICAL JOURNAL 24, 565-570.

Anderson, (1982). Testing goodness-of-fit of Cox's regression and life model. BIOMETRICAS 38, 67-77.

Ashar, VG. (1970). On the use of preliminary tests in regression. Unpublished Thesis, North Carolina State University, Raleigh, NC.

Bancroft, TA. (1944). On biases in estimation due to the use of preliminary tests of significance. ANN MATH STATIS 15, 192-204.

Bancroft, TA. (1964). Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance. BIOMETRICS 20, 427-442.

Bancroft, TA and Han, CP. (1980). Inference based on conditionally specified ANOVA models incorporating preliminary tests. HANDBOOK OF STATISTICS Vol 1, 407-441.

Basu, AP. (1981). The estimation of P(X < Y) for distributions useful in life testing. NAVAL RESEARCH LOGISTICS QUARTERLY, 28, 383-392.

Birnbaum, ZA. (1956). On a use of the Mann-Whitney statistic. PROCEEDINGS OF THE THIRD BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, Vol 1, Univ. of Calif. Press, 13-17.

Birnbaum, ZW. and McCarty, RC (1958). A distribution-free upper confidence bound for Pr (Y < X), based on independent samples of X and Y. ANN MATH STATIST 29, 558-562.

Bock, ME, Yancey, TA and Judge, CG. (1973). The statistical consequences of preliminary test estimators in regression. J AMER STATIST ASSOC 68, 107-116.

Brook, RJ. (1976). On the use of a regret function to set significance points in prior tests of estimation. J AMER STATISTIC ASSOC 71, 126-131.

Chernoff, H. and Savage, RI. (1958). Asymptotic normality and efficiency of certain non-parametric test statistics. ANNALS OF MATH. STAT., 29, 972-994.

Chipman, JS and Rao, MM. (1964). The testing of linear restrictions in regression analysis. ECONOMETRICA 32, 198-209.

Doksum, KA and Yandell, BS (1984). Tests for exponentiality. HANDBOOK OF STATISTICS, Vol. 4, North-Holland Press, 579-611.

Epstein, B. (1960). Tests for the validity of the assumption that the underlying distribution of life is exponential. TECHNOMETRICS 2, 83-101.

Govindarajulu, Z. (1968). Distribution-free confidence bounds for Pr (X < Y). ANN INSTITUTE STATIST MATH 20, 229-238.

Hajek, J. and Sidak, Z. (1967). Theory of rank tests. Academia, Prague.

Hill, RC, Judge, GG and Fomby, TB. (1978). On testing the adequacy of a regression model. TECHNOMETRICS 20, 491-494.

Hoel, DG. (1972). A representation of mortality data by competing risks. BIOMETRICS, 28, 475-488.

Huntsberger, DV. (1955). A generalization of a preliminary procedure for pooling data. ANNALS OF MATHEMATICAL STATISTICS 26, #4, 734-743.

Johnson, NL. (1975). Letter to the Editor, TECHNOMETRICS 17, 393.

Jureckova, J. (1969). Asymptotic linearity of a rank statistics in regression prameter. ANNALS OF MATH. STAT. 40, 1889-1900.

Kelley, GD, Kelley, JA, and Schucany, WR. (1976). Efficient estimation of P(Y < X) in the exponential case. TECHNOMETRICA 18, 359-360.

Larson, HJ and Bancroft, TA. (1963a). Sequential model building for prediction in regression analysis. I ANN MATH STAT 34, 462-479.

Larson, HJ. and Bancroft, TA. (1963b). Biases in prediction by regression for certain incompletely specified models. BIOMETRIKA 50, 391-402.

LeCam, L. (1960). Locally asymptotically normal families of distributions. Univ. of Calif. Publ. in STAT. 3, 37-98.

Lehmann, EL. (1950). Notes on the theory of estimation. Mimeographed lectures, Univeristy of California.

Lehmann, EL. (1951). Consistency and unbiasedness of certain nonparametric tests. ANNALS OF MATH STAT 22, 165-179.

McRae, KB. (1971). Inference procedures for pairs of distributions with proportional failure rate functions. Ph.D. Dissertation, Oregon State University.

Md. AK, Saleh, E. and Sen, PK (1984). Nonparametric preliminary test inference. HANDBOOK OF STATISTICS, Vol 4, North-Holland Press, 275-297.

Mosteller, F. (1948). On pooling data. JASA 43, 231-242.

Nagelkerke, NJD, Oosting, J and Hart, AAM (1984). A simple test for goodness-of-fit of Cox's proportional hazards model. BIOMETRICS, 40, 483-486.

Neyman, J. (1959). Optimal Asymptotic tests of composite hypotheses in Grenander, U. (Editor). Probability and Statistics, The Harold Cramer Volume, Stockholm, 1959.

Pathek, PK, Zimmer, WJ, and Williams, RE. (1979). Non-parametric estimation of an acceleration parameter. COMUN STATIST-THEOR MATH A8(4), 367-383.

Sathe, YS and Shah, SP. (1981). On estimating P(X < Y) for the exponential distribution. COMMUN STATIST-THEOR MATH A10(1), 39-47.

Schoenfeld, D. (1980). Chi-squared goodness-of-fit tests for proportional hazards regression model. BIOMETRIKA 67, 145-153.

Steck, GP, Zimmer, WJ, and Williams, RE. (1974). Estimation of parameters in acceleration models. PROC. 1974 ANNUAL RELIABILITY AND MAINTAINABILITY SYMPOSIUM, Los Angeles, California, 428-431.

Taulbee, JD. (1979). A general model for hazard rate with covariables. BIOMETRICS 35, 439-450.

Tong, H. (1974). A note on the estimation of Pr(Y < X) in the exponential case. TECHNOMETRICS 16, 625; errata, 17, 395.

Tsiatis, AA. (1981). A large sample tudy of Cox's regression model. ANNALS OF STATISTICS, Vol. 9(1), 93-108.

Ury, HK. (1972). On distribution-free confidence bounds for Pr(Y < X). TECHNOMETRICS, 14, 577-580.

Van Dantzig, D. (1951). On the consistency and the power of Wilcoxon's two-sample test. KONINKLIJKE NEDERLANDSE ADADEMIE van WETENSCHAPEN PROCEEDINGS, Series A, 54, 1-8.

RESULTS

Tables 2 through 9 display results for the combinations of scores and local alternatives for k = 1.1,1.3,1.5,1.7,1.9 and critical values C which produce one-sided tests of size α= .05, .10, .20.

TABLE 2 9LOGx-MAKEHAM)
C_α = .84, 1.282, 1.645

k	BIAS	BIASR	BIASW	RTMSR	RTMSW	C_2
1.1	-.006, -.023, -.032	-.042	0	.952, .953, .969	.714, .715, .727	-.310
1.3	-.006, -.004, -.031	-.041	0	.952, .953, .969	.716, .727, .729	-.304
1.5	-.006, -.022, -.031	-.040	0	.952, .953, .968	.719, .719, .731	-.302
1.7	-.006, -.021, -.030	-.039	0	.952, .952, .968	.720, .74, .733	-.301
1.9	-.006, -.021, -.029	-.037	0	.952, .953, .968	.722, .722, .734	-.301

TABLE 3 (PSEUDO-EFFICIENT MAKEHAM-MAKEHAM
C_α = .84, 1.282, 1.645

k	BIAS	BIASR	BIASW	RTMSR	RTMSW	C_2
1.1	-.023, -.032, -.037	-.042	0	.970, .972, .984	.728, .730, .738	-.430
1.3	-.021, -.031, -.036	-.041	0	.967, .970, .982	.727, .730, .739	-.413
1.5	-.019, -.030, -.035	-.040	0	.965, .968, .980	.728, .731, .740	-.399
1.7	-.018, -.028, -.033	-.039	0	.964, .966, .979	.730, .732, .741	-.387
1.9	-.016, -.027, -.032	-.037	0	.963, .963, .978	.731, .732, .742	-.377

TABLE 4 (SAVAGE - MAKEHAM)

C_{α} = .84, 1.282, 1.645

k	BIAS	BIASR	BIASW	RTMSR	RTMSW	C_2
1.1	-.026, -.033, -.036	-.042	0	1.124, 1.076, 1.047,	.844, .808, .786	+.494
1.3	-.025, -.031, -.035	-.041	0	1.123, 1.077, 1.049,	.845, .811, .789	+.483
1.5	-.023, -.029, -.034	-.040	0	1.122, 1.079, 1.051,	.847, .815, .794	+.474
1.7	-.021, -.027, -.032	-.039	0	1.122, 1.081, 1.054,	.849, .818 .797	+.464
1.9	-.019, -.026, -.030	-.037	0	1.122, 1.082, 1.056,	.851, .821, .801	+.455

TABLE 5 (WILCOXON - MAKEHAM)

C_{α} = .84, 1.282, 1.645

k	BIAS	BIASR	BIASW	RTMSR	RTMSW	C_2
1.1	.003, -.009, -.020	-.042	0	1.164, 1.148, 1.122,	.873, .862, .842	.515
1.3	.002, -.011, -.021	-.041	0	1.160, 1.143, 1.116,	.873, .860, .840	.503
1.5	.000, -.012, -.021	-.040	0	1.156, 1.137, 1.111,	.873, .858, .839	.491
1.7	-.001, -.012, -.021	-.039	0	1.152, 1.132, 1.106,	.872, .857 .837	.479
1.9	-.002, -.013, -.021	-.037	0	1.149, 1.128, 1.101,	.872, .855, .835	.468

TABLE 6 (LOGX-LINEAR HAZARD

C_{α} = .84, 1.282, 1.645

k	BIAS	BIASR	BIASW	RTMSR	RTMSW	C_2
1.1	-.032, -.060, -.079	-.107	0	1.01, .969, .973	.757, .727, .730	.182
1.3	-.014, -.039, -.055	-.080	0	1.00, .959, .961	.752, .721, .723	.151
1.5	-.003, -.024, -.039	-.060	0	.994, .954, .955	.751, .720, .721	.123
1.7	+.005, -.014, -.027	-.017	0	.990, .951, .953	.749, .720 .721	.100
1.9	+.010, -.007, -.019	-.019	0	.987, .950, .952	.749, .720, .722	.082

TABLE 7 (PSEUDO-EFFICIENT LINEAR HAZARD-LINEAR HAZARD

$C_\alpha = .84, 1.282, 1.645$

k	BIAS	BIASR	BIASW	RTMSR	RTMSW	C_2
1.1	+.021, -.017, -.048	-.107	0	1.140, 1.106, 1.085	.856, .830, .814	.430
1.3	+.030, -.004, -.031	-.080	0	1.115, 1.071, 1.049	.839, .806, .789	.323
1.5	+.037, +.005, -.019	-.060	0	1.096, 1.050, 1.029	.828, .793, .777	.247
1.7	+.042, +.012, -.010	-.017	0	1.083, 1.035, 1.016	.819, .784 .769	.195
1.9	+.045, +.017, -.004	-.019	0	1.722, 1.025, 1.001	.813, .777, .764	.156

TABLE 8 (Savage-Linear Hazard)

$C_\alpha = .84$ (1.282) 1.645

k	BIAS	BIASR	BIASW	RTMSR	RTMSW	C_2
1.1	-.086, -.097, -.102	-.107	0	1.10, 1.070 1.058	.822, .803, .794	-.018
1.3	-.063, -.071, -.076	-.080	0	1.083,1.056,1.041	.815, .794, .784	-.029
1.5	-.046, -.053, -.056	-.060	0	1.079,1.050,1.034	.815, .792, .780	-.029
1.7	-.034, -.040, -.043	-.017	0	1.078,1.048,1.031	.816, .793 .780	-.028
1.9	-.025, -.030, -.033	-.019	0	1.079,1.048,1.030	.818, .795, .781	-.027

TABLE 9 ((Wilcoxon-Linear Hazard)

$C_\alpha = .84, 1.282, 1.645$

k	BIAS	BIASR	BIASW	RTMSR	RTMSW	C_α
1.1	-.059, -.079, -.092	-.107	0	1.164,1.141 1.115	.874, .856, .836	.086
1.3	-.040, -.057, -.067	-.080	0	1.153,1.124,1.095	.867, .846, .824	.045
1.5	-.026, -.040, -.049	-.060	0	1.146,1.114,1.083	.865, .841, .818	.025
1.7	-.016, -.028, -.036	-.017	0	1.142,1.109,1.077	.864, .839 .815	.014
1.9	-.008, -.020, -.027	-.019	0	1.140,1.106,1.074	.864, .839, .814	.007

TABLE 10 (LogX)

k	BIAS	RTMSR	RTMSW
1.1	.042, .026, .015	.985, .951, .953	.739, .714, .715
1.3	.041, .026, .015	.984, .951, .953	.740, .715, .717
1.5	.039, .025, .014	.983, .950, .953	.742, .717, .719
1.7	.038, .024, .014	.982, .950, .953	.743, .719, .721
1.9	.036, .022, .013	.982, .950, .953	.744, .720, .723

TABLE 11 (Pseudo-Efficient Makeham)

k	BIAS	RTMSR	RTMSW
1.1	.023, .015, .009	1.00, .968, .967	.750, .727, .726
1.3	.025, .015, .009	1.00, .965, .965	.750, .726, .726
1.5	.025, .016, .009	.994, .963, .964	.751, .727 .727
1.7	.025, .016, .009	.993, .962, .963	.751, .728 .729
1.9	.025, .015, .009	.992, .962, .963	.752, .729, .730

TABLE 12 (Savage)

k	BIAS	RTMSR	RTMSW
1.1	.000, .000, .000	1.067, 1.034, 1.017	.801, .776, .763
1.3	.001, .001, .000	1.070, 1.036, 1.019	.805, .780, .767
1.5	.002, .001, .001	1.072, 1.040, 1.022	.810, .785, .771
1.7	.003, .002, .001	1.075, 1.043, 1.024	.814, .789 .775
1.9	.004, .003, .001	1.078 1.046, 1.026	.818, .793, .778

TABLE 13 (Wilcoxon)

k	BIAS			RTMSR			RTMSW		
1.1	.022,	.014,	.008	1.147,	1.109,	1.073	.861,	.832,	.805
1.3	.022,	.014,	.008	1.145,	1.108,	1.073	.862,	.833,	.807
1.5	.021,	.013,	.008	1.143,	1.106,	1.072	.863,	.835,	.809
1.7	.021,	.013,	.008	1.142,	1.105,	1.071	.864,	.836	.811
1.9	.020,	.013,	.007	1.141	1.104,	1.070	.865,	.837,	.812

TABLE 14 (Pseudo-Efficient Linear Hazard)

k	BIAS			RTMSR			RTMSW		
1.1	.070,	.044,	.026	1.126,	1.078,	1.048	.845,	.809,	.786
1.3	.069,	.043,	.025	1.101,	1.055,	1.030	.828,	.794,	.775
1.5	.067,	.042,	.025	1.085,	1.041,	1.019	.819,	.786,	.770
1.7	.066,	.041,	.024	1.074,	1.030,	1.012	.813,	.780	.766
1.9	.064,	.040,	.024	1.065	1.022,	1.006	.808,	.775,	.763

APPENDIX A

$$I \int_{0<H_N\leq 1}^{\infty} M(F_{n_1},G_{n_2})dH_N = \int_{0<H_N<1}^{\infty} M(F_{n_1},G_{n_2})dH_N + \int_{H_N=1}^{\infty} M(F_{n_1},G_{n_2})dH_N$$

$$M(F_{n_1},G_{n_2}) = \frac{(1-F_{n_1})}{p(1-G_{n_2}) + k\,(1-F_{n_1})}, \quad p = \frac{1-\lambda}{\lambda}$$

If the last observation is from G, then $M = 0$, if the last observation is from F, then $M = \frac{1}{k}$. Thus

$$\int_0^{\infty} M(F_{n_1},G_{n_2})dH_N \sim o_p(N^{-\frac{1}{2}}) .$$

On the interval $0 < H_N < 1$, all except one of the remainder terms require only the check on condition (4) of the Chernoff-Savage theorem.

Let $H = (1-\lambda)\,u + v\lambda$. $0 < u, v < 1$ and let C be a generic constant.

(i) $M(u,v) = \dfrac{1}{p(\frac{1-u}{1-v}) + k} \leq \dfrac{1}{k}$

(ii) $\left|\dfrac{\partial M}{\partial u}\right| = \dfrac{P}{1-v} \Big/ [P(\frac{1-u}{1-v}) + k]^2 = \dfrac{P(1-v)}{[P(1-u)+(1-v)]^2}$

Now $1-v \leq C(1-H)$. Thus

$$\frac{\partial M}{\partial v} \leq \frac{P(1-v)}{\min\,[(\frac{P}{\lambda})^2,(\frac{k}{1-\lambda})^2]\;(1-H)^2} \leq C(1-H)^{-1} \leq C\,[H(1-H)]^{-1}$$

$$\left|\frac{\partial M}{\partial v}\right| = \frac{P(1-u)}{(1-v)^2} \Big/ [P(\frac{1-u}{1-v}) + k]^2 = \frac{P(1-u)}{[P(1-u)+k(1-v)]^2} \leq C[H(1-H)]^{-1}$$

(iii) $\frac{\partial^2 M}{\partial u^2} = \frac{\frac{C}{(1-v)^2}}{[P(\frac{1-u}{1-v})+k]} \leq \frac{C(1-v)}{(1-H)^3} \leq C(1-H)^{-2} \leq C\,[H(1-H)]^{-2}$

$$\left|\frac{\partial^2 M}{\partial v^2}\right| \leq C\,\left[\frac{1-u}{1-v} + \left(\frac{1-u}{1-v}\right)^2\right]$$

Now $\frac{1-H}{1-v} = \lambda + P\,\frac{1-u}{1-v} \Rightarrow \frac{1-u}{1-v} \leq C$

Thus $\left|\frac{\partial^2 M}{\partial v^2}\right| \leq \frac{C}{(1-H)^2} \leq C\,[H(1-H)]^{-2}$.

The only term which needs a further result is the mixed partial term, R_{4N}. Note that

$$\frac{\partial^2 M}{\partial u \partial v} \leq \frac{\frac{C}{(1-v)^2}}{[P(\frac{1-u}{1-v})+k]^2} + \frac{\frac{C(1-u)}{(1-v)^3}}{[P(\frac{1-u}{1-v})+k]^3} \leq$$

$$\frac{C}{(1-H)^2} + \frac{C}{(1-H)^3} \leq \frac{C}{(1-H)^2} \leq k\ [H(1-H)]^{-2}.$$

Now $R_{4N} = \int_{0 \leq H_N < 1} (F_{n_1}-F)\,(G_{n_2}-G)\,\frac{\partial^2 M}{\partial F_{n_1} \partial G_{n_2}}\,(v_a)\,dH_N$.

By a result subsequent to the original Chernoff-Savage paper

$$\sup_{0 < H_N < 1} \frac{\sqrt{N}\,|u_n-u|}{\sqrt{u}(1-u)} = O_p(1).$$

Thus $|R_{4N}| \leq \int_{0 < H_N < 1} \frac{\sqrt{F}(1-F)}{\sqrt{n_1}} \cdot \frac{\sqrt{G}(1-G)}{\sqrt{n_2}}\,\frac{\partial^2 M}{\partial F_{n_1} \partial G_{n_2}}\,(v_a)\,dH_N$

$$\leq \frac{C}{N} \int_{0 < H_N < 1} H(1-H)\,[H(1-H)]^{-5/2+\delta}\,dH_N$$

$$\rightarrow \frac{C}{N} \int_{S_{N\varepsilon}} [H(1-H)]^{-3/2+\delta} dH$$

$$\leq \frac{C}{N} \int_{\frac{1}{CN}}^{1} \frac{dH}{H^{3/2-\delta}} \leq \frac{k}{N} \cdot N^{1/2-\delta}$$

then $R_{4_N} = o_p(N^{-1/2})$.

APPENDIX B

CONDITIONS FOR CALCULATION OF B

I. for log x, note that $E(\log x + \log k^* + c) = \log k^* - \log k$ where C = Euler's constant. Thus

$$B = \lim_{k^* \to k} \frac{\log k^* - \log k}{k^* - k} = \frac{d \log k^*}{dk^*}\Big|_{k^*=k} = \frac{1}{k}$$

For $J[H(x,k] = 2\,(1-e^{-kx}) + \frac{2x}{2} - \frac{3}{2}$, $B = \frac{1}{k}$

$$\left|\frac{\partial J}{\partial k}\right| = 2xe^{-k_x} + \frac{x}{2} < 2x + \frac{x}{2} \text{ which is integrable.}$$

For $J[H(x,k)] = -\frac{x^2}{2} + \frac{2x}{k} - \frac{1}{k^2}$, $B = 0$.

$$\left|\frac{\partial J}{\partial k}\right| \leq \frac{2x}{k^2} + \frac{2}{k^3} \leq 2x + 2 \text{ for } k > 1.$$

II. In order to simplify notation we let $J(H(x, k) = J(u)$ under H_o, the transformed random variables are exp(1) and exp(k). I.e.,

$$H(x) = (1-\lambda)\,(1-e^{-x}) + \lambda(1-e^{-k_x})$$

$$|H'(x)| = (1-\lambda)e^{-x} + k\lambda e^{-k_x}$$

$$|H''(x)| = (1-\lambda)e^{-x} + k^2\lambda e^{-k_x}$$

Note that if $k \geq 1$, then $1-e^{-x} \leq H(x) \leq 1-e^{-k_x} \to -\frac{1}{k}\log(1-u) \leq x \leq -\log(1-u)$

Log x

$$x \to \infty$$

$$x \le -\log(1-u) \Rightarrow \log x \le \log[-\log(1-u)] .$$

Let $u = \frac{T-1}{T}$, T large. Then $\log x \le \log\log T$ and

$$[u(1-u)]^{-1/2+\delta} = \left(\frac{T-1}{T^2}\right)^{-1/2+\delta} \simeq T^{1/2-\delta}, \text{ T large for}$$

some $\delta > 0$.

Clearly, $\log\log T \le \log T$, so it is sufficient to show that $\log T \le T^{-1/2-\delta}$. Let $T = e^v$. Then we must show that $v \le e^{v(1/2-\delta)}$, v large, or $\log v \le v\,(1/2-\delta)$. But $\frac{\log v}{v} \to 0$ as $v \to \infty$

$$\frac{\log v}{v} \to 0 < \frac{1}{2} - \delta \quad \text{where} \quad 0 < \delta < 1/2$$

$$|J'[u]|$$

$$x \to 0$$

$J'(u) = \frac{1}{xH'(x)}$ $H'(0) > 0$. This $J'(u) = \frac{c}{x} \le -\frac{c}{\log(1-u)}$

Near $u = 0$, $-\log(1-u) = u + \frac{u^2}{2!} + \frac{u^3}{3!} + \ldots$

Thus $\frac{c}{x} \le \frac{c}{u+\frac{u^2}{2}+\ldots} \le \frac{c}{u} \le \frac{c}{u(1-u)} = c[(1-u)u]^{-1}$

$$x \to \infty$$

$$H'(x) \ge 1-H(x) \text{ for } k \ge 1.$$

Thus

$$\frac{1}{H'(x)} \le \frac{1}{1-u} \le \frac{1}{u(1-u)} \Rightarrow \frac{1}{xH'(x)} \le \frac{c}{u(1-u)}$$

$$|J''(u)$$

$$x \to 0$$

$$|J''(u) = \frac{H''(x)}{x[H'(x)]^3} + \frac{1}{x^2[H'(x)]^2}$$

1^{st} Term: $H'(o) > 0$. Thus the first term reduces to the previous case.

2^{nd} Term: $$\frac{1}{x^2[H'(x)]^2} = \frac{c}{x^2} \le \frac{C}{\log^2(1-u)}$$

$$= \frac{c}{(u+\frac{u^2}{2} + \quad)^2} < \frac{k}{u^2} < \frac{c}{[u(1-u)]^2}$$

$x \to \infty$

1^{st} Term: $\frac{H''(x)}{H'(x)} \to 1$. Then $$\frac{H''(x)}{x[H'(x)]^3} < \frac{c}{x[H'(x)]^2}$$

Now $\frac{1}{H'(x)} < \frac{1}{1-u}$, so $$\frac{c}{x[H'(x)]^2} < \frac{c}{[u(1-u)]^2}$$

$$\frac{2[1-e^{-kx}] + \frac{k_x}{2} - \frac{3}{2}}{\quad}$$

$x \to \infty$: $x \le \log T$ where $T = \frac{1}{1-u}$ but $\log T \le T^{1/2 - \delta}$ as show earlier.

$$|J''(v)|$$

$$x \to \infty \; J'(H(x)) = \frac{2k_e^{-kx} + \frac{k}{2}}{H'(x)} < \frac{Ce^{-k_x}}{(1-\lambda)_e^{-x} + k\lambda_e^{-k_x}}$$

$$= \frac{C}{e^{(k-1)_x} + 1} < \infty$$

$|J''(v)|$

$$J''(H(x)) = \frac{[2ke^{-kx} + \frac{k}{2}]\ H''(x)}{[H'(x)]^3} + \frac{Ce^{-kx}}{[H'(x)]^2} \qquad x \to \infty : \frac{H''(x)}{H'(x)} \to C$$

Thus

$$|J''(H(x)| \leq \frac{C}{[H'(x)]^2} \leq \frac{1}{[u(1-u)]^2}, \text{ since}$$

$$\frac{1}{H'(x)} \leq \frac{1}{1-u} \leq \frac{1}{(1-u)} \qquad \text{when } k \geq 1.$$

$$-\frac{x^2}{2} \quad \frac{2x}{k} \quad -\frac{1}{k^2}$$

$|J(u)| \leq C(x^2+x)$. Now $x^2 \leq (\log T)^2$ as before. But $(\log T)^2 \leq T^{1/2-\delta}$ by letting $T = e^u$ as before.

$$x \to \infty: \ |J'(H(x)| < \frac{Cx}{H'(x)} . \ \text{Now} \ \frac{x}{H'(x)} \leq \frac{[u(1-u)]^{-1/2-\delta}}{u(1-u)} =$$

$$= [u(1-u)]^{-3/2+\delta}$$

$J''(u)$

$$|J''(H(x)| \leq C \left[\frac{x}{[H'(x)]^2} + \frac{1}{[H'(x)]^2} \right]$$

$$x \to \infty: \ \frac{x}{[H'(x)]^2} \leq \frac{[u(1-u)]}{[(1-u)]^2} = [u(1-u)]^{-5/2+\delta}$$

$-\log(1-u)$

$$J(u) = \log T \quad \text{so} \quad J(u) \leq [u(1-u)^{-1/2+\delta}$$

$$J'(u) = \frac{1}{1-u} \leq \frac{1}{(1-u)}$$

$$J''(u) = \frac{1}{(1-u)^2} \leq \frac{1}{[u(1-u)]^2}$$

APPENDIX C

We applied the linear rank statistic using Wilcoxon scores $(\frac{R_i}{N})$ to a data set produced by Hoel (1972). Two groups of male mice were given 300 rods of radiation and followed for cancer incidence. One group was maintained in a germ-free environment. Three causes of death were recorded and we compared the two groups with respect to "other causes." (See table on next page). Each group had had one tie, so the tied observations were deleted, so that $n_1 = 38$, $n_2 = 37$.

Figure 1 shows the survivorship curves for the two groups. Group 2 is the germ-free group. Figure 2 is a log(-log survival) plot for a visual assessment proportional hazards. If proportional hazards holds, then the two lines should be parallel. Such a relationship appear to hold generally.

The results are as follows. The rank maximum likelihood estimate of k is 6.08. The centering constant is thus calculated to be .3026, and $N \text{ Var } (T_N(\hat{k}_{NR}) - \mu)$ is .1116. Using the scores from Group 1 (since $k > 1$), we find that $T_N(\hat{k}_{NR}) = .3263$. Thus the asymptotically standard normal statistic under H_o is .751, clearly non-significant. In other words, the data is consistent with the notion that these two samples come from populations whose survival experience is related by proportional hazards with a proportionality constant of approximately six.

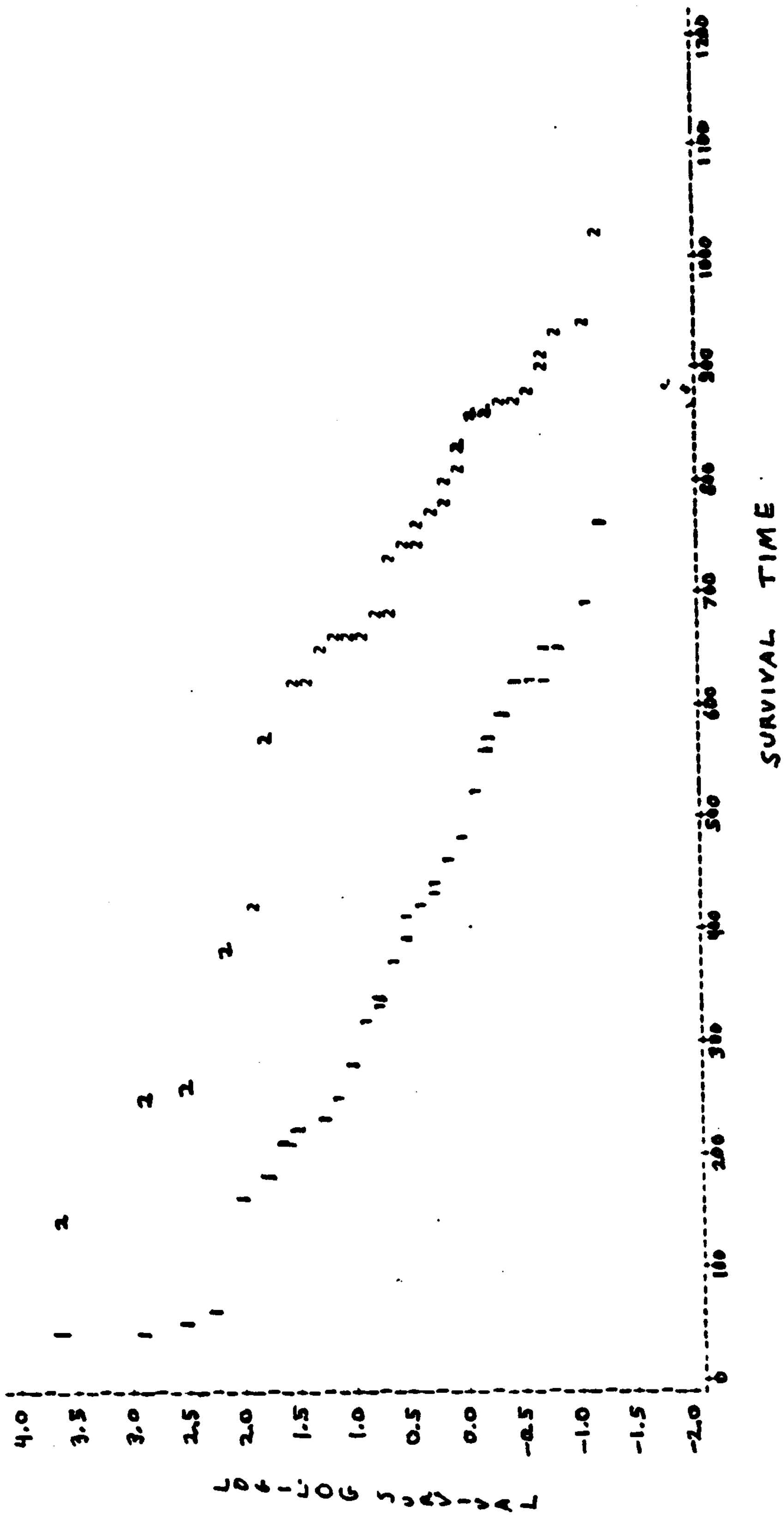
Data Set IV—*continued*

Other causes (39%)	40, 42, 51, 62, 163, 179, 206, 222, 228 252, 249, 282, 324, 333, 341, 366, 385, 407 420, 431, 441, 461, 462, 482, 517, 517, 524 564, 567, 586, 619, 620, 621, 622, 647, 651 686, 761, 763
Germ-free group	
Thymic lymphoma (22%)	158, 192, 193, 194, 195, 202, 212, 215, 229 230, 237, 240, 244, 247, 259, 300, 301, 321 337, 415, 434, 444, 485, 496, 529, 537, 624 707, 800
Reticulum cell sarcoma (18%)	430, 590, 606, 638, 655, 679, 691, 693, 696 747, 752, 760, 778, 821, 986
Other causes (46%)	136, 246, 255, 376, 421, 565, 616, 617, 652 655, 658, 660, 662, 675, 681, 734, 736, 737 757, 769, 777, 800, 807, 825, 855, 857, 864 868, 870, 870, 873, 882, 895, 910, 934, 942 1015, 1019

Source: Hoel (1972). For discussion see Sections 1.1.1 and 7.

Data Set V MOUSE LEUKEMIA DATA

T_1	T_2	J	δ	z_1	z_2	z_3	z_4	z_5	z_6
121175	122377	2	2	1	2	2	1	00.0	10000
121175	122377	3	2	1		2	2	07.4	02400
121175	060677	3	1	1		2	2	13.2	00000
121175	121476	2	1	1		2	2	00.0	10000
121175	122377	3	2	2	2	1	1	05.8	00000
121175	010577	2	1	2		1	1	00.0	08800
121175	111677	5	1	1	2	1	1	00.0	08000
121175	101477	1	1	2		1	2	00.0	10000
121175	122377	3	2	2	2	2	1	78.7	00000
121175	122377	3	2	2		2	2	05.3	00080
121175	122377	1	2	2	2	2	2	20.3	00000
121175	040777	1	1	1	1	2	2	04.7	10000
121175	081076	1	1	1		2	2		



SURVIVAL TIME
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0

FIGURE II

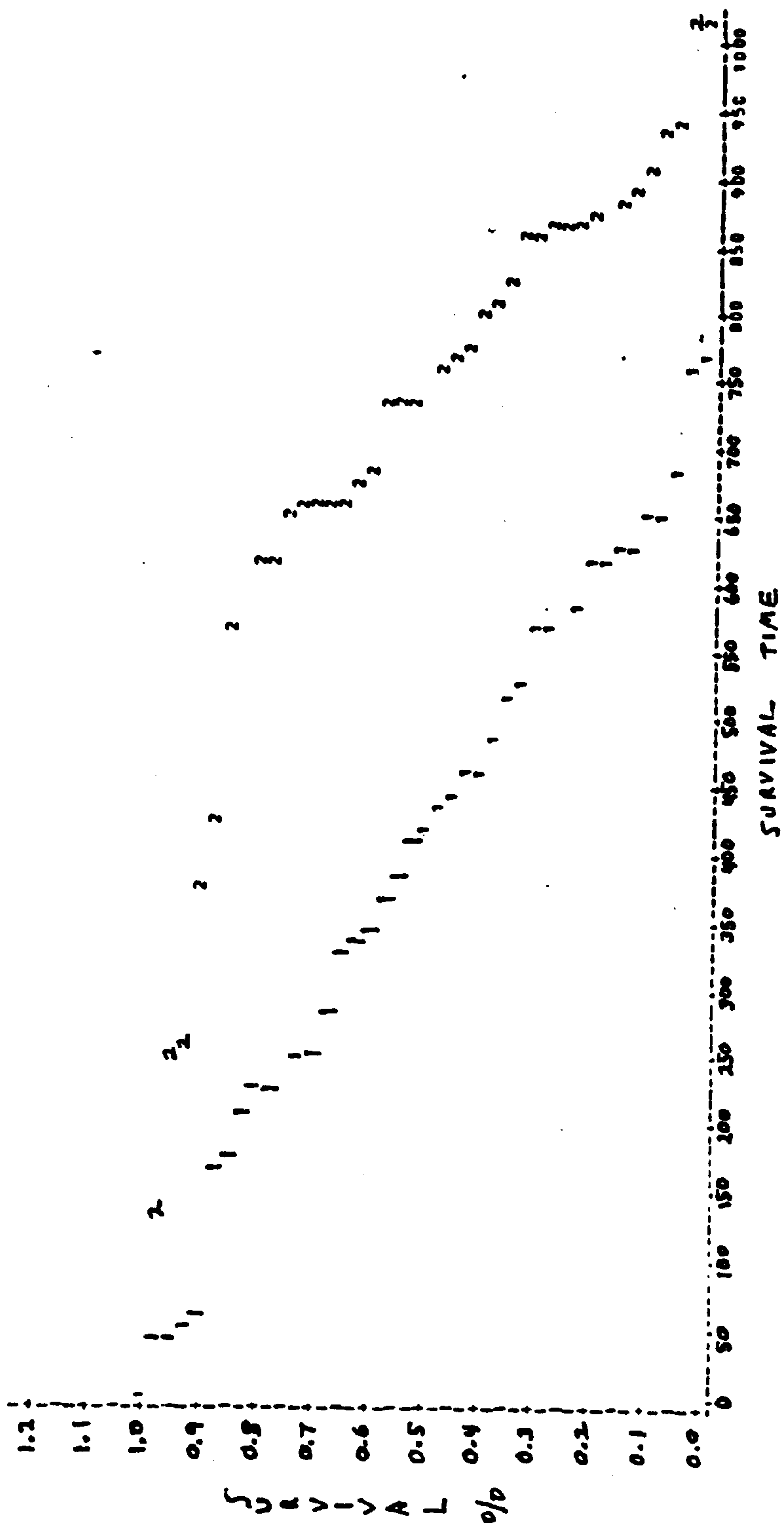

SURVIVAL TIME
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000
SURVIVAL %
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2

FIGURE I