Ann. Inst. Statist. Math.
30 (1978), Part A, 281-295

A NON-PARAMETRIC TEST FOR COMPOSITE HYPOTHESES IN SURVIVAL ANALYSIS

H. DENNIS TOLLEY

(Received June 13, 1977; revised Sept. 18, 1978)

Summary

For survival data with several concomitant (regressor) variables a large sample non-parametric procedure is presented which provides significance tests of hypotheses about a subset of the concomitant variables. This non-iterative procedure resembles linear model methodology in simplicity and form. The method is useful to eliminate unimportant concomitant variables prior to estimation of model parameters.

1. Introduction

In multiparameter experiments we are often mainly interested in some composite hypothesis in which only a subset of the parameters are specified. Parametric formulations, particularly the general linear model, have provided a unified approach to significance tests in the presence of nuisance parameters. Non-parametric methods, however, are considerably more restrictive in their applicability to this kind of problem. Extant non-parametric methods require either a designed experiment (e.g., a two-way layout such as random blocks and treatments) (see e.g., Friedman [7], Kruskal [12], Bhapkar [1]), or (iterative) techniques to eliminate nuisance parameter effects (see e.g., Puri and Sen [14]). One major area of application of non-parametric methodology has been in the analysis of lifetime data. Increasingly, however, multiparameter survival problems arise which do not lend themselves to analysis using either of the above non-parametric methodologies. For a special but important subset of survival models, Cox [3] has given a formulation in which the overall effect of the complete set of parameters may be tested using linear rank statistics. The purpose of this paper is to show that the general formulation of survival models as given by Cox can be used to provide large sample tests of composite hypothesis using linear rank methodology. This non-iterative approach is based on the work of Hájek [8] and Hoeffding [10]. We will see that the resulting methodology resembles the parametric linear model approach both in form and in simplicity.

In the next section of this paper, the Cox model is reviewed. It is assumed a priori that the concomitant variables are random variables, the realizations of which are inputs to the survival function. In the third section the test statistic is presented in a general form. The statement of assumptions, the preliminaries, and the proof for the results for the two parameter case are given in subsequent sections.

2. Cox survival model

Recently, increased interest has been given to the use of concomitant variables (covariables) which characterize experimental units (e.g., patients). In addition to the treatment received, concomitants, such as for example, weight, sex, and blood pressure, will often play a role in the probable lifetime of a patient. Several researchers have studied survival methodology in the presence of concomitants. Of these, Cox [3] gives a 'likelihood' type formulation which provides a fertile basis for analyzing survival data.

The formulation of Cox [3] for failure models is given as follows. Consider an experiment in which the failure time, or censoring time, is observed for each of n individuals. Suppose, in addition, that a set of p concomitants $\boldsymbol{Z}'=(Z_1,\cdots,Z_p)$ is observed for each individual. These can be regressor type variables such as age or dose level, or indicator variables. Following Cox, it is assumed that the relationship between the realizations $\boldsymbol{z}$ of $\boldsymbol{Z}$ to the failure time density function is expressed in the hazard function as

$$\lambda(t,\boldsymbol{z})=\lambda_0(t)\exp(\boldsymbol{z}'\boldsymbol{\beta}) \,, \tag{1}$$

where $\boldsymbol{\beta}$ is a vector of unknown coefficients. Let $R(t_i)$ be the set of labels attached to the individuals at risk just prior to time t_i, i.e., individuals who have neither failed nor been censored prior to time t_i. Assuming that censoring is independent of failure time, Cox modeled the likelihood of failure at time t_i as conditional on the set of individuals whose failure or censoring time is at least t_i. Explicitly, the likelihood associated with the ith failure conditional on $R(t_i)$ is

$$l_i=\frac{\exp(\boldsymbol{z}'_{(i)}\cdot\boldsymbol{\beta})}{\sum_{l\in R(t_i)}\exp(\boldsymbol{z}'_l\cdot\boldsymbol{\beta})} \,. \tag{2}$$

The resulting "likelihood", $L=\prod_{i=1}^{r} l_i$, gives the likelihood estimating equations

$$\frac{\partial \ln L}{\partial \beta_\zeta} = U_\zeta(\boldsymbol{\beta}) = \sum_{i=1}^{\tau} (z_{i\zeta} - A_{i\zeta}) \qquad \zeta = 1, 2, \cdots, p \tag{3}$$

where τ is the number of observed failures, and

$$A_{i\zeta} = \frac{\sum_{l \in R(t_i)} z_{l\zeta} \exp(\boldsymbol{z}_l' \boldsymbol{\beta})}{\sum_{l \in R(t_i)} \exp(\boldsymbol{z}_l' \boldsymbol{\beta})} . \tag{4}$$

The Fisher information matrix is given by

$$I(\boldsymbol{\beta}) = \{ I_{\zeta\gamma}(\boldsymbol{\beta}) \} , \tag{5}$$

where

$$I_{\zeta\gamma}(\boldsymbol{\beta}) = \sum_{i=1}^{\tau} \left\{ \frac{\sum_{l \in R(t_i)} z_{l\zeta} z_{l\gamma} \exp(\boldsymbol{z}_l' \boldsymbol{\beta})}{\sum_{l \in R(t_i)} \exp(\boldsymbol{z}_l' \boldsymbol{\beta})} - A_{i\zeta} A_{i\gamma} \right\} .$$

Under mild conditions the "maximum likelihood" estimate, $\hat{\boldsymbol{\beta}}$, of $\boldsymbol{\beta}$ will be asymptotically normally distributed with covariance matrix $I^{-1}(\boldsymbol{\beta}_0)$ where $\boldsymbol{\beta}_0$ obtains (see Cox [4]). Hence the statistic $U'(\boldsymbol{\beta}_0) = (U_1(\boldsymbol{\beta}_0), \cdots, U_p(\boldsymbol{\beta}_0))$ is asymptotically multinormal with covariance matrix $I(\boldsymbol{\beta}_0)$ when $\boldsymbol{\beta}_0$ obtains. Under the global hypothesis $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$ and (1) above, the statistic

$$Q_1 = U'(0) I^{-1}(0) U(0) \tag{6}$$

has asymptotically a chi-square distribution.

To relate this model to linear rank statistics, let us view the problem in a different perspective. Write equation (3) as

$$U_\zeta(\boldsymbol{\beta}) = \sum_{i=1}^{\tau} z_{(i)\zeta} \left(1 - \sum_{j : R_j \leq R_i} \frac{\exp(\boldsymbol{z}_{(i)}' \boldsymbol{\beta})}{\sum_{l \in R(t_j)} \exp(\boldsymbol{z}_l' \boldsymbol{\beta})} \right) \qquad \zeta = 1, 2, \cdots, p . \tag{7}$$

In this form, if we put $\boldsymbol{\beta} = 0$ then $U_\zeta(\boldsymbol{\beta})$ is a simple linear rank statistic with regressor variable $z_{(i)\zeta}$ and score function

$$a_{\zeta, z_i \beta}(R_i) = 1 - \sum_{j : R_j \leq R_i} \left(\frac{\exp(\boldsymbol{z}_{(i)} \boldsymbol{\beta})}{\sum_{l \in R(t_j)} \exp(\boldsymbol{z}_l \boldsymbol{\beta})} \right) , \tag{8}$$

where R_j is the rank of the failure time t_j associated with $\boldsymbol{z}_j$ among all the failures. This linear rank relationship and resulting connection with log rank tests has been pointed out by several researchers (see e.g., Downton [6] and Crowley [5]).

3. Tests of a general composite hypothesis

Consider now the composite or partial null hypothesis which specifies only a subset of the parameters as null, i.e., $H_1: \beta_1=\beta_2=\cdots=\beta_r=0$, where $r\leqq p$. To form the test, let $\hat{V}$ be the estimated covariance matrix of the observed concomitants z_i, $i=1,\cdots,n$, from a sample of size n. Let $PP'=\hat{V}^{-1}$ where P is a lower triangular matrix and $\hat{V}^{-1}$ is assumed to exist. Let P be partitioned as

$$P=[P_1, P_2] , \tag{9}$$

where P_1 is a $p\times r$ matrix and P_2 is a $p\times(p-r)$ matrix. Then, the multiparameter corollary of Lemma 1, given below, states that, under certain assumptions, the statistic

$$Q_2=U'(0)P_1(P_1'I(0)P_1)^{-1}P_1'U(0) \tag{10}$$

has asymptotically a (central) chi-square distribution when H_1 holds. Consequently, for large samples, significance tests for the partial null hypothesis H_1 can be simply formed using (10). In the following we make the foregoing more exact for the case $p=2$ and $r=1$.

4. Preliminaries

Let $Z=(X, Y)$ be a random vector of concomitant variables which has a bounded range and cumulative distribution function $G(x, y)$ such that the second moments of X and Y are non-null, and the covariance of X and Y is zero. Let T_i be a random variable with continuous cumulative distribution function $F(t, \beta_1 x_i, \beta_2 y_i)$, where x_i and y_i are realizations of X and Y. F is assumed to be absolutely continuous and such that all second partial derivatives of $\partial F/\partial t$ with respect to x and y are continuous. For a sample of n realizations of T_i, denoted by $\{t_i\}$, let R_i be the rank of t_i among $(t_1,\cdots,t_n)$.

Define the regressors $c_{in}^*=c_n^*(x_i)$ as functions of outcomes x_i and let

$$c_{in}=c_{in}^*-\bar{c}_n , \tag{11}$$

where

$$\bar{c}_n=n^{-1}\sum_{i=1}^{n} c_{in}^* .$$

Define the score $a_n(i)$ which is generated by the function $\phi(t)$, $0<t<1$ in one of the following two ways:

$$a_n(i)=\phi\left(\frac{i}{n+1}\right) \tag{12}$$

or

$$a_n(i) = \mathrm{E}\,\phi(U_i^{(n)}) \, , \tag{13}$$

where $U_i^{(n)}$ represents the ith order statistic from a sample of size n from a uniform distribution over (0, 1). The score function $\phi(\cdot)$ will be made more explicit below.

The test statistic of interest is based on the simple linear rank statistic

$$S = \sum_{i=1}^{n} c_{in} a_n(R_i) \, . \tag{14}$$

To show the asymptotic normality for a given fixed sequence $z_i = (x_i, y_i)$, $i = 1, \cdots$, of outcomes, we make the following assumptions.

ASSUMPTION 1. $\phi(t) = \phi_1(t) - \phi_2(t)$, $0 < t < 1$, where $\phi_1(t)$ and $\phi_2(t)$ are both non-decreasing, square integrable, and absolutely continuous inside (0, 1).

ASSUMPTION 2. For any $\eta > 0$ there exists an N_η such that for all $n > N_\eta$

$$\mathrm{Var}\,(S) > \eta n \sum_{i=1}^{n} c_{in}^2 \, . \tag{15}$$

ASSUMPTION 3.

$$\int_0^1 t^{1/2}(1-t)^{1/2} d\phi_k(t) < \infty \, , \qquad k = 1, 2 \, . \tag{16}$$

Then, (conditional upon the sequence of realizations z_i, $i = 1, \cdots$) the following theorem proved by Hájek [8] (Theorem 2.3) holds.

THEOREM 1 (Hájek). *Under Assumptions* 1 *and* 2, *for every* $\varepsilon > 0$ *and* $\eta > 0$ *there is an* $N_{\varepsilon\eta}$ *such that for* $n > N_{\varepsilon\eta}$

$$\sup_x \left| \Pr\,(S - \mathrm{E}\,S < x \cdot d) - (2\pi)^{-1/2} \int_{-\infty}^{x} \exp\,(-y^2/2) dy \right| < \varepsilon \, , \tag{17}$$

where

$$d^2 = \sum_{i=1}^{n} c_{in}^2 \int_0^1 (\phi(t) - \bar{\phi})^2 dt$$

and

$$\bar{\phi} = \int_0^1 \phi(t) dt \, .$$

To determine the value of $\mathrm{E}\,S$ we have the following theorem by Hoeffding [10].

THEOREM 2 (Hoeffding). *Under Assumptions* 1, 2 *and* 3, *Hájek's result* (17) *holds when* E S *is replaced by*

$$\mu=\sum_{i=1}^{n} c_{in}\int_{-\infty}^{\infty} \phi(H(t))dF(t,\, x_i\beta_1,\, y_i\beta_2)\ , \tag{18}$$

where

$$H(t)=\frac{1}{n}\sum_{i=1}^{n} F(t,\, x_i\beta_1,\, y_i\beta_2)\ .$$

Since Hájek's useful result is conditional upon any sequence z_i satisfying the three assumptions, an immediate corollary can be seen by making the following additional definition and assumption. Let $F^{(\infty)}$ be the sigma field of all possible series of outcomes of $z_1, z_2, \cdots$. Let $A_n \in F^{(\infty)}$ be such that for $n \geqq n_0$, and fixed η all possible series of outcomes whose first n components do not satisfy Assumption 2 are elements of A_n^c (complement of A_n). Due to the strong law of large numbers, the measure of A_n is greater than $1-a_n$ for some positive sequence of numbers a_n monotonically decreasing to zero as n increases.

ASSUMPTION 4. A_n is compact for any η.

COROLLARY. *Under Assumptions* 1, 3 *and* 4, *equation* (17) *holds unconditional upon the outcomes of* Z. *In addition,* E S *may be replaced by Hoeffding's centering constant* (18).

OUTLINE OF PROOF. For a fixed η, a value n_0 can be fixed such that given some series of possible outcomes, say $\delta \in A_n$, $n>n_0$, the theorem above holds for any fixed $\varepsilon>0$ with E $S=\mu$.
Assumption 4 implies that for $\varepsilon>0$, there exists an N_ε such that for $n>N_\varepsilon$

$$\sup_{\delta\in A_n}\sup_{w}\left|\Pr\left(S_\delta-\mathrm{E}\, S_\delta<wd_\delta\right)-(2\pi)^{-1/2}\int_{-\infty}^{w}\exp\left(-y^2/2\right)dy\right|<\varepsilon\ . \tag{19}$$

Thus

$$\Pr\left\{\sup_{w}\left|\Pr\left(S-\mathrm{E}\, S<wd\right)-(2\pi)^{-1/2}\int_{-\infty}^{w}\exp\left(-y^2/2\right)dy\right|<\varepsilon\right\}>1-a_n\ . \tag{20}$$

Hence asymptotically the linear rank statistic is distributed as a normal variable since $a_n \rightarrow 0$ as $n \rightarrow \infty$.

5. Testing a composite hypothesis

Using a rank statistic, the hypothesis that the score function $a(r)$ in (8) associated with $z_{(i)\zeta}$ is 'independent' of the value of $z_{(i)\zeta}$ may be tested in small samples by enumeration of the values of $U_\zeta(\beta)$ and invoking the permutation argument to determine the exact size of the test (see Hájek and Šidák [9]). It is important to note, however, that the 'independence' of $z_{(i)\zeta}$ with $a(R_i)$ is not equivalent to the hypothesis that $\beta_\zeta=0$ unless $Z_{(i)1}$ and $Z_{(i)2}$ are independent. Large sample approximate chi-square tests can also be formed. (See e.g., Puri and Sen [14]).

To test the composite hypothesis $H_1: \beta_1=0$ we would anticipate from Cox's global formulation that a test statistic would be of the form

$$S_{z'\boldsymbol{\beta}}=U_1(\boldsymbol{\beta})=\sum_{i=1}^{n}(x_i-\bar{x})a_{z'\boldsymbol{\beta}}(R_i) \tag{21}$$

as defined above by setting $\boldsymbol{\beta}=0$.

In order to obtain a test statistic, the following assumption limiting the degree of dependency of X and Y is made.

ASSUMPTION 5. The quantity

$$\mathrm{E}\left\{\sum_{i=1}^{n}\int_0^1 \phi(H(x))(x_i-\mu_x)(y_i-\mu_y)^2 f_y^{(2)}(t,0,\mu_y\beta_2)dt\right\} \tag{22}$$

is bounded as n increases, where

$$f_y^{(2)}=\frac{\partial^3 F}{\partial t\partial y^2}\ .$$

LEMMA 1. *Under Assumptions* 1, 3, 4 *and* 5, *the statistic* $Q=S_0^2/d^2$ *has asymptotically a (central) chi-square distribution with one degree of freedom when* H_1 *is true, where* $S_0=S_{z'\boldsymbol{\beta}}$ *evaluated at* $\boldsymbol{\beta}=0$, *and* d^2 *is as defined in Theorem* 1.

PROOF. Recall for fixed z and ε, there exists an n_0 such that if $n>n_0$ then

$$\sup_w\left|\Pr(S\leqq wd)-(2\pi)^{-1/2}\int_{-\infty}^{w}\exp\left(-\frac{(t-\mu_z)^2}{2}\right)dt\right|<\varepsilon\ , \tag{23}$$

where

$$\mu_z=\sum_{i=1}^{n}\frac{(x_i-\bar{x})}{d}\int_0^1\phi(H(t))dF(t,0,y_i\beta_2)\ .$$

By SLLN for large n we may replace d with $d_1=\sqrt{n\sigma_x^2}\cdot\left(\int_0^1(\phi(t)-\bar{\phi})^2dt\right)^{1/2}$, where $\sigma_x^2=\mathrm{Var}\,X$, and we may also replace $H(t)$ with $H_1(t)=F(t,0,\mu_y\beta_2)$ provided H_1 is true. To make (23) unconditional on the outcome of Z we have, suppressing η,

$$\sup_w \left| \sqrt{2\pi} \int_{A_n} \Pr(S \leqq w d_1) dG(x, y) - \int_{A_n} \int_{-\infty}^{w} \exp\left(-\frac{(t-\mu_z)^2}{2}\right) dt dG(x, y) \right| < \varepsilon \tag{24}$$

since ε is independent of z. Clearly we can change the order of integration in the second term. Expanding in a Taylor series about 0 we have

$$\begin{aligned} \int_{-\infty}^{w} \int_{A_n} \exp\left(-\frac{(t-\mu_z)^2}{2}\right) dG(x, y) dt &= \int_{-\infty}^{w} \int_{A_n} \Big\{ \exp(-t^2/2) + \mu_z t \exp(-t^2/2) \\ &\quad + \mu_z^2 \frac{t^2-1}{2} \exp\left(-\frac{(t-\theta)^2}{2}\right) \Big\} dG(x, y) dt , \end{aligned} \tag{25}$$

where $\theta \in (0, u_z)$. We now expand the expression μ_z about $\mathrm{E}Y = \mu_y$ to get

$$\begin{aligned} \mu_z = \frac{1}{\sqrt{n\sigma_x^2}} \Big\{ &\sum (x_i - \bar{x}) \int_{-\infty}^{\infty} \phi(H_1(t)) f(t, 0, \mu_y \beta_2) dt \\ &+ \sum (x_i - \bar{x})(y_i - \mu_y) \int_{-\infty}^{\infty} \phi(H_1(t)) f^{(1)}(t, 0, \mu_y \beta_2) dt \\ &+ \sum (x_i - \bar{x})(y_i - \mu_y)^2 \int_{-\infty}^{\infty} \phi(H_1(t)) f^{(2)}(t, 0, \theta_i) dt \Big\} \end{aligned} \tag{26}$$

where θ_i is between μ_y and y_i, $f^{(1)}(\cdot) = \partial f(\cdot)/\partial y$ and $f^{(2)}(\cdot) = \partial^2 f(\cdot)/\partial y^2$. Substituting (26) into (25) and integrating with respect to x and y we get

$$\int_{-\infty}^{w} \int_{A_n} \exp\left(-\frac{(t-\mu_z)^2}{2}\right) dG(x, y) = \int_{-\infty}^{w} \exp(-t^2/2) dt + \sum_{i=1}^{4} b_{in} \tag{27}$$

where

$$\begin{aligned} b_{1n} = \int_{-\infty}^{w} \int_{A_n} \frac{1}{\sqrt{n\sigma_x^2}} \sum (x_i - \bar{x})(y_i - \mu_y)^2 \left\{ \int_{-\infty}^{\infty} \phi(H_1(v)) f^{(2)}(v, 0, \theta_i) dv \right\} t \\ \cdot \exp\left(\frac{-t^2}{2}\right) dG(x, y) dt , \end{aligned}$$

$$b_{2n} = -a_n \int_{-\infty}^{w} \exp(-t^2/2) dt , \qquad (a_n \text{ is as described before}) ,$$

$$\begin{aligned} b_{3n} = \int_{-\infty}^{w} \int_{A_n} \frac{\sum (x_i - \bar{x})}{n} \left\{ \int_{-\infty}^{\infty} \phi(H_1(v)) f(v, 0, y_i \beta_1) dv \right\} \frac{t^2-1}{2} \\ \cdot \exp\left(-\frac{(t-\theta)^2}{2}\right) dG(x, y) dt , \end{aligned}$$

and

$$b_{4n}=\int_{-\infty}^{w}\int_{A_n^c}\sum(x_i-\bar{x})(y_i-\mu_y)\left\{\int_{-\infty}^{\infty}\phi(H_1(v))f(v,0,\mu_y\beta_2)dv\right\}t$$
$$\cdot\exp(-t^2/2)dG(x,y)dt\ .$$

Now as n increases, $b_{1n}\to 0$ by Assumption 5 and the fact that X and Y are bounded random variables. $b_{2n}\to 0$ since $a_n\to 0$ as n increases. $b_{3n}\to 0$ by the strong law of large numbers and the fact that X and Y are bounded. $b_{4n}\to 0$ as n increases since X and Y are bounded, and the measure of A_n^c goes to zero. Consequently, from (27), for large n, S_0/d has approximately a standard normal distribution. Hence the result.

At this point several remarks are in order concerning Assumption 5.

Remark 1. If essentially all of the dependency between two possible concomitant random variables X^* and Y^* can be accounted for by linear correlation, then the transformation of X^* and Y^* by their covariance matrix to give X and Y will satisfy Assumption 5.

Remark 2. Often the distribution of concomitant variables is well known before the actual experiment is done. For example, a clinic may be able to determine the distribution of age, sex, and perhaps, the extent of the disease of incoming patients. Although the effect of a treatment on survival will be unknown beforehand, the distribution of concomitants such as these may not be. In many of these cases, transformations exist which satisfy the conditions of Remark 1.

Remark 3. Cox proposed a global chi-square statistic for testing the hypothesis H_0. For asymptotic power considerations one often chooses a sequence of Pitman alternatives to H_0 indexed by n of the form

$$\text{ALT}_n:\quad \beta_1=\frac{\gamma_1}{\sqrt{n}}\ ,\quad \beta_2=\frac{\gamma_2}{\sqrt{n}}\ ,$$

where γ_1 and γ_2 are unknown, non-null constants. If the hypothesis H_1 is reformulated such that β_2 is either $\gamma_2/\sqrt{n}$ or null under either H_1 or its alternative, then Assumption 5 of boundedness is satisfied.

Remark 4. Often one of the 'concomitants' is the treatment administered. In this case treatments are randomized across the second concomitant (or a balanced blocking of the second concomitant), so that Assumption 5 is satisfied. This case resembles the conditions given by Ogawa in several papers (see e.g., Ogawa [13] for references).

Cox's global test statistic is based upon the 'information' matrix $I(0)$ as defined in equation (5). The manner in which the linear rank statistics are generated using the proportional hazard assumption (1) makes it much easier to calculate the variance estimate based upon (5) rather than using d^2. Thus, a simpler statistic for H_1, and one more in the vein of Cox's global chi-square, would be

$$Q_2 = U_1^2(0)/I_{11}(0) \ . \tag{28}$$

The asymptotic equivalence of this statistic with Q_1 can be shown assuming that the 'likelihood' estimates which maximize L are consistent. In this case, for a fixed sequence of outcomes $\boldsymbol{z}$,

$$\mathrm{Var}\,(U(\boldsymbol{\beta}_0)) = I(\boldsymbol{\beta}_0) + o(1) \tag{29}$$

when $\boldsymbol{\beta}_0$ obtains. Under Assumption 5, using the strong law of large numbers, we see that (29) implies $\mathrm{E}\,(I_{11}(0) - d^2)^2 \to 0$ when $\beta_1 = 0$ (see Theorem 3.1 of Puri and Sen [15]).

6. Application to survival problems

To apply the above results, consider a failure model with concomitants as described in Section 2. Let the observed concomitants be (x_i^*, y_i^*), $i=1,\cdots,n$ with associated failure times t_i, $i=1,\cdots,n$. Let $\hat{V}$ be the observed covariance of (x_i^*, y_i^*), and let P be a lower triangular matrix such that $PP' = \hat{V}^{-1}$. We assume that the effects of these concomitants can be related to the failure time distribution as given in the Cox model

$$\lambda(t) = \lambda_0(t) \exp\,(x^*\beta_1 + y^*\beta_2) \ .$$

The resulting vector of rank statistics is

$$U^*(0) = \begin{bmatrix} U_1^*(0) \\ U_2^*(0) \end{bmatrix} = \begin{bmatrix} \sum x_i^* a(R_i) \\ \sum y_i^* a(R_i) \end{bmatrix} . \tag{30}$$

Premultiplying by P' we have

$$P'U^*(0) = \begin{bmatrix} \sum x_i a(R_i) \\ \sum y_i a(R_i) \end{bmatrix} \tag{31}$$

where $x_i = p_{11}x_i^* + p_{21}y_i^*$ and $y_i = p_{22}y_k^*$. The effect of this transformation is that (asymptotically) X_i and Y_i are uncorrelated thus satisfying an assumption from Section 4. In addition, the linear rank statistic $U_1(0) = \sum x_i a(R_i)$ corresponds to the following reparametrization of the Cox hazard function;

$$\lambda(t) = \lambda_0(t) \exp\,(x\theta_1 + y\theta_2) \ , \tag{32}$$

where

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = P^{-1} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} p_{11}^{(-1)}\beta_1 \\ p_{21}^{(-1)}\beta_1 + p_{22}^{(-1)}\beta_2 \end{bmatrix} .$$

Hence using the result of Section 4 we can test the hypothesis $H_2: \theta_1 = 0$ under the lemma. However, for any given set of outcomes of Z we note that $\theta_1 = 0$ if and only if $\beta_1 = 0$. Hence for large sample sizes $Q_2 = U_1^2(0)/d_1^2$ is asymptotically distributed as chi-square when H_1 (or H_2) is true (under the assumptions of Lemma 1).

It is important to note that although Q_2 can be used to test if $\beta_1 = 0$, the quantity $U_2^2(0)/d_2^2$ derived in a similar manner has an approximate chi-square distribution if $\theta_2 = 0$ and Assumption 5 is satisfied when x and y are reversed. This is not equivalent to $\beta_2 = 0$ unless, additionally, either $\beta_1 = 0$ or $p_{21}^{(-1)} = 0$.

Using Lemma 1 we thus see that a simple test procedure for testing $H_1: \beta_1 = 0$ can be formulated based upon Hájek's results on linear rank statistics. Q_2 can be formed using matrix multiplication procedures and, for large samples, has an approximate chi-square distribution when $\beta_1 = 0$ obtains. This asymptotic result does not require an estimate of β_2 but instead depends upon Assumption 5. An indication of the necessary sample size and the magnitude of the effects of Assumption 5 were obtained from the following Monte Carlo study. For samples of size 8 and 16, the concomitant Y was selected from a truncated normal distribution with mean 1 and variance 1. Truncation was implicitly accomplished by the fact that the machine has finite word length (10 digits) in generating uniform random numbers. Conditional on the outcome of Y, the concomitant X was selected from a truncated normal distribution with mean $\mu_{x/y}$ and variance 1, where

$$\mu_{x/y} = 2 + \rho(y-1) + q(y-1)^2 . \tag{33}$$

The parameters ρ took the values 0, .2, and .4 and q took values 0, .2, .4 and .8. With each set of "observed" concomitants (x, y), the failure time was simulated. The exponential hazard model

$$\lambda(t) = \exp(x\beta_1 + y\beta_2) \tag{34}$$

was used where $\beta_2 = 1$ and β_1 took on values from 0 up. The test for the hypothesis that $\beta_1 = 0$ was calculated as above. This was repeated 1000 times for each set of β_1, ρ and q values.

Table I gives the proportion of times the test statistic was significant at the .05 significance level of the chi-square distribution. For $q = 0$ either X and Y are independent or only linearly correlated and hence satisfy Assumption 5. For $q = .2$, $q = .4$, or $q = .8$, Assumption 5

Table I Simulated power for the hypothesis H_0: $\beta_1=0$, for various values of β_1 when $\beta_2=1$. ρ represents correlation, q represents quadratic effect.

$N=8$

β_1	$q=0$			$q=.2$		
	$\rho=0$	$\rho=.2$	$\rho=.4$	$\rho=0$	$\rho=.2$	$\rho=.4$
0	.027	.025	.022	.027	.024	.028
.5	.070	.057	.049	.092	.056	.052
1.0	.222	.199	.159	.221	.200	.162
1.5	.417	.348	.294	.419	.363	.272
2.0	.587	.503	.426	.594	.484	.436
2.5	.667	.586	.477	.706	.602	.471
3.0	.768	.667	.538	.770	.641	.536

β_1	$q=.4$			$q=.8$		
	$\rho=0$	$\rho=.2$	$\rho=.4$	$\rho=0$	$\rho=.2$	$\rho=.4$
0	.028	.029	.022	.021	.029	.026
.5	.071	.080	.073	.112	.098	.100
1.0	.262	.218	.181	.334	.308	.252
1.5	.470	.395	.329	.513	.466	.461
2.0	.593	.514	.468	.617	.590	.504
2.5	.722	.624	.517	.709	.670	.622
3.0	.758	.667	.593	.751	.727	.638

$N=16$

β_1	$q=0$			$q=.2$		
	$\rho=0$	$\rho=.2$	$\rho=.4$	$\rho=0$	$\rho=.2$	$\rho=.4$
0	.015	.015	.012	.012	.018	.017
.5	.172	.137	.114	.150	.138	.097
1.0	.591	.504	.379	.564	.483	.369
1.5	.869	.785	.683	.847	.770	.635
2.0	.947	.898	.839	.952	.895	.802
2.5	.979	.943	.895	.966	.946	.893

β_1	$q=.4$			$q=.8$		
	$\rho=0$	$\rho=.2$	$\rho=.4$	$\rho=0$	$\rho=.2$	$\rho=.4$
0	.019	.022	.019	.021	.017	.022
.5	.171	.141	.115	.242	.208	.176
1.0	.630	.539	.454	.726	.672	.579
1.5	.874	.804	.719	.901	.846	.812
2.0	.945	.911	.823	.950	.930	.892
2.5	.970	.959	.894	.974	.954	.930

is violated. The effect of this on the size and power of the composite hypothesis can be readily seen in Table I.

The theoretical asymptotic power of the statistic, under a set of suitable alternatives, provides a useful comparison. Consider, for example, the set of Pitman type alternatives

$$(35)\qquad H_A:\ \beta_1=\gamma_1/\sqrt{n}\ .$$

To satisfy Assumption 5 let us assume that under H_1 or H_A, $\beta_2=\gamma_2/\sqrt{n}$ where γ_1 (non-null) and γ_2 are constants. If we expand $f(t, x_i\gamma_1/\sqrt{n}, y_i\gamma_2/\sqrt{n})$ about $\beta_1=0$ we have

$$(36)\qquad f(t, x_i\gamma_1/\sqrt{n}, y_i\gamma_2/\sqrt{n}) \\ =f(t, 0, y_i\gamma_2/\sqrt{n})+x_i\gamma_1/\sqrt{n}\cdot\frac{\partial f(t, a, y_i\gamma_2/\sqrt{n})}{\partial a}\bigg|_{a=0}+o(1)\ .$$

Proceeding as in the proof of Lemma 1 writing $f'=\partial f(t, a, b)/\partial a$ we have

$$(37)\qquad \mathrm{E}\frac{S_0}{d}\doteq\int_{A_n}\frac{1}{d}\sum(x_i-\bar{x})\int_{-\infty}^{\infty}\phi(H_1(t))\cdot\Big\{f(t, 0, y_i\gamma_2^2/\sqrt{n}) \\ +\frac{x_i\gamma_1}{\sqrt{n}}f'(t, 0, y_i\gamma_2/\sqrt{n})+o(1)\Big\}dtdG(x, y)$$

$$(38)\qquad \doteq\int_{A_n}\frac{\gamma_1}{d\sqrt{n}}\sum(x_i-\bar{x})x_i\int_{-\infty}^{\infty}\phi(H_1(t))\Big\{f'(t, 0, \mu_y\gamma_2\sqrt{n}) \\ +(y_i-\mu_y)\frac{\partial f'}{\partial y}(t, 0, y_i\gamma_2/\sqrt{n})|_{y_i=\mu_y}+o(1)\Big\}dtdG(s, y)\ .$$

Assuming that we can pass the limit through the two integrals, under the assumption that $(\partial/\partial y)f'(t, 0, \mu_y\gamma_2/\sqrt{n})$ is bounded as n increases, we have

$$(39)\qquad \lim_{n\to\infty}\mathrm{E}\frac{S_0}{d}=\frac{\gamma_1\sigma_x}{\left(\int_0^1(\phi(t)-\bar{\phi})^2dt\right)^{1/2}}\cdot\int_{-\infty}^{\infty}\phi(H_1(t))f'(t, 0, 0)dt\ .$$

In our case, with no censoring, $\phi(t)=\log(1-t)$. Hence, for the simulated example

$$(40)\qquad \lim_{n\to\infty}\mathrm{E}\frac{S_0}{d}=\gamma_1\sigma_x\ .$$

The resulting non-centrality parameter is $\gamma_1^2\sigma_x^2/2$. Table II gives the resulting expected power for the simulations in Table I based upon the above assumptions. It is reasonable to expect that with all of the large sample approximations made, the asymptotic power and the actual power

Table II Expected power under Pitman alternatives

$S_1/\sqrt{n}$	$n=8$		$n=16$	
	$S_1^2/2$	Power	$S_1^2/2$	Power
.5	1	.292	2	.516
1.0	4	.808	8	.979
1.5	9	.989	18	.999
2.0	16	.999	32	.999
2.5	25	1.000	50	1.000

will be substantially different. Comparisons of Table I with Table II shows this to be the case. Although Table I indicates that the test procedure may be useful, the power we would expect if all the linear approximations were correct is considerably higher. (Further simulations indicate that for $n=32$ the discrepancy between observed and theoretical asymptotic power is nearly negligible.)

7. Discussion

Use of the above results can be summarized by three points. First, if the number of concomitants in a study is very large, the above results suggest a method for eliminating possibly many of these concomitants before iterative techniques of parameter estimation are applied. This initial step does not require an estimate of parameters associated with nuisance concomitants but rather some degree of independence as given in Assumption 5. This assumption appears to be similar to, but stronger than, Cox's [4] requirement on U_j (see p. 274).

Secondly, sufficient conditions for the concomitants are given in Assumptions 1 through 4 to insure the quadratic forms Q_1 and Q_2 to be chi-square distributed for large samples. (Assumption 5 is necessary to obtain a null noncentrality parameter for Q_2 when H_1 obtains). Several papers addressing the Cox model allow the concomitants to be time-dependent. The assumption that $Z_1, Z_2, \cdots$ be identically distributed can be relaxed to provide for this case. However, the stability assumption that must be used as a replacement must include Assumption 2. It is the author's experience that this assumption is more prohibitive than it seems.

Kalbfleisch and Prentice [11] have proposed a marginal likelihood formulation of the proportional hazard model which, when no ties or censoring are present, is identical to Cox's model. When censoring or ties are present, however, the two formulations differ. The mathematical justification given by Kalbfleisch and Prentice suggests that in this case their model be used in lieu of Cox's. Unfortunately, the estimating equations in their model can be considerably more difficult. (With 10 failures at the same time the estimating equations to be minimized will contain literally millions of summands.) The procedure described above can be applied to this new likelihood, however, with about the same simplicity as with the Cox model. Thus estimating such a model as proposed by Kalbfleisch and Prentice could be greatly simplified if the procedure described above was used to initially discard unimportant parameters.

RADIATION EFFECTS RESEARCH FOUNDATION, HIROSHIMA, JAPAN*

REFERENCES

[1] Bhapkar, V. P. (1961). A nonparametric test for the problem of several samples, *Ann. Math. Statist.*, **32**, 1108-1117.

[2] Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain non-parametric test statistics, *Ann. Math. Statist.*, **29**, 972-994.

[3] Cox, D. R. (1972). Regression models in life tables (with discussion), *J. R. Statist. Soc.*, B, **34**, 187-220.

[4] Cox, D. R. (1975). Partial likelihood, *Biometrika*, **62**, 269-276.

[5] Crowley, J. (1974). A note on some recent likelihoods leading to the log rank test, *Biometrika*, **61**, 533-538.

[6] Downton, F. (1972). Discussion of Cox's regression models in life tables, *J. R. Statist. Soc.*, B, **34**, 202-205.

[7] Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance, *J. Amer. Statist. Ass.*, **32**, 675-701.

[8] Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives, *Ann. Math. Statist.*, **39**, 325-346.

[9] Hájek, J. and Sidák, Z. (1967). *Theory of Rank Tests*, Academic Press, New York.

[10] Hoeffding, W. (1973). On the centering of a simple linear rank statistic, *Ann. Statist.*, **1**, 54-66.

[11] Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model, *Biometrika*, **60**, 267-278.

[12] Kruskal, W. H. (1952). A nonparametric test for the several sample problems, *Ann. Math. Statist.*, **23**, 525-540.

[13] Ogawa, J. (1974). *Statistical Theory of the Analysis of Experimental Designs*, Marcel Dekker, Inc., New York.

[14] Puri, M. L. and Sen, P. K. (1973). A note on asymptotic distribution-free tests for subhypotheses in multiple linear regression, *Ann. Statist.*, **1**, 553-556.

[15] Puri, M. L. and Sen, P. K. (1969). A class of rank order tests for a general linear hypothesis, *Ann. Math. Statist.*, **40**, 1325-1343.

---

* Now at Texas A & M University.