The Canadian Journal of Statistics
Vol. 26, No. 1, 1998, Pages 57–68
La Revue Canadienne de Statistique

Restricted tests for testing independence of time to failure and cause of failure in a competing-risks model*

Richard DYKSTRA, Subhash KOCHAR and Tim ROBERTSON

The University of Iowa, Indian Statistical Institute, and *The University of Iowa*

Key words and phrases: Isotonic regression, chi-bar-square distribution, likelihood-ratio dependence, multinomial distribution.
AMS 1991 subject classifications: Primary 62905, 62410; secondary 62E20.

ABSTRACT

We consider the competing-risks problem without making any assumption concerning the independence of the risks. Maximum-likelihood estimates of the cause-specific hazard rates are obtained under the condition that their ratio is monotone. We also consider the likelihood-ratio test for testing the proportionality of two cause-specific hazard rates against the alternative that the ratio of these two hazard rates is monotonic. This testing problem is equivalent to testing independence against likelihood-ratio dependence of the time to failure and the cause of failure in the competing-risks setup. We allow for random censoring on the right. The asymptotic null distribution of the test statistic is obtained and is found to be of the chi-bar-square type. The problem is extended to the case of more than two risks. A numerical example is given to illustrate the procedure.

RÉSUMÉ

Nous considérons le problème des risques concurrents sans formuler d'hypothèses au sujet de l'indépendance des risques. Des estimés du maximum de vraisemblance des taux de hasard spécifiques à la cause sont obtenus sous la condition que leur ratio est monotone. Nous considérons également le test du ratio de vraisemblance afin de tester la proportionalité de deux taux de hasard spécifiques à la cause contre l'alternative selon laquelle le ratio de ces deux taux de hasard est monotone. Ce problème de test est équivalent à tester l'indépendance contre la dépendance au ratio de vraisemblance du temps avant l'échec et la cause de l'échec dans l'organisation des risques concurrents. Nous permettons la censure aléatoire à droite. La distribution asymptotique nulle de la statistique de test est obtenue et est découverte comme étant de type chi-barre carré. Le problème est étendu au cas de plus de deux risques. Un exemple numérique est donné afin d'illustrer la procédure ici développée.

1. INTRODUCTION

In competing-risks survival analysis each unit under study is exposed to a number of different risks, but the actual failure of each unit results from just one of those risks. Suppose that there are only two possible causes of failure, labeled 1 and 2, and that the notional times to failure of a unit under those two risks are denoted by the random variables X and Y, respectively. We assume that $\mathrm{pr}[X = Y] = 0$, so that the cause of failure is 1 if and only if $X < Y$. The variables X and Y cannot be observed. The observed information is the time of failure $T = \min(X, Y)$ and the corresponding cause of death C.

*This research was supported by NSF Grant DMS 95-04891.

Later, we will allow censoring information of the form $T > t$, but for current purposes of discussion, we will assume that no censoring takes place. Such data occur frequently in studies of the relative risk of death from various causes such as cancer and heart disease and they also occur in multicomponent series systems in reliability. See Cox (1959) for more applications of such a model.

It is common in the literature to do the statistical analysis assuming that X and Y are stochastically independent. However, as noted by Gail (1975), among others, this assumption is often unrealistic because the risks are acting within the same environment. This problem is compounded by the fact that independence of X and Y cannot be tested on the basis of data observed on T and C (cf. Cox 1959).

To quantify the risks of failure from the various causes, the notion of cause-specific hazard rates is used. This is an extension of the usual concept of hazard rate. We assume that the time to failure, T, is discrete with k possible values labeled $t_1, t_2, \ldots, t_k$. Continuous data would be collected into a finite number of cells and analyzed as if they were discrete. In that case $t_1, t_2, \ldots, t_k$ would be points identifying those cells.

The cause-specific hazard rate corresponding to the ith cause is

$$g_i(t_j) = \mathrm{pr}(C = i,\ T = t_j | T \geq t_j). \tag{1.1}$$

Thus $g_i(t_j)$ is the rate of failure at time t_j from the ith cause given that the unit has been working up to time t_j. It is easy to see that if X and Y are independent, then g_1 and g_2 are simply the hazard rates corresponding to the marginal distributions of X and Y. However, even without independence, the sum, $g_1(t_j) + g_2(t_j)$ is equal to the hazard rate $r_T(t_j)$ of T. Prentice *et al.* (1978) show that only probabilities expressible as functions of the g_i's can be estimated from the observable data.

It is often of interest to compare the relative risks of failure from various causes over time. Aly, Kochar and McKeague (1994) and Dykstra, Kochar and Robertson (1995b) have proposed distribution-free tests for testing the equality of cause-specific hazard rates against ordered alternatives. In this paper, we consider a somewhat different problem. We are concerned with the question of whether the ratio of cause-specific hazard rates is constant versus the alternative that the ratio changes monotonically.

Such a one-sided alternative is often of great interest. For example, it is generally recognized that the cause-specific hazard rate for heart disease in women increases much faster than those for many other risks beyond the onset of menopause. This fact has led to significant changes in medical care for women in this age group.

On the basis of a random sample, $(C_1, T_1), (C_2, T_2), \ldots, (C_n, T_n)$ from (C, T) we wish to test the null hypothesis

$$\mathcal{H}_0 : g_2(t_j)/g_1(t_j) \text{ is constant} \tag{1.2}$$

against the alternative $\mathcal{H}_1 - \mathcal{H}_0$, where

$$\mathcal{H}_1 : g_2(t_j)/g_1(t_j) \text{ is nondecreasing in } t_j. \tag{1.3}$$

Another way of expressing these hypotheses is to define the functions $\pi_i(t_j) = \mathrm{pr}(C = i | T = t_j)$, $i = 1, 2$, $j = 1, 2, \ldots, k$. The functions $\pi_i(t_j)$ are related to the cause-specific hazard rates by $\pi_i(t_j) = g_i(t_j)/r_T(t_j)$, $i = 1, 2$, and $\pi_1(t_j) + \pi_2(t_j) = 1$. In terms of the π_i's the above hypotheses are

$$\mathcal{H}_0 : \pi_1(t) \text{ and } \pi_2(t) \text{ are constant in } t$$

and

$$\mathcal{H}_1 : \pi_2(t) \text{ is nondecreasing and } \pi_1(t) \text{ is nonincreasing in } t.$$

Thus, given that a death has occurred at time t, the conditional probability that it is due to cause 2 remains constant in t under $\mathcal{H}_0$ but is nondecreasing in t (and somewhere increasing) under the alternative $\mathcal{H}_1$.

Let $p_{ij} = \text{pr}(C = i, T = t_j)$, $i = 1, 2$, $j = 1, 2, \ldots, k$, so that the cause-specific hazard rates are given by

$$g_i(t_j) = \frac{p_{ij}}{\sum_{i=1}^{2} \sum_{l=j}^{k} p_{il}}. \tag{1.4}$$

Then the hypotheses written in terms of the p_{ij} are

$$\mathcal{H}_0 : \frac{p_{21}}{p_{11}} = \frac{p_{22}}{p_{12}} = \cdots = \frac{p_{2k}}{p_{1k}} \tag{1.5}$$

and

$$\mathcal{H}_1 : \frac{p_{21}}{p_{11}} \leq \frac{p_{22}}{p_{12}} \leq \cdots \leq \frac{p_{2k}}{p_{1k}}. \tag{1.6}$$

Thus, $\mathcal{H}_0$ is equivalent to the assumption that T and C are independent, while $\mathcal{H}_1$ assumes that T and C are likelihood-ratio-dependent (see Lehmann 1966). The observation that T and C are independent under the assumption analogous to (1.5) for the continuous case was made by Kochar and Proschan (1991). Their observation extends the well-known characterization result of Allen (1963), Armitage (1959), and Sethuraman (1965) concerning the proportionality of hazard rates of two independent risks in the competing-risks model.

Although this paper only considers the discrete situation, there are connections to the continuous case. In particular, if one interprets maximum likelihood in the generalized sense of Kiefer and Wolfowitz (1956) for the types of constraints discussed in this paper, probability will only occur on the observed values. Thus the MLEs of the cause-specific hazard rates given in Section 2 will yield (generalized) MLEs where the family of interest consists of all bivariate distributions with well-defined cause-specific hazard rates (including the continuous ones).

Though these generalized MLEs of the cause specific hazard rates must correspond to discrete distributions, one might hope that the associated CDFs (of the estimated casue specific hazard rates) would converge to the CDFs corresponding to the true cause specific hazard rates when the ratio of the true rates satisfies the monotonicity condition given in (1.3) even in the continuous case. This is not obvious, since the generalized MLEs under similar orders (i.e., hazard rate order or uniform stochastic order) need not have this property in the continuous case (Rojo and Samienego, 1991).

Fortunately, a variation of the proof given in Dykstra, Kochar and Robertson (1995a) can be employed to show the desired consistency under very general conditions regardless of whether the cause-specific hazard rates are discrete or continuous. This would indicate that the continuous case is a limit of the discrete situation (in a natural sense). Of course, the asymptotic distribution theory derived in Section 3 for hypothesis testing would not be appropriate for the continuous case unless grouping was done (similar to the chi-square goodness-of-fit tests).

These hypotheses are of interest in parametric models, too. For example, Gumbel's bivariate exponential distribution (Gumbel 1960) has joint survival function

$$\bar{F}(x, y) = \exp\{-(\lambda_1 x + \lambda_2 y + \lambda_3 xy)\}, \ , x, y \geq 0$$

for $\lambda_1, \lambda_2 > 0$, $\lambda_3 \geq 0$. Since the cause-specific hazard rates are $g_1(t) = \lambda_1 + \lambda_3 t$ and $g_2(t) = \lambda_2 + \lambda_3 t$, it easily follows that

$$\frac{g_1(t)}{g_2(t)} \text{ is strictly increasing in } t \quad \Leftrightarrow \quad \lambda_1 < \lambda_2 \text{ and } \lambda_3 > 0$$

while

$$\frac{g_1(t)}{g_2(t)} \equiv \text{constant} \quad \Leftrightarrow \quad \lambda_1 = \lambda_2 \text{ or } \lambda_3 = 0.$$

Dykstra, Kochar and Robertson (1996) have proposed distribution-free tests for testing the above type of alternatives in the case of continuous data.

In Section 2, we find maximum-likelihood estimates of the p_{ij}'s and the cause-specific hazard rates under the null and alternative hypotheses.

In Section 3, the asymptotic distribution of the likelihood-ratio test of $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ is derived. This asymptotic distribution is a version of the chi-bar-square distribution. Critical values for this test have been tabled, and it is well known that a test which restricts alternatives to those in $\mathcal{H}_1$ is significantly more powerful than a test for general omnibus alternatives. Extension to the case of more than two risks is also discussed in Section 3. A data set is analyzed in Section 4 illustrating the techniques developed in this paper.

2. MAXIMUM-LIKELIHOOD ESTIMATION

We allow for the possibility that our observations are censored on the right, but we do assume that the censoring mechanism acts independently of the failure times. In a competing-risks problem the various causes are censoring one another and we are not assuming that the two causes are acting independently. The assumption that the censoring mechanism acts independently may be unrealistic in some applications, but in general, the censoring mechanism gives no information about the relative seriousness of the causes, and the independence assumption is necessary to model the problem.

Our data set consists of a $2 \times k$ contingency table $(n_{ij})_{2\times k}$ together with a k-dimensional vector $\mathbf{d} = (d_1, d_2, \ldots, d_k)$. The number n_{ij} is the number of failures from cause i in the time period indexed by j, and d_j is the number of right-censored items during that time period.

Although the data could be displayed in a series of ordered 2×2 contingency tables specifying the number of failures at each time point along with the size of the risk set, our formulation is more concise.

This problem was studied by Armitage (1955) and Cox (1959), though their procedures were not based on any optimum or likelihood considerations. Lawless (1982) discussed the problem of testing $\mathcal{H}_0$ against the general unrestricted alternative. The problem of testing independence against positive dependence in a contingency table has been considered by many researchers, including Agresti, Wackerly and Boyett (1979), Grove (1980), Patefield (1982), and Lee (1989). Most of these tests do not allow for censoring and are actually testing independence against an alternative that is implied by likelihood-ratio dependence. The specific nature of the alternative may not be clear from the test, and it is also not clear how close these alternatives are to likelihood-ratio dependence. Most of these papers propose tests conditioned on the column totals $n_{1j} + n_{2j}$, $j = 1, 2, \ldots, k$. The conditional distribution of the n_{2j}'s given the column totals is that of k independent binomials with parameters $n_{1j} + n_{2j}, \theta_j$, where $\theta_j = p_{2j}/(p_{1j} + p_{2j})$. The null hypothesis of independence is equivalent to equality of the θ_j's, and the alternative of likelihood-ratio dependence is equivalent to $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$. Thus asymptotically, a conditional test in our problem reduces to the bioassay problem studied by Bartholomew (1959). We shall derive the unconditional likelihood-ratio test, study its asymptotic properties and discuss its relationship to the conditional test.

Let $\mathbf{P} = [p_{ij}]$ be the $2 \times k$ matrix of unknown probabilities associated with the n_{ij}'s. The likelihood function is proportional to

$$L(\mathbf{P}) = \prod_{j=1}^{k} \prod_{i=1}^{2} p_{ij}^{n_{ij}} \cdot \prod_{j=1}^{k} \left[\sum_{l=j+1}^{k} (p_{1l} + p_{2l}) \right]^{d_j}. \tag{2.1}$$

We make the change of variables

$$\theta_{1j} = p_{1j} + p_{2j} \quad \text{and} \quad \theta_{2j} = \frac{p_{2j}}{p_{1j} + p_{2j}} = \left(\frac{p_{1j}}{p_{2j}} + 1 \right)^{-1}, \tag{2.2}$$

so that

$$p_{1j} = \theta_{1j}(1 - \theta_{2j}) \text{ and } p_{2j} = \theta_{1j}\theta_{2j}, \qquad j = 1, 2, \ldots, k. \tag{2.3}$$

The constraint that $\mathbf{P}$ is a matrix of probabilities is equivalent to $0 \leq \theta_{ij} \leq 1$, $\sum_{j=1}^{k} \theta_{1j} = 1$. The hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_1$, written in terms of the θ_{ij}'s, are

$$\mathcal{H}_0 : \theta_{21} = \theta_{22} = \cdots = \theta_{2k},$$
$$\mathcal{H}_1 : \theta_{21} \leq \theta_{22} \leq \cdots \leq \theta_{2k} \qquad (\theta_{21} < \theta_{2k}).$$

The likelihood function, written in terms of the θ_{ij}'s, after some combining and rearranging of factors, becomes

$$L(\Theta) = \prod_{j=1}^{k} \theta_{2j}^{n_{2j}} (1 - \theta_{2j})^{n_{1j}} \cdot \prod_{j=1}^{k} \theta_{1j}^{n_{1j}+n_{2j}} \left(\sum_{l=j+1}^{k} \theta_{1l} \right)^{d_j}. \tag{2.4}$$

Because the constraints in $\mathcal{H}_0$ and $\mathcal{H}_1$ only involve the θ_{2j}'s, the two factors in the likelihood product can be maximized separately. The second factor in the likelihood (2.4),

$$\prod_{j=1}^{k} \theta_{1j}^{n_{1j}+n_{2j}} \left(\sum_{l=j+1}^{k} \theta_{1l} \right)^{d_j},$$

is similar to the expression for the likelihood on p. 873 of Dykstra, Kochar and Robertson (1991). Following their approach, we find the m.l.e.'s of

$$1 - \eta_j = \frac{\theta_{1j}}{\sum_{l=j}^{k} \theta_{1l}} \tag{2.5}$$

as

$$1 - \hat{\eta}_j = \frac{n_{1j} + n_{2j}}{m_j}, \qquad j = 1, 2, \ldots, k-1, \tag{2.6}$$

where

$$m_j = \sum_{l=j}^{k} (n_{1l} + n_{2l} + d_l)$$

is the number of items on test at the beginning of the jth time period.

From this, we obtain the m.l.e.'s of the θ_{1j}'s recursively by

$$\hat{\theta}_{11} = \frac{n_{11} + n_{21}}{m_1}, \tag{2.7}$$

$$\hat{\theta}_{1j} = \frac{n_{1j} + n_{2j}}{m_j}\left(1 - \sum_{r=1}^{j-1} \hat{\theta}_{1r}\right), \qquad j = 2, 3, \ldots, k-1, \tag{2.8}$$

$$\hat{\theta}_{1k} = 1 - \sum_{r=1}^{k-1} \hat{\theta}_{1r}.$$

Clearly, the maximum-likelihood estimates of the θ_{1j}'s remain the same under both $\mathcal{H}_0$ and $\mathcal{H}_1$.

The unrestricted maximum-likelihood estimator of θ_{2j} is

$$\hat{\theta}_{2j} = \frac{n_{2j}}{n_{1j} + n_{2j}}. \tag{2.9}$$

Using (2.3), (2.8) and (2.9), we obtain the unrestricted m.l.e.'s of p_{1j} and p_{2j} as

$$\hat{p}_{1j} = \hat{\theta}_{1j}(1 - \hat{\theta}_{2j}), \tag{2.10}$$

$$\hat{p}_{2j} = \hat{\theta}_{1j}\hat{\theta}_{2j} \tag{2.11}$$

Using (1.4) and the above estimates, we find that the unrestricted maximum-likelihood estimates of the cause-specific hazard rates at time t_j are

$$\hat{g}_i(t_j) = \frac{n_{ij}}{m_j}, \qquad i = 1, 2, \quad j = 1, 2, \ldots, k. \tag{2.12}$$

This is the Kaplan-Meier estimator of the cause-specific hazard rate and agrees with that obtained by Davis and Lawrence (1989).

The maximum-likelihood estimate of the common value of $\theta_{21}, \theta_{22}, \ldots, \theta_{2k}$ under $\mathcal{H}_0$ is

$$\hat{\theta}^0_{2j} = \frac{1}{n}\sum_{j=1}^{k} n_{2j}, \qquad j = 1, 2, \ldots, k,$$

where

$$n = \sum_{i=1}^{2}\sum_{j=1}^{k} n_{ij}.$$

Now we consider the problem of maximizing, under $\mathcal{H}_1$, the first factor in the likelihood (2.4). This is the bioassay problem as discussed in Robertson, Wright and Dykstra (1988) (denoted by RWD in the sequel). The solution is the least-squares projection of the vector $\hat{\boldsymbol{\theta}}_2 = (\hat{\theta}_{21}, \hat{\theta}_{22}, \ldots, \hat{\theta}_{2k})$ with weights $w_j = n_{1j} + n_{2j}$ onto the cone of nondecreasing vectors, $\mathcal{K}$. We denote these ordered values by $\theta^*_{2j} = P_w(\hat{\boldsymbol{\theta}}_{2\cdot}|\mathcal{K})_j$, $j = 1, 2, \ldots, k$ (see p. 99 of RWD). The *pool-adjacent violators algorithm* (PAVA) (see p. 8 of RWD) provides an easy method of obtaining the solution $(\theta^*_{21}, \theta^*_{22}, \ldots, \theta^*_{2k})$.

The maximum-likelihood estimators of $g_2(t_j)$ under $\mathcal{H}_1$ are given by

$$g^*_2(t_j) = \theta^*_{2j}(1 - \hat{\eta}_j)$$
$$= \frac{\theta^*_{2j}(n_{1j} + n_{2j})}{m_j}, \qquad j = 1, 2, \ldots, k$$

with a similar expression for $g^*_1(t_j)$.

3. HYPOTHESIS TESTING

We now consider our testing problem. We propose the test that rejects $\mathcal{H}_0$ in favour of $\mathcal{H}_1$ for large values of $T = -2 \ln \Lambda$, where Λ is the likelihood-ratio statistic. The asymptotic distribution of T can be derived in a manner similar to the derivation in Dykstra, Kochar and Robertson (1991). We give a brief sketch of the derivation here. The $\hat{\theta}^*_{1j}$'s and the $\hat{\theta}^0_{1j}$'s cancel in the likelihood ratio, so that Λ depends only on the $\hat{\theta}^*_{2j}$'s and the $\hat{\theta}^0_{2j}$'s. Expand the logarithms in this expression about $\ln \hat{\theta}_{2j}$ and $\ln (1 - \hat{\theta}_{2j})$ respectively, using a second-degree Taylor's expansion. The linear terms drop out because of Theorem 1.3.3 in RWD.

Using the corollary on the middle of page 376 and the last part of Corollary C on p. 379 of RWD, T can be rewritten (assuming $\mathcal{H}_0$ is true) as

$$T = \sum_{j=1}^{k} \left\{ \{\sqrt{n}(\hat{\theta}_{2j} - \theta_0)\}^2 \left(\frac{n_{1j}}{n\alpha_j^2} + \frac{n_{2j}}{n\beta_j^2} \right) - \{ P_w(\sqrt{n}(\hat{\theta}_{2\cdot} - \theta_0)|\mathcal{K})_j - \sqrt{n}(\hat{\theta}_{2j} - \theta_0)\}^2 \left(\frac{n_{1j}}{n\gamma_j^2} + \frac{n_{2j}}{n\delta_j^2} \right) \right\},$$

where θ_0 is the common value of the θ_{2j}'s under $\mathcal{H}_0$, the α's, β's, γ's, δ's come from Taylor's expansion, and the rest of the notation is the same as in Section 2.

Let $\hat{\mathbf{P}}$ be the $2 \times k$ matrix of relative frequencies ($\hat{p}_{ij} = n_{ij}/n$). Using the multivariate central limit theorem for $\sqrt{n}(\hat{\mathbf{P}} - \mathbf{P})$ and then the δ-method, the random vector $\sqrt{n}(\hat{\theta}_{2\cdot} - \theta_0)$ converges in distribution to the random vector $\mathbf{U} = (U_1, U_2, \ldots, U_k)^{\mathrm{T}}$, where U_j are independent, mean zero, normal random variables and

$$\mathcal{V}ar \; U_j = \frac{p_{1j}p_{2j}}{(p_{1j} + p_{2j})^3}.$$

Using the strong law of large numbers and standard facts concerning the isotonic regression operators, it follows that

$$T \xrightarrow{\mathcal{L}} \sum_{j=1}^{k} \left(P_w(Y|\mathcal{K})_j - \frac{\sum_{j=1}^{k} w_j Y_j}{\sum_{j=1}^{k} w_j} \right)^2 w_j,$$

where

$$Y_j = \frac{U_j}{\sqrt{\theta_{2j}(1 - \theta_{2j})}} \sim \mathrm{N}(0, w_j^{-1}) \quad \text{and} \quad w_j = p_{1j} + p_{2j}.$$

By applying the Corollary on p. 70 of RWD, the asymptotic distribution of T must be that of a chi-bar-square distribution. Specifically, for any t

$$\lim_{n \to \infty} \mathrm{pr}[T \geq t] = \sum_{l=1}^{k} P(l, k; w) \, \mathrm{pr}[\chi^2_{l-1} \geq t], \tag{3.1}$$

where χ^2_{l-1} is a chi-square variable with $l - 1$ degrees of freedom ($\chi^2_0 \equiv 0$) and $P(l, k; w)$ is the probability that the isotonic regression with weights $\mathbf{w}$ of k independent $\mathrm{N}(0, 1/w_i)$ random variables takes on l distinct values. This asymptotic distribution depends upon the unknown matrix $\mathbf{P}$ through the level probabilities $P(l, k; w)$. A least-favourable distribution can be found using Theorem 3.6.1 and the remark proceding it in RWD. We summarize our findings in the following theorem.

THEOREM 3.1. *If $T = -2 \ln \Lambda$, where Λ is the likelihood-ratio statistic for testing $\mathcal{H}_0$ vs. $\mathcal{H}_1 - \mathcal{H}_0$, then assuming that $\mathcal{H}_0$ is true, the asymptotic distribution of T is given by* (3.1). *The least favourable asymptotic distribution is given by*

$$\sup_{P \in \mathcal{H}_0} \lim_{n \to \infty} \operatorname{pr}[T \geq t] = \sum_{l=1}^{k} \binom{k-1}{l-1} \left(\frac{1}{2}\right)^{k-1} \operatorname{pr}[\chi^2_{l-1} \geq t]. \tag{3.2}$$

Although critical values for a test based upon this least favourable distribution can be found in Table 5.3.1 of RWD for $k = 2, \ldots, 15$, we do not recommend its use, since these tests are quite conservative.

A much better approach is to use critical values associated with the distribution on the right side of (3.1) with the $P(l, k; \mathbf{w})$ replaced by the equal-weights level probabilities. Critical values for this test are given in Table A.4 of RWD. Unless the true values of $w_1, \ldots, w_k$ have extreme variability (say the ratio of the largest to the smallest exceeds 4), this should give a very good approximation (see Chapter 3 of RWD).

Another possibility when k is small ($k \leq 5$) is to use critical values obtained from the right-side distribution of (3.1) where the w_j are replaced by the estimates $\hat{w}_j = \hat{p}_{1j} + \hat{p}_{2j}$. This procedure gives approximate asymptotic values and works extremely well (see the example in Section 4). Expressions for $P(l, k; w)$ can be calculated from formulae given in Section 2.4 of RWD. It is interesting that this is exactly the asymptotic distribution associated with the conditional test scenario discussed early in Section 2, which occurs when one approximates the binomial distributions by normals. This strongly supports the conclusion that the unconditional and conditional tests are very similar and nearly always lead to the same conclusion (at least when the amount of data is of reasonable size).

When k exceeds 5, we recommend using the pattern approximation discussed in Section 3.4 of RWD. This works very well (see the example in Section 4) in nearly all situations. Pillers, Robertson and Wright (1984) provide computer code for implementing this approximation.

The results in this paper extend directly to problems involving more than two causes of failure. (A disadvantage of the conditional approach is that it is not at all clear how to extend it to more than two causes of failure.) Suppose we have c causes of failure ($c \geq 2$), so that our data consist of a $c \times k$ table of frequencies, n_{ij}, representing the number of items that failed from cause i in time period j together with a k-dimensional vector $\mathbf{d}$ containing the numbers of items censored during the k time periods. If we make the change of parameters

$$\theta_{1j} = \sum_{l=1}^{c} p_{lj}, \qquad j = 1, 2, \ldots, k,$$

$$\theta_{ij} = \frac{\sum_{l=i}^{c} p_{lj}}{\sum_{l=i-1}^{j} p_{lj}}, \qquad i = 2, 3, \ldots, c, \quad j = 1, 2, \ldots, k,$$

and consider testing $\mathcal{H}_0$ vs. $\mathcal{H}_1$ where

$$\mathcal{H}_0 : \theta_{ij} = \theta_{i,j+1}, \ i = 2, 3, \ldots, c, \ j = 1, 2, \ldots, k-1,$$

$$\mathcal{H}_A : \theta_{ij} \leq \theta_{i,j+1}, \ i = 2, 3, \ldots, c, \ j = 1, 2, \ldots, k-1,$$

then the arguments in this paper carry through to this more general problem and the likelihood ratio statistic $T = -2 \ln \Lambda$ again has an asymptotic chi-bar-square distribution.

The least-favourable distribution is the same as the distribution of the sum of $c-1$ independent random variables each having the distribution given in (3.2). Since the chi-bar-square distribution is approximately normal for large k and the central limit theorem applies for large values of c, a normal approximation should serve well for many values of c and k encountered in practice. See Dykstra, Kochar and Robertson (1991) for more details.

The null hypothesis $\mathcal{H}_0$ is equivalent to independence of the time to failure and the cause of failure. However, for values of c larger than 2 the alternative hypothesis is not equivalent to likelihood-ratio dependence between these two variables. Under this alternative hypothesis the ratio $P[C > i+1,\ T = t_j]/P[C > i,\ T = t_j]$ is nondecreasing in j for $i = 1,2,\ldots,c-1$. Shaked (1977) calls such a relationship "dependence by total positivity of order (1,0)" and says that the pair (C,T) is $DPT(1,0)$. This type of dependence is implied by likelihood-ratio dependence but is stronger than positive regression dependence of C on T (see Shaked 1977). Grove (1980) proposed tests for testing independence of T and C against positive regression dependence. However, positive regression dependence is also implied by likelihood-ratio dependence.

In terms of the cause-specific hazard rates, the alternative hypothesis is equivalent to the assumption that the ratio $\sum_{i=l+1}^{c} g_i(t_j)/g_l(t_j)$ is nondecreasing in j for $l = 1,2,\ldots,c-1$ [$\sum_{i=l+1}^{c} g_i(t)$ is the cause-specific hazard rate of the combined risks $l+1, l+2,\ldots,c$].

4. AN EXAMPLE

Cox (1959) considers a data set originally analyzed by Medenhall and Hader (1958) involving failure times of radio receivers. These data are reproduced in Table 1. They are the failure times for 369 radio transmission receivers. These failures are classified as confirmed on arrival at the maintenance center (type I) or unconfirmed (type II). One of the problems posed by Cox (1959) is to test the null hypothesis that the hazard rates are proportional (that is, type of failure and time to failure are independent) against alternatives of type $\mathcal{H}_1$. Cox (1959) assumes that the two risks are independent, but in our analysis we do not make any such assumption.

Forty-four of the 369 receivers did not fail during the test period (630 hours), and these were the only censored items. The information contained in these 44 items gives no additional insight into the relationship between the cause-specific hazard rates for the two types of failure, so that we have chosen to ignore them [as was done by Cox (1959)].

As before, p_{ij} denotes the probability that we have a type-i malfunction during time period j, $j = 1,2,\ldots,13$. We consider the problem of testing H_0 against the restricted alternative $\mathcal{H}_1{}' : g_2(t)/\{g_1(t)+g_2(t)\}$ is nonincreasing in t [or equivalently, $g_1(t)/\{g_1(t)+g_2(t)\}$ is nondecreasing in t]. This alternative is reasonable, since it seems quite plausible that unconfirmed failures would become less prevalent over time. Since the theory developed above is designed for testing $\mathcal{H}_0$ against the alternative $\mathcal{H}_1 : g_2(t)/\{g_1(t)+g_2(t)\}$ is nondecreasing in t, we switch the roles of type I and type II failures, without any loss of generality. The unrestricted maximum-likelihood estimates of the θ_{2j}'s are given in column 2 of Table 2, and the estimates under $\mathcal{H}_1$ are given in Column 3 of that table. The value of the likelihood-ratio statistic, $-2\ln\Lambda$, for testing $\mathcal{H}_0$ against all alternatives (unrestricted) is 9.92, with a corresponding p-value $\text{pr}[\chi^2_{02} \geq 9.92] = 0.62$ based on 12 d.f. We mention that the p-value reported by Cox (1959) using a different chi-square test was 0.67.

For the restricted test, the value of the likelihood-ratio test statistic is 6.11. The p-value

TABLE 1: Frequency Distributions of Times of Failure of Radio Transmitters (Medenhall and Hader).

Time (h)	Observed frequency: Type I (confirmed)	Type II (unconfirmed)	Total
0–	26	15	41
50–	29	15	44
100–	28	22	50
150–	35	13	48
200–	17	11	28
250–	21	8	29
300–	11	7	18
350–	11	5	16
400–	12	3	15
450–	7	4	11
500–	6	1	7
550–	9	2	11
600–629	6	1	7
Not failed at 630 h	—	—	44
Total	218	107	369

TABLE 2: Computation of restricted m.l.e.'s

j	$\hat{\theta}_{2j}$	θ^*_{2j}
1	$26/41 = 0.634$	0.615
2	$29/44 = 0.659$	0.615
3	$28/50 = 0.560$	0.615
4	$35/48 = 0.729$	0.634
5	$17/28 = 0.607$	0.634
6	$21/29 = 0.724$	0.634
7	$11/18 = 0.611$	0.634
8	$11/16 = 0.6875$	0.6875
9	$12/15 = 0.800$	0.731
10	$27/11 = 0.636$	0.731
11	$6/\ 7 = 0.857$	0.833
12	$9/11 = 0.818$	0.833
13	$6/\ 7 = 0.857$	0.857

computed under the least-favourable distribution given in Theorem 3.1 is

$$\sum_{l=1}^{13} \binom{12}{l-1} \left(\frac{1}{2}\right)^{12} \mathrm{pr}(\chi^2_{l-1} \geq 6.11) = 0.4139.$$

Although neither test would lead to a rejection of $\mathcal{H}_0$, the p-value for the restricted test is substantially less than that of the unrestricted one.

Although standard procedure, tests based upon such least-favourable distributions are generally quite conservative and can be very misleading. The true asymptotic distribution depends upon the unknown parameters $\theta_{1j} = p_{1j} + p_{2j} = w_j$. A p-value computed from this distribution would be given by

$$\sum_{l=1}^{13} P(l, 13; \mathbf{w}) \operatorname{pr}(\chi^2_{l-1} \geq 6.11),$$

where $P(l, k; \mathbf{w})$ are the simple order level probabilities discussed in Chapters 2 and 3 of RWD. One could estimate this p-value by replacing $\mathbf{w}$ with its estimate $\hat{\mathbf{w}}$ where $\hat{w}_j = (n_{1j} + n_{2j})/325$. No closed-form expressions exist for $P(l, k; \mathbf{w})$ for k larger than 5, but several studies have confirmed that the *pattern* approximation discussed in Section 3.4 of RWD produces quite satisfactory approximations except in very unusual cases. Computer code for implementing this approximation is given in Pillers, Robertson and Wright (1984). The thirteen approximate level probabilities given by this approximation are 0.063, 0.209, 0.298, 0.242, 0.129, 0.047, 0.012, 0.002 for the first eight values and 0.000 for the last five values.

The p-value for this approximation is given by

$$\sum_{l=1}^{13} P(l, 13; \text{ pattern weights}) \operatorname{pr}(\chi^2_{l-1} \geq 6.11) = 0.087.$$

Although this would not lead to rejection at the 5% level, it indicates much stronger evidence in the data against independence of type of failure and time to failure when the alternative is restricted than when testing against an omnibus alternative.

Another commonly used approximation is the equal-weights approximation discussed in Section 3.1 of RWD. This yields a similar p-value of 0.076.

As a check on the pattern approximation, we estimated the $P(l, k; \hat{\mathbf{w}})$ by conducting a Monte Carlo simulation with 20,000 repetitions. That is, 20,000 weighted least-squares projections from normal vectors of length 13 with respective variances $1/\hat{w}_i$ onto the cone of nondecreasing vectors were computed, and the total number of level sets tallied. The empirical values for the $P(l, k; \hat{\mathbf{w}})$ were 0.062, 0.210, 0.303, 0.241, 0.130, 0.043, 0.010 and 0.002 for the first eight values and 0.000 for the last five. Note that these values are extraordinarily close to the values given by the pattern approximation.

The p-value associated with the simulated $P(l, k, \hat{\mathbf{w}})$ is given as

$$\sum_{l=1}^{13} \hat{P}(l, 13; \hat{\mathbf{w}}) \operatorname{pr}(\chi^2_{l-1} \geq 6.11) = 0.085,$$

very nearly the same as that given by the pattern approximation (0.087) and close to the equal-weights approximation (0.076), even though the estimated weights vary considerably.

This example would seem to support the robustness of the $P(l, k)$'s, and also indicates the inherent difficulties of constructing critical points based on least-favourable distributions.

REFERENCES

Aly, E., Kochar, S., and McKeague, I. (1994). Some tests for comparing cumulative incidence functions and cause specific hazard rates. *J. Amer. Statist. Assoc.*, 89, 994–999.

Agresti, A., Wackerly D., and Boyett, J. (1979). Exact conditional tests for cross classifications: Approximation of attained significance levels. *Psychometrika*, 44, 75–83.

Allen, W.R. (1963). A note on the conditional probabilities of failure rates when hazards are proportional. *Oper. Res.*, 11, 658–659.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics*, 11, 375–386.

Armitage, P. (1959). The comparison of survival curves. *J. Roy. Statist. Soc. Ser. A*, 122, 279–300.

Bartholomew, D.J. (1959). A test of homogeneity for ordered alternatives. *Biometrika*, 46, 36–48.

Cox, D.R. (1959). The analysis of exponentially distributed lifetimes with two types of failures. *J. Roy. Statist. Soc. Ser. B*, 21, 411–421.

Davis, T.P., and Lawrence, A.J. (1989). The likelihood for competing risks survival analysis. *Scand. J. Statist.*, 16, 23–30.

Dykstra, R., Kochar, S., and Robertson, T. (1991). Statistical inference for uniform stochastic ordering in several populations. *Ann. Statist.*, 19, 870–888.

Dykstra, R., Kochar, S., and Robertson, T. (1995a). Inference for likelihood ratio ordering in the two-sample problem. *J. Amer. Statist. Assoc.*, 90, 1034–1040.

Dykstra, R., Kochar, S., and Robertson, T. (1995b). Likelihood based inference for cause specific hazard rates under order restrictions. *J. Multivariate Anal.*, 54, 263–274.

Dykstra, R., Kochar, S., and Robertson, T. (1996). Testing whether one risk progresses faster than another is a competing risk problem. *Statist. Decisions*, 14, 209–222.

Gail, M. (1975). A review and critique of some models used in competing risk analysis. *Biometrics*, 31, 209–222.

Grove, D.M. (1980). A test of independence against a class of ordered alternatives in a $2 \times c$ contingency table. *J. Amer. Statist. Assoc.*, 75, 454–459.

Gumbel, E.J. (1960). Bivariate exponential distributions. *J. Amer. Statist. Assoc.*, 55, 698–707.

Kiefer, J., and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Z. Wahrsch. Verw. Geb.*, 34, 89–95.

Kochar, S.C., and Proschan, F. (1991). Independence of time and cause of failure in the dependent competing risks model. *Statist. Sinica*, 1, 295–299.

Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

Lee, M.T. (1989). Some cross product difference statistics and a test for trends in ordered contingency tables. *Statist. Probab. Lett.*, 7, 41–46.

Lehmann, E.L. (1966). Some concepts of dependence. *Ann. Math. Statist.*, 37, 1137–1153.

Medenhall, W., and Hader, R.J. (1958). Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. *Biometrika*, 45, 504–520.

Patefield, W.M. (1982). Exact tests for trends in ordered contingency tables. *Appl. Statist.*, 31, 32–43.

Pillers, C., Robertson, T., and Wright, F.T. (1984). A FORTRAN program for the level probabilities of order restricted inference. *J. Roy. Statist. Soc. Ser. C*, 33, 115–119.

Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Flourney, N., Farewell, V.T., and Breslow, N.E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34, 541–554.

Robertson, T., Wright, T., and Dykstra, R. (1988). *Order Restricted Statistical Inference*. Wiley, Chichester.

Rojo, J., and Samaniego, F.J. (1991). On nonparametric maximum likelihood estimation of a distribution uniformly stochastically smaller than a standard. *Statist. Probab. Lett.*, 11, 267–271.

Sethuraman, J. (1965). On a characterization of the three limiting types of the extreme. *Sankhyā*, 27, 357–364.

Shaked, M. (1977). A family of concepts of dependence for bivariate distributions. *J. Amer. Statist. Assoc.*, 72, 642–655.

Received 27 March 1996
Revised 22 May 1997
Accepted 4 June 1997

Department of Statistics and Actuarial Science
University of Iowa
Iowa City, Iowa
U.S.A. 52242-1466

email: rdykstra@stat.uiowa,edu
troberts@stat.uiowa.edu

Indian Statistical Institute
7, SJS Sansanual Marg
New Delhi - 110016
India

email: kochar@isid.ernet.in