Statistica Sinica **12**(2002), 725-750

ANALYSIS OF PARAMETRIC MODELS FOR COMPETING RISKS

Ross A. Maller and Xian Zhou

University of Western Australia and Hong Kong Polytechnic University

Abstract: "Competing risk" or "multiple cause" survival data arise in medical, criminological, financial, engineering, and many other contexts when death or failure of an individual or unit is classified into one of a variety of types or causes. Important issues in the analysis of such data range from basic properties, such as consistency of estimation of parameters, through more complex boundary hypothesis-testing problems, such as whether a specified list of causes is "exhaustive" − as opposed to the possibility that some individuals may be "immune" to all of these causes. We give a carefully formulated parametric mixture model for competing risk data which allows for censoring and immune individuals, and for which a large-sample analysis can be developed. Under some mild assumptions, we are able to show the existence, uniqueness (local to the true parameter values with probability approaching 1), consistency and asymptotic normality of the maximum likelihood estimators when the parameters are interior to the parameter space. A formulation using "cause-specific hazards" can be treated in the same way.

Consistent estimators also exist when the parameters are on the boundary of the parameter space, as is the case for example when testing for exhaustiveness of causes. The "deviance" statistic for testing this hypothesis is shown to have as its large-sample distribution a 50-50 mixture of a chi-square distribution with 1 degree of freedom, and a point mass at 0. Competing risks data with no censoring can be analyzed similarly.

The large-sample results we give allow many of the data-analytic questions for competing risks data to be formulated and answered in a satisfying way. The methods and approaches are illustrated on a set of criminological (re-arrest) data from Western Australia.

Key words and phrases: Competing risks, survival analysis, mixture models, censored data, causes of death, maximum likelihood estimation, likelihood ratio test.

1. Introduction

The analysis of competing risks data goes to the heart of modern preoccupations in survival analysis, touching as it does on many of the major areas of importance in the subject. The classification of death or failure by "type" or "cause" is a natural extension of (single-cause) survival analysis, once we consider alternative or "competing" causes of death or failure. The effort expended in fitting and interpreting competing risks models of some sort, evident from a glance at the current literature, recommends them as a worthwhile object of study. See, for example, Larson and Dinse (1985), David and Moeschberger (1978), Kalbfleisch and Prentice (1980), Gaynor et al. (1993), Escarela, Francis and Soothill (2000), Tai, Machin, White and Gebski (2001), and their references.

While there have been many investigations, often from the point of view of counting process theory (e.g., Andersen, Borgan, Gill and Keiding (1993), Fleming and Harrington (1991), Kalbfleisch and Prentice (1980)), of properties of various formulations of competing risks models, a comprehensive large-sample analysis of certain interesting aspects has not been given so far, especially for the cases where the "true" values of parameters may be on the boundary of the parameter space. This applies to a class of "mixture models", used for example in a paper by Larson and Dinse (1985) (see also Elandt-Johnson and Johnson (1980, p.288), which plays a prominent role in the analysis of competing risks and directly addresses the data-analytic questions of interest. To fill this need we provide a rigorous analysis of the parametric mixture models and derive useful large-sample properties of maximum likelihood estimators and test statistics, which cover both interior and boundary cases.

Under an i.i.d. censoring model, we show that the mixture model approach produces consistent estimates which are asymptotically normally distributed when the parameters are in the interior of the parameter space (we refer to this as the "interior" case). However, our main emphasis is on the large-sample distributions of the likelihood ratio statistics which can be used to test naturally arising hypotheses of interest. Some of the parameters of interest in the model, namely the mixing probabilities, are constrained to a J-dimensional simplex, and one of the interesting hypotheses requires those parameters to lie on the face of the simplex. Recently derived theory of Vu and Zhou (1997) can be applied in these "boundary-value" situations. Thus, in addition to "interior" results, we derive the large sample distribution of the likelihood ratio statistic used to test the boundary hypothesis of the exhaustiveness of failure causes − that is, that the causes under consideration account for all causes of death or, in other words, no individuals are "immune" to the currently considered causes of death. Note that here the term "death" or "failure" could have a much wider sense which includes, for example, re-arrest of a former prisoner (as in the criminology example in Section 5), divorce in a marriage, bankruptcy of a company, failure of a motor, etc.

The results can be easily applied, in particular, when the survival distributions are those commonly adopted in survival analysis, such as exponential, Weibull or Gamma distributions, as we show. They are illustrated on a criminological data set in which second arrests of a number of releases from West Australian prisons are classified into three categories: less, similar, and more "serious", according to some criminological criteria, than the first arrest. In a discussion section, some other applications and extensions of the results are mentioned.

The paper is organized as follows. Section 2 describes the models. Large sample properties of the maximum likelihood estimators and the boundary test for the exhaustiveness of failure causes are provided in Sections 3 and 4, respectively. The data analysis example is in Section 5. Section 6 gives some further discussions, while the proofs of the theory in Sections 3−4 are deferred to Section 7.

2. Parametric Mixture Models for Competing Risks

Following the approach in Larson and Dinse (1985), suppose that competing risks data consist of observations $t_1, \ldots, t_n$ on the lifetimes of n individuals and $c_1, \ldots, c_n$ on censor indicators, where

$$c_i = \begin{cases} 1 & \text{if individual } i \text{ is uncensored} \\ 0 & \text{if individual } i \text{ is censored,} \end{cases} \tag{2.1}$$

and, if $c_i = 1$, i.e., individual i dies, we also observe the cause of death, $j(i)$, which takes values in $\{1, \ldots, J\}$, say. Thus we can define, and observe, the indicators

$$c_{ij} = \begin{cases} 1 & \text{if individual } i \text{ dies of cause } j \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

for $1 \le i \le n$, $1 \le j \le J$.

Suppose that the index set $\{1, \ldots, J\}$ is a classification of those causes of death currently of interest to the researcher, and that we can define, for $1 \le j \le J$,

$$p_j = P\{\text{individual } i \text{ dies (will die) from cause } j\} \tag{2.3}$$

(and these do not depend on i; covariate models will be discussed in Section 6). Let

$$p = \sum_{j=1}^{J} p_j = P\{ \text{ an individual is susceptible to some risk } \} \tag{2.4}$$

with $0 < p \le 1$. If $p < 1$, some individuals are not susceptible (are immune) to any of the risks currently under consideration.

We assume an *i.i.d. censoring model* throughout, generalizing Farewell's (1977) formulation of a long-term survivor model. Associate with each individual i random variables (rvs) B_i, $(t^*_{i1}, \ldots, t^*_{iJ})$ and u_i. For each i, u_i is assumed independent of $(t_{i1}^*, \ldots, t_{iJ}^*)$ and of B_i, and $(B_i, t_{i1}^*, \ldots, t_{iJ}^*, u_i)_{1 \le i \le n}$ are i.i.d. The u_i are "censoring random variables" for the individuals (which may degenerate to constants, for example in the case of a fixed followup time for each), and have a common cumulative distribution function (c.d.f.) G, say. For most of the paper, G is assumed *proper*, i.e., to have total mass 1, except in the remark following Theorem 5, where it will be degenerate at infinity. The B_i, which are not observed, are assumed discrete with probabilities

$$P\{B_i = j\} = p_j, \ 1 \le j \le J, \quad P\{B_i = 0\} = 1 - \sum_{j=1}^{J} p_j = 1 - p. \tag{2.5}$$

They can be interpreted as indicators of the cause of death of an individual: $B_i = j$ if i will die from cause j, whereas $B_i = 0$ if i is immune to all risks under consideration. Let

$$F_j(t) = P\{t_{ij}^* \le t | B_i = j\} \tag{2.6}$$

denote the conditional c.d.f. of the t_{ij}^*, given that the death is of type j. Each $F_j(t)$ is proper. Suppose that t_i^* are random variables defined by

$$t_i^* = \begin{cases} t_{ij}^* & \text{on } \{B_i = j\}; \\ \infty & \text{on } \{B_i = 0\}. \end{cases}$$

Then t_i^* is independent of u_i. For each i, t_i^* has a c.d.f. $F(t)$ given by

$$P\{t_i^* \le t\} = \sum_{j=0}^{J} P\{t_i^* \le t | B_i = j\} P\{B_i = j\} = \sum_{j=1}^{J} p_j P\{t_{ij}^* \le t | B_i = j\} = \sum_{j=1}^{J} p_j F_j(t). \tag{2.7}$$

Note that if $p = \sum_1^J p_j < 1$ then F is improper. Finally, the observations t_i, c_{ij} and c_i are rvs satisfying $t_i = \min(t_i^*, u_i)$, $c_{ij} = 1_{\{t_{ij}^* \le u_i, B_i = j\}}$ and $c_i = 1_{\{t_i^* \le \, u_i\}} = \sum_j c_{ij}$ for $j = 1, \ldots, J$; $i = 1, \ldots, n$ (where 1_E, the indicator of an event E, takes value 1 if E occurs and 0 otherwise). The t_i^* represent the "true" lifetimes of the individuals, and the t_i represent the observed, possibly censored, lifetimes.

We suppose that the data at hand constitute, for individual i, observations on the random variables t_i and c_i, with c_{ij} also observed when $c_i = 1$. Thus (2.1)−(2.2) are satisfied if we follow the usual abuse of notation whereby observations on rvs are identified with the rvs themselves. At this stage we will make a further assumption that each $F_j(t)$ has a density function $f_j(t)$; this is not really essential but simplifies the analysis.

The setup described in (2.3)−(2.7) provides a probabilistic foundation for a mixture model approach to competing risks. An equivalent formulation can be given via *cause-specific distributions*. This approach is described in Prentice et al. (1978) and Kalbfleisch and Prentice (1980, p.167). Like the mixture model, it avoids the need to make assumptions about the joint distribution of the survival times under the different causes. This model starts with the *cause-specific hazards*

$$\lambda_j(t)dt = P\{t_i^* \in (t, t+dt),\ \text{individual } i \text{ dies from cause } j \mid t_i^* > t\},\ 1 \le j \le J, \tag{2.8}$$

where, as above, t_i^* is the random variable representing the observed lifetime of individual i. Adding over $1 \le j \le J$ in (2.8) produces the hazard function associated with the c.d.f. $F(t)$ of t_i^* as $\lambda(t) = \sum_j \lambda_j(t) = F'(t)/(1-F(t))$. From these the probability p_j that individual i ultimately dies of cause j and the survival distribution function $F_j(t)$ under cause j can be expressed respectively as

$$p_j = \int_0^\infty \lambda_j(t)(1-F(t))dt \quad \text{and} \quad F_j(t) = \frac{1}{p_j}\int_0^t \lambda_j(y)(1-F(y))dy.$$

These correspond to (2.3) and (2.6). The (improper) c.d.f. $p_jF_j(t)$ is referred to as the *cumulative incidence function*.

Conversely, given (2.3)$-$(2.6), we can define cause-specific hazards by

$$\lambda_j(t) = \frac{p_jf_j(t)}{1-\sum_{j=1}^J p_jF_j(t)} = \frac{p_jf_j(t)}{1-F(t)}, \quad j = 1,\ldots,J,$$

which have the same interpretation as in (2.8). Thus the two approaches give rise to equivalent formulations.

3. Maximum Likelihood Estimators

The likelihood of the sample (conditional on $\{B_1,\ldots,B_n\}$, and apart from a multiplicative constant) is

$$L_n = \prod_{i=1}^n (p_{j(i)}f_{j(i)}(t_i))^{c_i}\,(P(t_i^* > t))^{1-c_i} = \prod_{i=1}^n (p_{j(i)}f_{j(i)}(t_i))^{c_i}\Big(1-\sum_{j=1}^J p_jF_j(t_i)\Big)^{1-c_i}.$$

Letting $t_{ij} = t_i$ when $c_{ij} = 1$ (the t_{ij} need not be defined otherwise), we can write

$$L_n = L_n(\theta) = \prod_{i=1}^n \Big(\prod_{j=1}^J (p_jf_j(t_{ij}))^{c_{ij}}\Big)\Big(1-\sum_{j=1}^J p_jF_j(t_i)\Big)^{1-c_i}. \tag{3.1}$$

The p_j, $1 \le j \le J$, constitute parameters to be estimated from the data. Further parameters arise in the specification of the survival distributions for those susceptible to one of the risks $\{1,\ldots,J\}$. We assume a parameterization for F_j of the form

$$F_j(t) = F(t,\psi_j), \quad t \ge 0. \tag{3.2}$$

Here ψ_j is an s-vector of parameters varying over an open subset Ψ of $\mathbf{R}^s$. Collect all parameters into an $(s+1)J\times 1$ vector: $\theta=(\psi_1,\ldots,\psi_J,p_1,\ldots,p_J)\in\mathbf{R}^{(s+1)J}$. The "true" value of θ is taken to be $\theta_0=(\psi_{10},\ldots,\psi_{J0},p_{10},\ldots,p_{J0})$, with ψ_{j0} in Ψ, $1\le j\le J$, and $(p_{10},\ldots,p_{J0})$ constrained to lie in the simplex S_J defined by

$$S_J=\{(p_1,\ldots,p_J),\ p_j>0,\ 1\le j\le J,\ p_1+\cdots+p_J\le 1\}. \tag{3.3}$$

The parameter space, the range of θ, can then be written as $\Theta=\Psi^J\times S_J$.

We can write $f_j(t)=f(t,\psi_j)=\partial F(t,\psi_j)/\partial t$, $t\ge 0$. Assume that the partial derivatives $\partial^2 f(t,\psi)/\partial\psi\partial\psi^T$ and $\partial^2 F(t,\psi)/\partial\psi\partial\psi^T$ (as $s\times s$-matrices) exist and are finite at each $\psi\in\Psi$, and are continuous at each ψ_{j0} for each $t\ge 0$. Then, letting $\mathcal{L}_n(\theta)=\log(L_n(\theta))$ be the log-likelihood function, the derivatives

$$S_n(\theta)=\frac{\partial\mathcal{L}_n(\theta)}{\partial\theta}\quad\text{and}\quad\mathcal{F}_n(\theta)=-\frac{\partial^2\mathcal{L}_n(\theta)}{\partial\theta\partial\theta^T}, \tag{3.4}$$

the $(s+1)J\times 1$ score vector and $(s+1)J\times(s+1)J$ negative second derivative matrix of $\mathcal{L}_n(\theta)$, exist and are finite for θ in Θ. Further regularity assumptions needed are that, for any $t\ge 0$,

$$\frac{\partial F(t,\psi)}{\partial\psi}=\int_0^t\frac{\partial f(y,\psi)}{\partial\psi}dy\quad\text{and}\quad\frac{\partial^2 F(t,\psi)}{\partial\psi\partial\psi^T}=\int_0^t\frac{\partial^2 f(y,\psi)}{\partial\psi\partial\psi^T}dy. \tag{3.5}$$

Our aim is to maximize $L_n(\theta)$ for θ in Θ, or in some restricted subset of Θ, and to derive the large-sample distributions of the resulting estimators and test statistics. Our first result, given in Theorem 1, shows the kinds of conditions under which a maximum likelihood estimator (MLE) exists local to an interior "true" point, and provides the large sample properties for such an MLE. Let $\tau_G=\sup\{y>0:G(y)<1\}$ be the right extreme of G.

Theorem 1. *Assume, for each $j=1,\ldots,J$, that*

$$F(\tau_G,\psi_{j0})>0, \tag{3.6}$$

that, for each $z\in\mathbf{R}^s$ and $a\in\mathbf{R}$, $(z,a)\ne 0$,

$$P\left\{z^T\frac{\partial\log f(t^*_{ij},\psi_{j0})}{\partial\psi}+a=0\right\}=0, \tag{3.7}$$

and that, for some $\delta>0$,

$$E\left(\int_{[0,u]}\sup_{|\psi-\psi_{j0}|\le\delta}\left\|\frac{\partial^2\log f(y,\psi)}{\partial\psi\partial\psi^T}\right\|F(dy,\psi_{j0})\right)<\infty, \tag{3.8}$$

$$E\left(\sup_{|\psi-\psi_{j0}|\le\delta}\left(\left\|\frac{\partial^2 F(u,\psi)}{\partial\psi\partial\psi^T}\right\|+\left|\frac{\partial F(u,\psi)}{\partial\psi}\right|^2\right)\right)<\infty. \tag{3.9}$$

Then, with probability approaching 1 ($WPA1$) *as* $n \to \infty$, *there exists an interior maximum* $\hat{\theta}_n^{(1)}$ *of* $\mathcal{L}_n(\theta)$ in Θ *which is unique in a neighborhood of* θ_0 *and is a consistent estimator of* θ_0. *Moreover, as* $n \to \infty$, $\sqrt{n}(\hat{\theta}_n^{(1)} - \theta_0) \xrightarrow{D} N(0, D)$, *where* $D = E(\mathcal{F}_n(\theta_0))/n = E(\mathcal{F}_1(\theta_0))$ *is finite, nonsingular, and independent of* n, *and*

$$(\mathcal{F}_n(\hat{\theta}_n^{(1)}))^{T/2}(\hat{\theta}_n^{(1)} - \theta_0) \xrightarrow{D} N(0, I_{(s+1)J}). \tag{3.10}$$

In (3.8)−(3.10), $A^{\frac{1}{2}}$ denotes the left Cholesky square root of a positive definite matrix A (the lower triangular matrix such that $A^{\frac{1}{2}}(A^{\frac{1}{2}})^T = A$) and $A^{T/2}$ is its transpose, "$\xrightarrow{D}$" denotes convergence in distribution, $N(0, I_d)$ denotes a standard normal random vector in d dimensions, and "$\|\cdot\|$" is any matrix norm.

Theorem 1 tells us that $\hat{\theta}_n^{(1)}$ is $\sqrt{n}$-consistent for θ_0 and, when normed by an appropriate square root of the sample information matrix, the local MLE of an interior true parameter vector θ_0 is asymptotically standard normal. The "studentized" form of (3.10) is preferred for practical purposes, as it is directly usable, while D has to be estimated.

We next give a result for interior hypotheses. Let χ_ν^2 denote a chi-square random variable with ν degrees of freedom.

Theorem 2 (Interior hypothesis). *Let* $\mathbf{S}_r$ *be an* r*-dimensional subspace of* $\mathbf{R}^{(s+1)J}$, $0 \le r < (s+1)J$, *and let* θ^* *be any specified point in the interior of* Θ. *Consider the null hypothesis*

$$H_0^{(1)} : \theta_0 \in \Omega_r = (\mathbf{S}_r + \theta^*) \cap \Big(\Psi^J \times \Big\{(p_1, \ldots, p_J) \in S_J,\ \sum_{j=1}^{J} p_j < 1\Big\}\Big).$$

Then, under the conditions of Theorem 1,

(i) WPA1, *there exists a unique maximizer* $\hat{\theta}_n^{(2)}$ *over* Ω_r *in a neighborhood of* θ_0 *within* Ω_r, *which is consistent for* θ_0; *and*

(ii) *for testing* $H_0^{(1)}$ *versus an unrestricted interior alternative* (so $p_{10} + \cdots + p_{J0} < 1$), *the deviance statistic*

$$d_n^{(1)} = 2(\mathcal{L}_n(\hat{\theta}_n^{(1)}) - \mathcal{L}_n(\hat{\theta}_n^{(2)})) \xrightarrow{D} \chi^2_{(s+1)J-r} \tag{3.11}$$

where $\hat{\theta}_n^{(1)}$ *is as given in Theorem* 1.

Remarks. (i) Theorem 2 tells us that the deviance statistic ($-2\times$ log-likelihood ratio statistic) for testing "interior" hypotheses is asymptotically distributed as chi-square, as we would expect. The hypothesis $H_0^{(1)}$ in Theorem 2 allows testing not only a hypothesis where the null parameter has some or all components specified, but also relationships of practical interest between the components. As a simple example, consider the hypothesis $H_0^{(1)}: \theta_0 = \theta^*$, which specifies that θ_0 equal some given vector θ^*. Then Theorem 2 applies with $\Omega_r = \{\theta^*\}$ (*i.e.*, $\mathbf{S}_r = \{0\}$, $r = 0$) to give that $d_n^{(1)} = 2(\mathcal{L}_n(\hat{\theta}_n^{(1)}) - \mathcal{L}_n(\theta^*))$ is asymptotically distributed as chi-square with $(s+1)J$ degrees of freedom.

For another example, consider the null hypothesis $p_1 = p_2 = p_3$, with $p_1 + p_2 + p_3 < 1$, which specifies equal probabilities of death from each of three causes in a 3-cause study. Assume a Weibull mixture model, so that $s = 2$, $J = 3$, $(s+1)J = 9$ and $\theta = (\theta_1, \ldots, \theta_9)$. Then the above null hypothesis can be formulated as $\theta_7 = \theta_8 = \theta_9$, which is a version of $H_0^{(1)}$ with $\mathbf{S}_r = \mathbf{S}_7 = \{(x_1, \ldots, x_9) \in \mathbf{R}^9 : x_7 = x_8 = x_9\}$ and $\theta^* = 0$. The asymptotic distribution of the deviance for testing this hypothesis is, according to Theorem 2, $\chi^2_{(9-7)} = \chi^2_2$. Similarly, other hypotheses such as $\theta_1 = 2\theta_2 = 4\theta_3$ or $\theta_2 - \theta_1 = \theta_3 - \theta_2$, etc., can be formulated in the form required in Theorem 2, as well.

(ii) The conditions of Theorem 1 are quite mild. To begin with, (3.6) is minimal: it merely specifies that uncensored survival times of those dying from each cause can be observed with positive probability. Condition (3.7) guarantees D to be nonsingular, but is much easier to verify (for most models) than checking this fact directly. (In many treatments, this problem is simply assumed away.) For absolutely continuous $F(\cdot, \psi)$ (as we have assumed), (3.7) holds if the function $z^T \partial \log f(t, \psi_{j0})/\partial\psi + a$ has no zeroes or only isolated zeros in $t \in (0, \infty)$, which is often easy to verify. Conditions (3.8)$-$(3.9) are similar to those usually encountered in asymptotic analyses of MLE's in parametric (and other) models, but are considerably easier to check than most; in some treatments, conditions imposed include, for example, uniform bounds on third derivatives of the log-likelihood, or even assumptions on the (uniform) convergence of some stochastic quantities.

It is not difficult to verify that conditions (3.6)$-$(3.9) are satisfied for commonly used survival distributions such as exponential, Weibull and Gamma, with no further assumptions than that the censoring distribution G is not degenerate at 0, as we show in the next theorem.

Theorem 3. *Suppose the censoring distribution G is not degenerate at* 0*. If*
1. $F(t, \psi) = 1 - e^{-\lambda t}$ *is exponential, so* $s = 1$, $\psi = (\lambda)$, $\lambda > 0$, *or*
2. $F(t, \psi) = 1 - \exp(-(\lambda t)^\alpha)$ *is Weibull, so* $s = 2$, $\psi = (\lambda, \alpha)$, $\lambda > 0$, $\alpha > 0$, *or*
3. $f(t, \psi) = \lambda^r t^{r-1} e^{-\lambda t}/\Gamma(r)$ *is Gamma, so* $s = 2$, $\psi = (\lambda, r)$, $\lambda > 0$, $r > 0$,

then the conclusions of Theorems 1 *and* 2, *including* (3.10) *and* (3.11), *hold.*

Thus our results are directly applicable to a wide variety of practical situations.

4. Testing for Exhaustiveness of Causes

The above approach allows for the possibility that not all causes of death in the population are represented among $\{1, \ldots, J\}$. This is necessary in general because, with censored data, we may not know whether a censored individual will die from one of the causes under consideration. To investigate this, we may wish to test the hypothesis that the J specified causes do in fact include all those to which individuals are susceptible, i.e., that $\sum_1^J p_j = 1$. This hypothesis restricts the p_j's to a boundary of the parameter space - one face of the simplex S_J - so we do not expect a limiting chi-square distribution for the likelihood ratio statistic. But the required result can be derived from the Vu and Zhou (1997) results. We need the following notation: let $F_0(t) = \sum_{j=1}^J p_{j0}F_{j0}(t)$ and

$$H_\delta(t) = \max_{1\le j\le J} \sup_{|\psi-\psi_{j0}|\le\delta} \left|\frac{\partial H(t,\psi)}{\partial\psi}\right|, \ \delta > 0, \ t \ge 0. \tag{4.1}$$

Theorem 4 (Boundary hypothesis, no immunes). *Consider the hypothesis*

$$H_0^{(2)} : \theta_0 = (\psi_{10}, \ldots, \psi_{J0}, p_{10}, \ldots, p_{J0}) \in \Psi^J \times S_J, \quad p_{10} + \cdots + p_{J0} = 1. \tag{4.2}$$

Assume (3.6)−(3.8), *and that for some* $\delta > 0$ *and for each* $1 \le j \le J$,

$$E\left(e^{\delta H_\delta(u)} \sup_{|\psi-\psi_{j0}|\le\delta} \left\|\frac{\partial^2 F(u,\psi)}{\partial\psi\partial\psi^T}\right\|\right) < \infty, \tag{4.3}$$

$$E\left(\frac{e^{\delta H_\delta(u)}}{1-F_0(u)}\left(1 + \sup_{|\psi-\psi_{j0}|\le\delta}\left|\frac{\partial F(u,\psi)}{\partial\psi}\right|^2\right)\right) < \infty. \tag{4.4}$$

Then with probability approaching 1 *there exist, uniquely* (*local to* θ_0),
(i) *an unrestricted maximizer* $\hat\theta_n^{(1)}$ *of* $\mathcal{L}_n(\theta)$ *over the parameter space* Θ*; and*
(ii) *a restricted maximizer* $\hat\theta_n^{(3)}$ *of* $\mathcal{L}_n(\theta)$*, subject to the restriction* $H_0^{(2)}$*.*
Both $\hat\theta_n^{(1)}$ *and* $\hat\theta_n^{(3)}$ *are consistent for* θ_0 *as* $n \to \infty$*. Furthermore, the deviance statistic* $d_n^{(2)} = 2(\mathcal{L}_n(\hat\theta_n^{(1)}) - \mathcal{L}_n(\hat\theta_n^{(3)}))$ *has, asymptotically, the same distribution as* $N_1^2 I(N_1 \ge 0)$*, where* $N_1 \sim N(0,1)$*.*

Theorem 4 tells us that the limiting distribution of the likelihood ratio statistic for testing that no individuals are immune to all of the causes $\{1, \ldots, J\}$ is that of a 50-50 mixture of a chi-square random variable with 1 degree of freedom and a point mass at 0. Thus, for example, its 95th percentile is 2.71 (the 90th percentile of the χ_1^2 distribution).

We have not written down the limiting distribution of $\hat\theta_n^{(3)}$ itself, in Theorem 4. In the 1-cause case it can be seen in Theorem 3.2 of Zhou and Maller (1995). This distribution is probably not of much interest in itself; more important we think is the large-sample distribution of $d_n^{(2)}$. This allows us to accept $H_0^{(1)}$ (if the observed value of $d_n^{(2)}$ does not exceed 2.71), thus concluding that all (significant) causes of death in the study are included among $\{1, \ldots, J\}$, or to reject $H_0^{(1)}$, in which case we conclude that some individuals are "immune to" or "cured of" causes $\{1,\ldots,J\}$.

As a parallel to Theorem 3, we have

Theorem 5. *Suppose the censoring distribution G is not degenerate at* 0. *If $F(t,\psi)$ is exponential or Gamma and, for some $\eta > 0$,*

$$E(e^{(\lambda_{m0}+\eta)u}) < \infty, \tag{4.5}$$

where $\lambda_{m0} = \min(\lambda_{10},\ldots,\lambda_{J0})$*; or $F(t,\psi)$ is Weibull and, for some $\eta > 0$,*

$$E\left(\exp\left(\eta u^{\alpha_{M0}+\eta}\right)\right) < \infty, \tag{4.6}$$

where $\alpha_{M0} = \max(\alpha_{10},\ldots,\alpha_{J0})$, *then the conclusions of Theorem* 4 *hold.*

Remark. (iii) When there is no censoring and all failure times are observed, we must have $G(y) = 0$ for all $y \geq 0$ and a proper F, i.e., $p_1 + \cdots + p_J = 1$. This situation is "boundary" with respect to the original setup but has true value "interior" to the boundary subspace specified by $p_1 + \cdots + p_J = 1$ under both the null and alternative hypotheses. Assuming in addition that (3.7) and (3.8) (with $u = \infty$ a.s. in (3.8)) hold, then we can show that, with probability approaching 1, there exist, uniquely (local to the true value, $\tilde{\theta}_0$, say), a maximizer $\tilde{\theta}_n^{(4)}$ of $\mathcal{L}_n(\theta)$ subject to $p_1 + \cdots + p_J = 1$, which is consistent for $\tilde{\theta}_0$ and satisfies $(\mathcal{F}_n(\tilde{\theta}_n^{(4)}))^{T/2}(\tilde{\theta}_n^{(4)} - \tilde{\theta}_0) \xrightarrow{D} N(0, I_{(s+1)J-1})$. Similarly, we can obtain a consistent maximizer $\tilde{\theta}_n^{(5)}$ restricted to an r-dimensional subspace of $\mathbf{R}^{(J+1)s}$ and subject to $p_1 + \cdots + p_J = 1$, such that $d_n^{(3)} = 2(\mathcal{L}_n(\tilde{\theta}_n^{(4)}) - \mathcal{L}_n(\tilde{\theta}_n^{(5)}) \xrightarrow{D} \chi^2_{(s+1)J-r}$. Moreover, similar to Theorem 3, we can obtain asymptotic normality for MLE's when there is no censoring for exponential, Weibull, or Gamma models, with no further assumptions required. We omit the proofs of these, which are similar to those of Theorems 1−3.

5. Example: Time to First Re-Arrest

Figure 5.1 below shows "cumulative incidence curves" and fitted Weibull distributions for 3,636 male Australian aborigines who had incurred at least one arrest in the West Australian legal system over the period 1984-1996. The data is extracted from a much larger database held by the University of Western Australia Crime Research Center, which contains the entire arrest records of the population over that period. As the diagram shows, followup of as much as 12 years is thus available for some individuals. The "survival time" of interest is the time between the first and second arrests, if a second arrest occurred within the limit of followup, otherwise the re-arrest time is censored.

The second arrest, if it occurred, was classified into one of three types: less, equal, or more serious than the first, where "seriousness" is based on criminological criteria which are not especially relevant here, but give a classification of failure which we can use as an interesting (and important) example of a competing risks analysis.

The empirical cumulative incidence function (c.i.f.) is an estimator of the theoretical c.i.f., and can be used as a convenient form of data display. For its definition see Kalbfleisch and Prentice (1980, p.169), and for a good discussion of its properties and an example of its use in a medical context see Gaynor et al. (1993). The empirical c.i.f. for a given type of failure is constant except for jumps at failure times of that type.

In Figure 5.1, empirical cumulative incidence functions of the three types are shown with dots at their jump points for three rearrest types:

Type 1: Second offense less serious than first;
Type 2: Second offense equally as serious as first;
Type 3: Second offense more serious than first.

At around 2 years, the order is Type 2 > Type 1 > Type 3. They each level at about 0.25 at their right-hand ends. The three incidence functions add to the Kaplan-Meier (1958) estimator (KME) of the overall re-arrest times, without regard to type and thus represent a disjoint decomposition of the KME into the three competing types.

The three continuous curves in Figure 5.1 are Weibull distributions fitted by the likelihood method of Section 3 (they are not fitted to the incidence functions directly). As can be seen from the figure, the Weibull mixture model provides an extremely good description of each of the three types of failure. The estimated Weibull parameters with 95% confidence intervals in brackets (failure time is measured in years) are:

Type 1: $\hat{p}_n = 0.263$ (0.244, 0.282), $\hat{\alpha}_n = 0.781$ (0.728, 0.838), $\hat{\lambda}_n = 0.531$ (0.456, 0.618);
Type 2: $\hat{p}_n = 0.244$ (0.230, 0.261), $\hat{\alpha}_n = 0.765$ (0.719, 0.813), $\hat{\lambda}_n = 0.907$ (0.804, 1.023);
Type 3: $\hat{p}_n = 0.251$ (0.234, 0.268), $\hat{\alpha}_n = 0.792$ (0.740, 0.846), $\hat{\lambda}_n = 0.655$ (0.574, 0.747).

The fitted Weibull sub-distributions are easy to distinguish in Figure 5.1; they each level at about 0.25, as indicated by their $\hat{p}_n$ values, but their $\hat{\lambda}_n$ values appear different, and a test of $H_0^{(3)} : \lambda_{10} = \lambda_{20} = \lambda_{30}$ gives a deviance value of 29.54. According to Theorem 2 we can take this as approximately χ_2^2, and thus it is highly significant. So the three types have significantly different rates of failure. This is apparent from the figure, where the incidence function for Type

2 rises significantly faster than for the other two types in the first two years or so after the initial arrest. The α-parameters of the Weibull are significantly less than 1, so the distribution of the data is exponential-like, with a rapid failure (re-arrest) rate at small times. These kinds of conclusions have useful criminological applications.

The test of $H_0^{(2)}: p_{10} + p_{20} + p_{30} = 1$ gives a deviance value of 89.7 which is very significant by comparison with its critical value of 2.71, according to Theorem 5. We deduce that there is a large "immune" component in the population, about 25% of the population, who will not be expected to fail again (be rearrested) and clearly the amount of followup available in the data set contributes to our confidence in this conclusion.

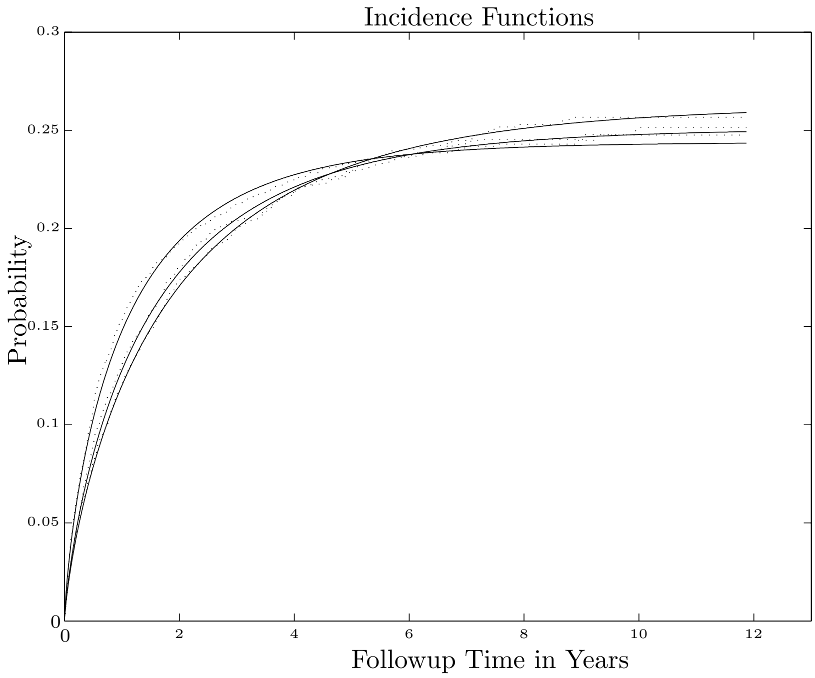

Figure 5.1. Empirical cumulative incidence functions and fitted Weibull curves for a set of criminological data.

6. Discussion

The theoretical results provide the basic large-sample foundations for data analysis using the mixture model in the single-sample case. The main findings are that the basic properties of consistency and asymptotic normality hold in "interior" cases, while the large-sample 50-50 chi-square distribution is appropriate for the boundary hypothesis test that none of the population is immune to the J causes of death currently under consideration. With the intuition gained from these results, we can extrapolate to more general setups, and we now explore some of these. A useful backdrop to this discussion is Chapter 7 of Kalbfleisch and Prentice (1980).

Covariates and More General Survival Distributions.

We restricted ourselves to identically distributed observations, which enables us to give the clear formulations of Theorems 1−5, and thereby to build up intuition as to which elements are important in the analysis of competing risks data. The role of the censoring distribution becomes apparent from our analysis − apart from the integrability and non-degeneracy assumptions, we need only the minimal assumption (3.6) that uncensored survival times of those dying from each cause can be observed, with positive probability. These are ways of representing that the censoring not be too "heavy".

Although we assumed in (3.2) that the survival distribution for each cause is of the same type, there is no difficulty in principle in allowing different types for different causes; e.g., we might have a Weibull for Cause 1 and a Gamma for Cause 2. The methods can be extended to cover such cases.

The assumption of identical distributions can be relaxed by allowing the parameters p_j and ψ_j in Section 2 to depend on covariate information specific to individual i by way of, e.g., logistic-linear and log-linear functions of covariate vectors for p_j and λ_j, in the exponential model. The manner of doing this is detailed for example in Ghitany, Maller and Zhou (1994) and Vu, Maller and Zhou (1998), for single-cause models, and much the same approach will work here. Results can be obtained under conditions on the covariates which are not too far from requiring that they be uniformly asymptotically negligible in the sense spelled out in Maller (1993), for example; and under conditions for the censoring distributions which reveal their role in the asymptotic analysis by showing, in some way, the necessity for "sufficient followup" to exist in the sample. Analyses like these add greatly to understanding the way censoring and covariate information affect the inferences we can draw from the data, as discussed in detail, for example in Ghitany, Maller and Zhou (1994) and Vu, Maller and Zhou (1998) and Ghitany and Maller (1992).

Goodness of Fit and Sufficient Followup

Goodness of fit of parametric distributions to competing risks data can be assessed informally using the cumulative incidence function as in Figure 5.1. To proceed more formally we could use individual-cause PP plots, as exemplified in Maller and Zhou (1996, Section 5.4, p.115) for the one-cause case, and simulate percentage points of some measure of the linearity of the plots, as is done there.

Similarly, "sufficient followup" in censored competing risks data could be assessed by generalising the methods of Maller and Zhou (1996, p.37) for single-cause survival data. These are important areas for future research.

Additional Causes, and Elimination of a Cause

The boundary hypothesis test proposed in Theorem 4 is essentially a test for whether all causes of mortality affecting the population are listed among $\{1, \ldots, J\}$; or are there other potential causes? As a vivid illustration of the possibilities here see the analysis by Larson and Dinse (1985) of the Stanford heart transplant data, where they conjecture, in explanation for the fact that their model fits poorly in respect to "other" causes of death, that at least one other major cause should be listed. Larson and Dinse (1985) constrain their $p_j, 1 \leq j \leq J$, to add to 1, so they are operating in the restricted parameter space of Theorem 4. It would be interesting to extend their analysis to the situation of Theorem 1 to see if the fitting can be improved significantly by allowing the sum of the p_j to be less than 1.

The above discussion relates to the possibility of *additional* causes. The obverse question is: can we test for the *elimination* of a cause? We may suppose for example that a researcher believes a "cure" from Cause 1 has been effected, and indeed in his/her data observes very few deaths from this cause. This cause still exists in the population, since we do observe some deaths from it (otherwise there is nothing to test), but is it "significant" in some sense, relative to the other causes? This question is of great practical interest since surely a "cure" is the ultimate aim or at least hope of any treatment. We have not addressed this question here but it can be approached as a test for a significant increase in the log-likelihood when Cause 1 is dropped from the list. This creates a degenerate situation in that $p_1 = 0$ implies complete elimination of ψ_1 as well. Lemdani and Pons (1997) have considered this issue in a related context (see below).

Early researchers debated ways of estimating the effects of remaining causes if it were possible to eliminate one or more causes. Daniel Bernoulli, in 1760, considered the effect on population mortality of the eradication of smallpox (just recently become a reality!), for example. On this issue we concur with Prentice et al. (1978, p.546), who stress that "the interpretation of such effects is ... restricted to actual study conditions and there is no implication that the same regression estimates would prevail ... (if) certain causes of failure have been eliminated". In line with this philosophy, and with the data-driven approach we adopt, our analysis does not address this issue as such.

The Lemdani and Pons Model

Lemdani and Pons (1997) give an analysis of a "multiple cause" survival model which differs from our formulation in that the different causes of death or failure are not identified. Their model is in fact an extension of the "long-term survivor" model discussed in Maller and Zhou (1996). The different causes or "susceptibility factors" are manifest in the model as a vector of probabilities to be estimated. A major issue is then the identifiability of these mixing probabilities, from data in which the different causes of death are not identified. Lemdani and Pons' analysis is by martingale methods and requires very stringent assumptions, apart from identifiability assumptions. They obtain consistency and asymptotic normality of the parameters in a parametric setup, and an analogue of Theorem 4 for the boundary value case when individuals are "totally susceptible" to death, among other results. They also address the problem of the elimination of a cause, as discussed above, providing an ingenious approach to it in their context.

Parametric or Nonparametric?

We have restricted ourselves to parametric models for competing risks but this is not to deny the important role played by nonparametric methods, not only for displaying the data, as in Section 6, but also as a vehicle for the estimation and testing of effects. We expect that a development of nonparametric techniques along the lines of that in Maller and Zhou (1996) for single-cause models would throw a great deal of light on the important issue of sufficient followup, for example. Including covariates in a nonparametric approach suggests a version of the Cox proportional hazards model (Kalbfleisch and Prentice (1980, p.183), Cox and Oakes (1984, p.143)). Useful approaches to nonparametrics in the competing risks context such as that of Lagakos, Sommer and Zelen (1978) and Chapter IV of Andersen et al. (1993) provide a basis for this development. For an interesting discussion of some of the data-display issues, see Pepe and Mori (1993). Lin (1997) gives an analysis of such methods using counting process methods, with illuminating data displays as well.

7. Proofs of Theorems 1−5

For the most part our proofs will be by appeal to the results in Vu and Zhou (1997) (see also Self and Liang (1987) and Lemdani and Pons (1997)), where a very general approach to the boundary (and interior) hypothesis testing problem for estimating functions is set out under some natural and fairly mild conditions. Apart from second order differentiability of the log likelihood, the parameter spaces specified by both the alternate and null hypotheses are required to be (locally) cone-like near the true parameter point θ_0, in the sense that there is a cone which coincides with the parameter space under consideration on a closed neighborhood of θ_0. See conditions (A1), (A2), (A2′), (A3) of Vu and Zhou (1997). These conditions are trivially satisfied in our setup once the parameter spaces have been specified appropriately. The third type of condition, (B1)−(B5) of Vu and Zhou (1997), concerns the probabilistic properties of the first and second derivatives of the log likelihood. Since our models have true likelihoods (Vu and Zhou allow a more general estimating equation setup) we can take $D_n = G_n$ and $V = I$ in their notation, thus simplifying their (B1) and making their (B4) redundant. Furthermore, under our i.i.d. censoring model, the (expected) information matrix, $D_n = E(\mathcal{F}_n(\theta_0))$, equals nD, where D is the (expected) information contributed by one individual. Thus we can recast Vu and Zhou's (B1)−(B5) as requiring, in our context:

(B1) $E(S_n(\theta_0)) = 0$ and $E(S_n(\theta_0)(S_n(\theta_0))^T) = E(\mathcal{F}_n(\theta_0))$, a finite matrix; (7.1)

(B2) $\lambda_{\min}(D) > 0$, i.e., D is nonsingular; (7.2)

(B3) $$\sup_{\theta\in N_n(A)} \left\| \frac{1}{n}\mathcal{F}_n(\theta) - D \right\| \xrightarrow{P} 0 \text{ as } n\to\infty, \text{ for each } A > 0 \qquad (7.3)$$

(here $N_n(A) = \left\{\theta \in \Theta : n(\theta-\theta_0)^T D(\theta-\theta_0) \le A^2\right\}$, for $n \ge 1$ and $A > 0$); and

(B5) (B2) holds and $$D^{-1/2}\frac{S_n(\theta_0)}{\sqrt{n}} \xrightarrow{D} N(0, I). \qquad (7.4)$$

(Here $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a symmetric matrix.) The first equality in (B1) gives the "unbiasedness" of the estimating function derived from the likelihood, while the second equality allows us to take $D_n = G_n$ in Vu and Zhou's original (B1). (B2) is a necessary requirement for non-degeneracy of the model, as is apparent from (B5). (B3) is the hardest condition to check. Our i.i.d. assumptions and the Weak Law of Large Numbers immediately give $\mathcal{F}_n(\theta_0)/n \xrightarrow{P} D$ as $n \to \infty$, but (7.3) requires some uniformity of convergence in a neighborhood of θ_0.

We stress that (B1)−(B5) (and likewise, (A2) and (A3)) must be checked anew for each null hypothesis under test, and this may necessitate different considerations and result in quite different conditions (such as the stronger integrability conditions required in Theorems 4 and 5).

For the proofs we need the derivatives in $S_n(\theta)$ and $\mathcal{F}_n(\theta)$. Let $\mathcal{L}_{(i)}$ be the contribution to the likelihood by individual i, so from (3.1) and (3.2) we can write

$$\mathcal{L}_{(i)}(\theta) = \sum_{j=1}^{J} (c_{ij}(\log p_j + \log f(t_{ij}, \psi_j))) + (1-c_i)\log(1-F(t_i))$$

where $F(t)$ is defined in (2.7). Thus, for $1 \le j \le J$,

$$\frac{\partial \mathcal{L}_{(i)}(\theta)}{\partial \psi_j} = c_{ij}\frac{\partial \log f(t_{ij}, \psi_j)}{\partial \psi} - \frac{(1-c_i)p_j}{1-F(t_i)}\frac{\partial F(t_i, \psi_j)}{\partial \psi}, \qquad (7.5)$$

$$\frac{\partial \mathcal{L}_{(i)}(\theta)}{\partial p_j} = \frac{c_{ij}}{p_j} - \frac{(1-c_i)F(t_i,\psi_j)}{1-F(t_i)}. \tag{7.6}$$

These constitute the $(s+1)J$ components of $S_n(\theta)$ defined in (3.4). Second derivatives are as follows:

$$-\frac{\partial^2 \mathcal{L}_{(i)}(\theta)}{\partial \psi_j \partial \psi_k} = -c_{ij}\frac{\partial^2 \log f(t_{ij},\psi_j)}{\partial \psi \partial \psi^T} 1_{\{j=k\}} + \frac{(1-c_i)p_j}{1-F(t_i)}\frac{\partial^2 F(t_i,\psi_j)}{\partial \psi \partial \psi^T} 1_{\{j=k\}}$$
$$+\frac{(1-c_i)p_j p_k}{(1-F(t_i))^2}\frac{\partial F(t_i,\psi_j)}{\partial \psi}\frac{\partial F(t_i,\psi_k)}{\partial \psi^T}, \tag{7.7}$$

$$-\frac{\partial^2 \mathcal{L}_{(i)}(\theta)}{\partial \psi_j \partial p_k} = \left\{\frac{(1-c_i)}{1-F(t_i)}\frac{\partial F(t_i,\psi_j)}{\partial \psi}\right\} 1_{\{j=k\}} + \frac{(1-c_i)p_j F(t_i,\psi_k)}{(1-F(t_i))^2}\frac{\partial F(t_i,\psi_j)}{\partial \psi}, \tag{7.8}$$

$$-\frac{\partial^2 \mathcal{L}_{(i)}(\theta)}{\partial p_j p_k} = \frac{c_{ij}}{p_j^2} 1_{\{j=k\}} + \frac{(1-c_i)F(t_i,\psi_j)F(t_i,\psi_k)}{(1-F(t_i))^2}, \tag{7.9}$$

for $1 \leq j,k \leq J$. When added over $1 \leq i \leq n$, these form the components of $\mathcal{F}_n(\theta)$, defined in (3.4).

Before proceeding, we give three lemmas, the first of which establishes (B1) for the assumed model, and the next two of which contain sufficient conditions for (B2) and (B3).

Lemma 7.1. *We have*

$$E\left[\frac{\partial \mathcal{L}_{(1)}(\theta)}{\partial \theta}\right]_{\theta=\theta_0} = 0 \quad \textit{and} \quad E\left[-\frac{\partial^2 \mathcal{L}_{(1)}(\theta)}{\partial \theta \partial \theta^T}\right]_{\theta=\theta_0} = E\left[\frac{\partial \mathcal{L}_{(1)}(\theta)}{\partial \theta}\frac{\partial \mathcal{L}_{(1)}(\theta)}{\partial \theta^T}\right]_{\theta=\theta_0},$$

and the latter is a finite matrix.

Proof. The following formulae are useful. Let $Q(\cdot)$ be a measurable function on $\mathbf{R}$ and write $F_{j0}(y) = F(y,\psi_{j0})$ for brevity. Then for any i, j, $1 \leq i \leq n$, $1 \leq j \leq J$,

$$E(c_{ij}Q(t_i)) = E(c_{ij}Q(t_{ij}^*)) = E\Big(Q(t_{ij}^*)1_{\{t_{ij}^* \leq u_i, B_i = j\}}\Big)$$
$$= p_{j0}E(E(Q(t_{ij}^*)1_{\{t_{ij}^* \leq u_i\}}|u_i, B_i = j)) = p_{j0}E\Big(\int_0^u Q(y)dF_{j0}(y)\Big), \tag{7.10}$$

$$E((1-c_i)Q(t_i)) = E\Big(E(Q(u_i)1_{\{t_i^* > u_i\}}|u_i, B_i)\Big) = E\left(Q(u)(1-F_0(u))\right), \tag{7.11}$$

where u is a rv with c.d.f. G. The expectations on the right sides of (7.10)−(7.11) are with respect to u, while those on the left sides are with respect to any distributional setup (and specification of parameters) under which the i.i.d. censoring model holds. Note that the independence of the t_{ij}^* from u_i, and of t_i^* from u_i, is crucial in these calculations.

The lemma follows straightforwardly by applying these formulae together with the regularity assumptions in (3.5) to (7.5)−(7.9).

Lemma 7.2. *Conditions* (3.6) *and* (3.7) *imply* (7.2).

Proof. For brevity let $Y_i = \partial\mathcal{L}_{(i)}(\theta_0)/\partial\theta$. Suppose by way of contradiction that D is singular, so $w^T D w = 0$ for some $w = (w_1, \ldots, w_J, w_{Js+1}, \ldots, w_{(J+1)s}) \in \mathbf{R}^{(J+1)s}$, $w \neq 0$, where $w_j \in \mathbf{R}^s$ for $j = 1, \ldots, J$ and $w_j \in \mathbf{R}$ for $j = Js + 1, \ldots, (J+1)s$. Then $w^T Y_1 Y_1^T w$ is a nonnegative r.v. with $E(w^T Y_1 Y_1^T w) = w^T D w = 0$ by Lemma 7.1, which implies $w^T Y_1 = 0$ a.s. As $w \neq 0$, there exists $k \in \{1, \ldots, J\}$ such that $(w_k, w_{Js+k}) \neq 0$. For this k we define an event $A_k = \{c_{1k} = 1\} = \{t^*_{1k} \le u_1, B_1 = k\}$ and let

$$Z_k = w_k^T \frac{\partial \log f(t^*_{1k}, \psi_{k0})}{\partial\psi} + \frac{w_{Js+k}}{p_{k0}}.$$

On A_k, $c_{1j} = 0$ for $j \neq k$, $c_1 = 1$, and $t_{1k} = t^*_{1k}$, so that $w^T Y_1 = Z_k$ by (7.5) and (7.6). It follows that $P(Z_k \neq 0, A_k) = P(w^T Y_1 \neq 0, A_k) \le P(w^T Y_1 \neq 0) = 0$ and so, by (3.6),

$$\begin{aligned} P(Z_k = 0) &\ge P(Z_k = 0, A_k) = P(A_k) = P(t^*_{1k} \le u_1, B_1 = k) \\ &= \int_{[0,\infty)} P(t^*_{1k} \le y, B_1 = k) dG(y) = p_{k0} \int_{[0,\infty)} F(y, \psi_{k0}) dG(y) > 0. \end{aligned}$$

This contradicts (3.7), so D is non-singular.

Lemma 7.3. *Suppose* (7.2) *holds. Then* (7.3) *holds if, for some* $\delta > 0$,

$$\text{(B3}^*\text{)} \qquad E\Big(\sup_{|\theta-\theta_0|\le\delta} \Big\| \frac{\partial^2 \mathcal{L}_{(1)}(\theta)}{\partial\theta\partial\theta^T} \Big\| \Big) < \infty. \tag{7.12}$$

Proof. Assume (7.12) and write $\mathcal{F}_n(\theta) = -\partial^2\mathcal{L}_n(\theta)/\partial\theta\partial\theta^T = \sum_i X_i(\theta)$, say, where the $X_i(\theta)$ are $(J+1)s \times (J+1)s$ symmetric random matrices constructed from the right hand sides of (7.7)−(7.9). Take $\delta > 0$ and let $\mathcal{N}(\delta) = \{|\theta - \theta_0| \le \delta\}$ be a closed δ-neighborhood of θ_0 in $\mathbf{R}^{(s+1)J}$. Note that

$$\begin{aligned} \frac{1}{n} \sup_{\theta\in\mathcal{N}(\delta)} \Big\| \mathcal{F}_n(\theta) - nD \Big\| &= \frac{1}{n} \sup_{\theta\in\mathcal{N}(\delta)} \Big\| \sum_{i=1}^n (X_i(\theta) - EX_i(\theta_0)) \Big\| \\ &\le \Big\| \frac{1}{n} \sum_{i=1}^n (X_i(\theta_0) - EX_i(\theta_0)) \Big\| + \frac{1}{n} \sum_{i=1}^n \sup_{\theta\in\mathcal{N}(\delta)} \| X_i(\theta) - X_i(\theta_0) \| . \end{aligned} \tag{7.13}$$

The first term on the right hand side of (7.13) is the average of n i.i.d. random matrices with expected value 0, hence tends to 0 as $n \to \infty$ by the Weak Law of

Large Numbers. The second term on the right hand side of (7.13) is the average of n non-negative i.i.d. rvs, $Y_i(\delta)$, say, which have finite expectations by (7.12) and are monotone in δ. By the Weak Law of Large Numbers this term converges in probability to $EY_1(\delta)$. Take a sequence $\delta_k \downarrow 0$, and fix $\varepsilon > 0$. As $X_1(\theta)$ is a.s. continuous at θ_0,

$$\begin{aligned}
&\limsup_{k\to\infty} P(Y_1(\delta_k) > \varepsilon)\\
&= \limsup_{k\to\infty} P\Big(\sup_{\theta\in\mathcal{N}(\delta_k)} \|X_1(\theta) - X_1(\theta_0)\| > \varepsilon\Big)\\
&\le \limsup_{k\to\infty} P(\|X_1(\theta)-X_1(\theta_0)\|>\varepsilon \text{ for some } \theta \in \mathcal{N}(\delta_k) \text{ with rational coordinates})\\
&\le P(\|X_1(\theta) - X_1(\theta_0)\| > \varepsilon \text{ infinitely often for } \theta \in \mathcal{N}(\delta_k) \text{ as } \delta_k \downarrow 0) = 0.
\end{aligned}$$

Thus $Y_1(\delta) \xrightarrow{P} 0$ as $\delta \downarrow 0$ and hence, by monotone convergence, $EY_1(\delta) \to 0$ as $\delta \downarrow 0$.

Now, given $\varepsilon > 0$, we can choose $\delta = \delta(\varepsilon) > 0$ so small that $EY_1(\delta) \le \varepsilon$. Recall the definition of $N_n(A)$ following (7.3), let $A > 0$ and take $\theta \in N_n(A)$. In view of (7.2) and $A^2 \ge n(\theta - \theta_0)^T D(\theta - \theta_0) \ge n|\theta - \theta_0|^2\lambda_{\min}(D)$, we see that $\theta \in \mathcal{N}(\delta)$ if we choose $n \ge A^2/(\delta^2\lambda_{\min}(D))$.
Then for n this large,

$$\begin{aligned}
&P\Big(\frac{1}{n}\sup_{\theta\in N_n(A)} \|\mathcal{F}_n(\theta) - nD\| > 3\varepsilon\Big) \le P\Big(\frac{1}{n}\sup_{\theta\in\mathcal{N}(\delta)} \|\mathcal{F}_n(\theta) - nD\| > 3\varepsilon\Big)\\
&\le P\Big(\frac{1}{n}\Big\|\sum_{i=1}^n X_i(\theta_0) - EX_i(\theta_0)\Big\| > \varepsilon\Big) + P\Big(\frac{1}{n}\sum_{i=1}^n Y_i(\delta) > EY_1(\delta) + \varepsilon\Big).
\end{aligned}$$

The last expression tends to 0 as $n \to \infty$, establishing (7.3), i.e., (B3).

Proof of Theorem 1. The Vu and Zhou (1997) formulation specifies subsets τ and Ω of Θ over which maximization takes place. For Theorem 1 it suffices to consider $\tau = \Theta$. Now τ is required to be locally cone-like in the sense that there is a closed neighborhood $\mathcal{N}(\delta)$ of θ_0 in $\mathbf{R}^{(s+1)J}$ with δ small enough, and a *containing cone* C_τ with vertex at θ_0, such that $C_\tau \cap \mathcal{N}(\delta) = \tau \cap \mathcal{N}(\delta)$. This holds with $C_\tau = \mathbf{R}^{(s+1)J}$ here, as clearly $\mathcal{N}(\delta) \subseteq \tau$ for sufficiently small δ, since θ_0 is an interior point of Θ. This establishes (A2) of Vu and Zhou (1997), for τ. (A3) of Vu and Zhou (1997) requires C_τ to be rescaled and centered at 0, giving a cone $\tilde{C}_{\tau_n}$ which must "asymptotically coincide" with a cone $\tilde{C}_\tau$ in the sense of (A3), p.903 of Vu and Zhou (1997). Here $\tilde{C}_\tau = \mathbf{R}^{(s+1)J}$ meets the requirement as C_τ is trivially invariant under rescaling and centering.

Now suppose we have verified (B1)−(B5). Then Theorem 2.1 of Vu and Zhou (1997) gives a local MLE $\hat{\theta}_n^{(1)}$ (maximizer of $\mathcal{L}_n(\theta)$ over τ), which is locally uniquely determined interior to τ WPA1, and is consistent for θ_0 as $n \to \infty$.

Asymptotic distributions of the maximum estimators are not spelled out in Vu and Zhou (1997) since they can be very complicated in boundary cases and depend heavily on the particular situation. However, since $\hat{\theta}_n^{(1)}$ is an interior maximum, i.e., $S_n(\hat{\theta}_n^{(1)}) = 0$, it is a straightforward result of (B5) that $\sqrt{n}(\hat{\theta}_n^{(1)} - \theta_0)$ converges in distribution to $N(0, D)$, and then (3.10) follows from (B3) and an application of the general studentization theorem of Vu, Maller and Klass (1996).

It remains to verify (B1)−(B5). (B1) was proved in Lemma 7.1, and (B2) follows from Lemma 7.2 and (3.6) and (3.7). For (B3) we use Lemma 7.3 and verify that (B3*) holds for the present setup under (3.8)−(3.9). (7.10) and (7.11) are used to evaluate the integrals. Thus to deal with the first term on the right hand side of (7.7), we need that

$$E\left(c_{1j} \sup_{|\psi-\psi_{j0}|\le\delta} \left\| \frac{\partial^2 \log f(t_{1j},\psi)}{\partial\psi\partial\psi^T} \right\| \right) < \infty \tag{7.14}$$

for some $\delta > 0$ and for each $1 \le j \le J$. But by (7.10) this follows immediately from (3.8).

The second term on the right hand side of (7.7) requires checking

$$E\left((1-c_1) \sup_{|\psi-\psi_{j0}|\le\delta} \frac{1}{1-F(t_1)} \left\| \frac{\partial^2 F(t_1,\psi)}{\partial\psi\partial\psi^T} \right\| \right) < \infty, \quad 1 \le j \le J, \tag{7.15}$$

for some $\delta > 0$. Now in this interior case we can ignore the factor of $1 - F(t_1)$ in the denominator of (7.15), and likewise in the other terms in (7.7)−(7.9), because when $|\psi - \psi_{j0}| \le \delta$,

$$1-F(t_1) = 1-\sum_{j=1}^J p_j+\sum_{j=1}^J p_j\,(1-F(t_1,\psi_j)) \ge 1-\sum_{j=1}^J p_j \ge 1-\sum_{j=1}^J (p_{j0}+\delta) \ge a_0 > 0, \tag{7.16}$$

once $\delta < (1/2)(1-\sum_{j=1}^J p_{j0}) = a_0$, say. Now, using (7.11), we see that (7.15) is implied by (3.9). The third term on the right hand side of (7.7) requires

$$E\left(\sup_{|\psi-\psi_{j0}|\le\delta} \left\| \frac{\partial F(t_1,\psi)}{\partial\psi} \frac{\partial F(t_1,\psi)}{\partial\psi^T} \right\| \right) < \infty,$$

and by the Cauchy-Schwarz inequality this follows from (3.9) in a similar way as for the second term. We can similarly deal with (7.8), and (7.9) is integrable in the interior case with no assumptions at all. Putting these together we get (B3*) and hence (B3).

Finally, under our i.i.d assumptions, (B5) follows immediately from the multivariate Central Limit Theorem, since $S_n(\theta_0)$ is a sum of i.i.d random vectors, each with mean 0 and finite covariance matrix D.

Proof of Theorem 2. Again we use the Vu and Zhou (1997) formulation, but for Theorem 2 we need to deal with $\Omega = \Omega_r$, the subset of Θ to which θ_0 is confined under the null hypothesis. Ω is required to have similar properties as τ. In the present proof this is straightforward. As θ_0 and θ^* are in the interior of Θ, and $\mathbf{S}_r$ can be rescaled to $\mathbf{R}^r \times \{0\}^{(J+1)s-r}$, $C_\Omega = \mathbf{S}_r + \theta^*$ and $\tilde{C}_{\Omega_n} = \tilde{C}_\Omega = \mathbf{R}^r \times \{0\}^{(J+1)s-r}$ satisfy the requirements in Vu and Zhou (1997). Therefore, Theorem 2.1 of Vu and Zhou (1997) gives the required MLE $\hat{\theta}_n^{(2)}$ over Ω_r, while their Theorem 2.2, in conjunction with the τ and $\hat{\theta}_n^{(1)}$ in the proof of Theorem 1, concludes that

$$\begin{aligned} d_n^{(1)} &= 2\{\mathcal{L}_n(\hat{\theta}_n^{(1)}) - \mathcal{L}_n(\hat{\theta}_n^{(2)})\} \xrightarrow{D} \inf_{\theta\in\mathbf{R}^r\times\{0\}^{(J+1)s-r}} |N-\theta|^2 - \inf_{\theta\in\mathbf{R}^{(s+1)J}} |N-\theta|^2 \\ &= N_{r+1}^2 + \cdots + N_{(s+1)J}^2 - 0 \sim \chi^2_{(s+1)J-r}, \end{aligned}$$

where $N = (N_1, \ldots, N_{(s+1)J})$ is a standard normal random vector in $(s+1)J$ dimensions. This proves (3.11).

Proof of Theorem 3. The proof is basically a straightforward checking of the conditions of Theorem 1, though the details are lengthy. Here we give an outline for the Weibull distribution, which of course covers the exponential distribution as a special case. The arguments for the Gamma distribution are omitted for the sake of brevity.

For the Weibull distribution, (3.6) is obvious provided G does not degenerate at 0. Next, $\log f(t,\psi) = \log\alpha + \alpha\log\lambda + (\alpha-1)\log t - (\lambda t)^\alpha$, so

$$\frac{\partial \log f(t,\psi)}{\partial\psi} = \begin{bmatrix} \frac{\alpha}{\lambda} - \alpha(\lambda t)^{\alpha-1}t \\ \frac{1}{\alpha} + \log(\lambda t) - (\lambda t)^\alpha \log(\lambda t) \end{bmatrix}.$$

Thus for $(z,a) = (z_1, z_2, a) \in \mathbf{R}^3 - \{0\}$,

$$z^T \frac{\partial \log f(t,\psi_{j0})}{\partial\psi} + a = z_1 \frac{\alpha_{j0}}{\lambda_{j0}}(1-y) + \frac{z_2}{\alpha_{j0}}(1 + (1-y)\log y) + a,$$

where $y = (\lambda_{j0}t)^{\alpha_{j0}}$, which clearly has no zeroes or only isolated zeroes in $t \in (0,\infty)$, so (3.7) holds. Furthermore,

$$\frac{\partial^2 \log f(t,\psi)}{\partial\psi\partial\psi^T} = \begin{bmatrix} -\frac{\alpha}{\lambda^2} - \alpha(\alpha-1)(\lambda t)^{\alpha-2}t^2 & \frac{1}{\lambda} - (\lambda t)^{\alpha-1}t - \alpha t(\lambda t)^{\alpha-1}\log(\lambda t) \\ \frac{1}{\lambda} - (\lambda t)^{\alpha-1}t - \alpha t(\lambda t)^{\alpha-1}\log(\lambda t) & -\frac{1}{\alpha^2} - (\lambda t)^\alpha \log^2(\lambda t) \end{bmatrix}.$$

For t near 0, the above matrix is $O(1)$ and, for large t, it is $O(t^{\alpha+\eta})$ for some $\eta > 0$. On the other hand $F(dt,\psi) = f(t,\psi)dt$ is $O(t^{\alpha-1})$ for t near 0, and exponentially small as $t \to \infty$. Now for $|\psi - \psi_{j0}| \le \delta$, λ and α are bounded above and below by $\lambda_{j0} \pm \delta$ and $\alpha_{j0} \pm \delta$ respectively, provided we keep $0 < \delta < \min_j(\lambda_{j0}, \alpha_{j0})$. As a result, the integral in (3.8) is uniformly bounded in u and $|\psi - \psi_{j0}| \le \delta$, which implies (3.8).

It remains to show (3.9). Since $h(t,\psi) = \alpha\lambda(\lambda t)^{\alpha-1}$ and $H(t,\psi) = (\lambda t)^\alpha$, we obtain

$$\partial H(t,\psi)/\partial\psi = [\,\alpha(\lambda t)^{\alpha-1}t \quad (\lambda t)^\alpha \log(\lambda t)\,]^T \quad \text{and}$$

$$\frac{\partial^2 H(t,\psi)}{\partial\psi\partial\psi^T} = \begin{bmatrix} \alpha(\alpha-1)(\lambda t)^{\alpha-2}t^2 & (\lambda t)^{\alpha-1}t + \alpha t(\lambda t)^{\alpha-1}\log(\lambda t) \\ (\lambda t)^{\alpha-1}t + \alpha t(\lambda t)^{\alpha-1}\log(\lambda t) & (\lambda t)^\alpha \log^2(\lambda t) \end{bmatrix}.$$

These are bounded in norm by terms that are $O(t^{\alpha-\eta})$ for t near 0 and $O(t^{\alpha+\eta})$ for t near ∞, where $\eta > 0$ can be arbitrarily small. Thus, as

$$\frac{\partial F}{\partial\psi} = e^{-H}\frac{\partial H}{\partial\psi} \qquad \text{and} \qquad \frac{\partial^2 F}{\partial\psi\partial\psi^T} = e^{-H}\frac{\partial^2 H}{\partial\psi\partial\psi^T} - e^{-H}\frac{\partial H}{\partial\psi}\frac{\partial H}{\partial\psi^T},$$

there exists a constant C such that

$$\left\|\frac{\partial^2 F(u,\psi)}{\partial\psi\partial\psi^T}\right\| + \left|\frac{\partial F(u,\psi)}{\partial\psi}\right|^2 \le C\exp(-(\lambda u)^\alpha)(u^{\alpha-2\eta} + u^{2\alpha+2\eta}) \qquad (u > 0).$$

The last term is uniformly bounded in u and $|\psi - \psi_{j0}| \le \delta$, provided δ and η are sufficiently small. This proves (3.9).

Proof of Theorem 4. We proceed as in the proof of Theorem 2, but in order to apply the Vu and Zhou (1997) results it turns out to be better here to transform the parameter space from $\Theta = \Psi^J \times S_J$ to $\Theta' = H\Theta$ by the nonsingular linear transform

$$H = \begin{bmatrix} I & 0 \\ 0 & h \end{bmatrix}_{(s+1)J\times(s+1)J} \qquad \text{where} \qquad h = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{J\times J}.$$

This takes $\theta = (\psi_1,\ldots,\psi_J,p_1,\ldots,p_J)$ to $\theta' = H\theta = (\psi_1,\ldots,\psi_J,p_1,\ldots,p_{J-1},p_1 + \cdots + p_J)$ with corresponding "true" value $\theta_0' = H\theta_0 = (\psi_{10},\ldots,\psi_{J0},p_{10},\ldots,p_{J-1,0}$ $,1)$ under $H_0^{(2)}$ (see (4.2)).

The parameter set under the alternative, $\tau = \Theta$, transforms to $\tau' = \Theta' = H\Theta = \{\theta' \in \Psi^J \times (0,\infty)^J : \theta'_{sJ+1} + \cdots + \theta'_{(s+1)J-1} < \theta'_{(s+1)J} \le 1\}$. The corresponding neighborhood of θ_0 is

$$\mathcal{N}'(\delta) = \prod_{j=1}^{J}\prod_{\ell=1}^{s}[\psi_{j0}^{(\ell)} - \delta, \psi_{j0}^{(\ell)} + \delta] \times \prod_{j=1}^{J-1}[p_{j0} - \delta, p_{j0} + \delta] \times [1-\delta, 1+\delta]. \quad (7.17)$$

Here $\psi_{j0}^{(\ell)}$, $1 \le \ell \le s$, denote the components of ψ_{j0}. When δ is sufficiently small, $\tau' \cap \mathcal{N}'(\delta)$ has the same form as (7.17) except that the range of $\theta'_{(s+1)J}$ in $\tau' \cap \mathcal{N}'(\delta)$ is $[1-\delta, 1]$ instead of $[1-\delta, 1+\delta]$. The containing cone $C_{\tau'}$, with vertex at θ'_0, can be taken as $C_{\tau'} = \mathbf{R}^{(s+1)J-1} \times (-\infty, 1]$, which coincides with τ' on $\mathcal{N}'(\delta)$, so (A2) of Vu and Zhou (1997) is satisfied for τ', $C_{\tau'}$, and $\mathcal{N}'(\delta)$.

We must check that conditions (B1)$-$(B5) hold in the transformed parameter space; call them (B1$'$)$-$(B5$'$). We will deduce them from (B1)$-$(B5), for which much of the groundwork was laid in the proof of Theorem 1 because with H a linear transform it is easy to go from the (B)-conditions to the (B$'$)-conditions. For (B1$'$) we get

$$E\left(S'_n(\theta'_0)\right) = E\left(\frac{\partial \mathcal{L}_n(\theta')}{\partial \theta'}\right)\bigg|_{\theta'=\theta'_0} = H^{-T} E\left(\frac{\partial \mathcal{L}_n(\theta)}{\partial \theta}\right)\bigg|_{\theta=\theta_0} = H^{-T} E\left(S_n(\theta_0)\right) = 0$$

(by (B1)), and the matrix $nD' = D'_n = E\left(S'_n(\theta'_0)\right)\left(S'_n(\theta'_0)\right)^T = nH^{-T}DH^{-1}$ is obviously finite and nonsingular, so (B2$'$) holds. For (B3$'$), let $A > 0$, $n \ge 1$, and note that

$$\begin{aligned} N'_n(A) &= \left\{\theta' : n(\theta' - \theta'_0)^T D'(\theta' - \theta'_0) \le A^2\right\} \\ &= \left\{\theta : n(\theta - \theta_0)^T H^T (H^{-T} D H^{-1}) H(\theta - \theta_0) \le A^2\right\} = N_n(A), \end{aligned}$$

so that, letting $\mathcal{F}'_n(\theta') = \partial^2 \mathcal{L}_n(\theta')/\partial(\theta')\partial(\theta')^T = H^{-T}\mathcal{F}_n(\theta)H^{-1}$, we have

$$\sup_{\theta' \in N'_n(A)} \left\| \frac{1}{n}\mathcal{F}'_n(\theta') - D' \right\| \le \|H^{-1}\|^2 \sup_{\theta \in N_n(A)} \left\| \frac{1}{n}\mathcal{F}_n(\theta) - D \right\|.$$

Hence (B3$'$) will follow from (B3); but we must check this for the current choice of θ_0 on the boundary. (3.8) is used to deal with (7.14) just as in the proof of Theorem 1. But when it comes to the terms in (7.7)$-$(7.9) with the factor $(1 - F(t_i))$ in the denominator, we can no longer use (7.16) to keep this positive since $\sum p_{j0} = 1$ in the present case. This is where the more stringent integrability conditions (4.3)$-$(4.4) are needed. To bound $1 - F(t_i)$ below we proceed as follows. Take $\delta > 0$ and define $H_\delta(t)$ as in (4.1). Use Taylor's expansion to write, for $\psi \in \Psi$,

$$\log\left(\frac{1 - F(t,\psi)}{1 - F(t,\psi_{j0})}\right) = (\psi - \psi_{j0})^T \frac{\partial \log(1 - F(t,\xi))}{\partial \psi} = -(\psi - \psi_{j0})^T \frac{\partial H(t,\xi)}{\partial \psi}.$$

Here $\xi = a\psi + (1-a)\psi_{j0}$ for some $0 \le a \le 1$. Keep $|\psi - \psi_{j0}| \le \delta$, so $|\xi - \psi_{j0}| \le \delta$. Then $1 - F(t,\psi) \ge e^{-\delta H_\delta(t)}(1 - F(t,\psi_{j0}))$. Now choose $\delta < (\min_{1 \le j \le J} p_{j0})/2$, so that $p_j \geq p_{j0} - \delta \geq (1/2)p_{j0}$, $1 \leq j \leq J$. Thus

$$1 - F(t) = 1 - \sum_{j=1}^{J} p_j + \sum_{j=1}^{J} p_j \left(1 - F(t, \psi_j)\right) \geq \frac{1}{2} e^{-\delta H_\delta(t)} \sum_{j=1}^{J} p_{j0} \left(1 - F(t, \psi_{j0})\right)$$
$$= e^{-\delta H_\delta(t)} (1 - F_0(t))/2, \tag{7.18}$$

where $F_0(t)$ is defined just before (4.1). By (7.11), any measurable function integrated against $1-c_i$ brings a factor of $1-F_0(u)$ into the numerator, cancelling any similar factor introduced by a factor of $1-F(t_i)$ in the denominator via (7.18). Thus the second term on the right hand side of (7.7) requires us to check the finiteness of

$$E\Big[\sup_{|\psi-\psi_{j0}|\leq\delta} \frac{1-c_1}{1-F(t_1)} \Big\| \frac{\partial^2 F(t_1,\psi)}{\partial\psi\partial\psi^T} \Big\|\Big] \leq E\Big[\frac{2e^{\delta H_\delta(t_1)}(1-c_1)}{1-F_0(t_1)} \sup_{|\psi-\psi_{j0}|\leq\delta} \Big\| \frac{\partial^2 F(t_1,\psi)}{\partial\psi\partial\psi^T} \Big\|\Big]$$
$$= 2E\left[e^{\delta H_\delta(u)} \sup_{|\psi-\psi_{j0}|\leq\delta} \left\| \frac{\partial^2 F(u,\psi)}{\partial\psi\partial\psi^T} \right\|\right].$$

The last is finite by (4.3). For the third term on the right hand side of (7.7) we look at

$$E\Big[\frac{e^{2\delta H_\delta(t_1)}(1-c_1)}{(1-F_0(t_1))^2} \sup_{|\psi-\psi_{j0}|\leq\delta} \Big| \frac{\partial F(t_1,\psi)}{\partial\psi} \Big|^2\Big] = E\Big[\frac{e^{2\delta H_\delta(u)}}{1-F_0(u)} \sup_{|\psi-\psi_{j0}|\leq\delta} \Big| \frac{\partial F(u,\psi)}{\partial\psi} \Big|^2\Big]$$

and this is finite by (4.4) (note that we can replace 2δ by δ just by taking a larger δ, if necessary). For (7.9) we need

$$E\Big[(1-c_1) \sup_{|\psi-\psi_{j0}|\leq\delta} \frac{1}{(1-F(t_1))^2}\Big] \leq E\Big[\frac{e^{2\delta H_\delta(u)}}{1-F_0(u)}\Big]$$

and this is covered by the "1" in (4.4). Similarly we deal with (7.8) (use the Cauchy-Schwartz inequality and (4.4) for the second term on the right hand side of (7.8)). This completes the verification of (B3*). Finally for (B5′), simply note that (B5) implies

$$(D')^{-1/2} \frac{S_n'(\theta_0')}{\sqrt{n}} = (H^{-T} D H^{-1})^{-1/2} H^{-T} D^{1/2} \, \frac{D^{-1/2} S_n(\theta_0)}{\sqrt{n}} \xrightarrow{D} N(0, I).$$

Now we return to the A-conditions. We found $C_\tau' = \mathbf{R}^{(s+1)J-1} \times (-\infty, 1]$, so the centered, rescaled cone for the alternative is $\tilde{C}_{\tau_n'} = \{\tilde{\theta}' = \sqrt{n}(H^{-T} D H^{-1})^{T/2} \times (\theta' - \theta_0') : \theta' \in C_\tau'\}$. Because the Cholesky square root $(H^{-T} D H^{-1})^{T/2}$ is upper triangular, $\tilde{C}_{\tau_n'}$ asymptotically coincides with $\tilde{C}_{\tau'} = \mathbf{R}^{(s+1)J-1} \times (-\infty, 0]$. Now $H_0^{(2)}$ transforms to $H_0^{(2)\prime} : \theta' = \theta_0' \in \Omega' = \Psi^J \times S_{J-1} \times \{1\}$. $\mathcal{N}'(\delta)$

is already defined in (7.17) and for the containing cone for the null we take $C_{\Omega'} = \mathbf{R}^{(s+1)J-1} \times \{1\}$, which we re-center/scale to $\tilde{C}_{\Omega'_n} = \mathbf{R}^{(s+1)J-1} \times \{0\}$, asymptotically coincident with $\tilde{C}_{\Omega'} = \mathbf{R}^{(s+1)J-1} \times \{0\}$. Clearly (A2′) is satisfied for Ω', $\tilde{C}_{\Omega'}$, $\mathcal{N}'(\delta)$, so Theorems 2.1 and 2.2 of Vu and Zhou (1997) give the local uniqueness and consistency of $\hat{\theta}_n^{(1)}$ and $\hat{\theta}_n^{(3)}$, and that

$$\begin{aligned} d_n^{(2)} &\xrightarrow{D} \inf_{\theta\in\mathbf{R}^{(s+1)J-1}\times\{0\}} |N-\theta|^2 - \inf_{\theta\in\mathbf{R}^{(s+1)J-1}\times(-\infty,0]} |N-\theta|^2 \\ &= N^2_{(s+1)J} - N^2_{(s+1)J} I(N_{(s+1)J}>0) = N^2_{(s+1)J} I(N_{(s+1)J}\le 0) \sim N_1^2 I(N_1\ge 0). \end{aligned}$$

Proof of Theorem 5. The proof is similar to those of Theorems 3 and 4, with condition (4.5) or (4.6) taking care of (4.3)−(4.4). Details are omitted.

Acknowledgement

This work is partially supported by Hong Kong Polytechnic University Internal Research Grant No. A-PB50, and by ARC Grant No. A69803103. We are grateful to two anonymous referees for their valuable comments and suggestions which led to the improvement of this paper.

References

Andersen, P. K., Borgan, O., Gill, R. G. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Cox, D. R. and Oakes, D. (1984). *Analysis Survival Data.* Chapman and Hall, London.

David, H. A. and Moeschberger, M. L. (1978). *The Theory of Competing Risks.* Griffin, London.

Elandt-Johnson, R. C. and Johnson, N. L. (1980). *Survival Models and Data Analysis.* Wiley, New York.

Escarela, G. Francis, B. and Soothill, K. (2000). Competing risks, persistence and desistence in crime: a case study of ex-offenders convicted of indecent assault on a female. *J. Quantitative Criminology* **16**, 385-414.

Farewell, V. T. (1977). A model for a binary variable with time censored observations. *Biometrika* **64**, 43-46.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

Gaynor, J. J., Feuer, E. J., Tan, C. C., Wu, D. H., Little, C. R., Straus, D. J., Clarkson, B. D. and Brennan, M. F. (1993). On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. *J. Amer. Statist. Soc.* **88**, 400-409.

Ghitany, M. E. and Maller, R. A. (1992). Asymptotic results for exponential mixture models with long term survivors. *Statistics* **23**, 321-336.

Ghitany, M. E., Maller, R. A. and Zhou, X. (1994). Exponential mixture models with long term survivors and covariates. *J. Multivariate Anal.* **49**, 218-241.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, N.Y.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Lagakos, S. W., Sommer, C. J. and Zelen, M. (1978). Semi-Markov models for partially censored data. *Biometrika* **65**, 311-317.

Larson, M. G. and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Appl. Statist.* **34**, 201-211.

Lemdani, M. and Pons, O. (1997). Estimation and tests in finite mixture models for censored survival data. *Statistics* **29**, 363-388.

Lin, D. Y. (1997). Nonparametric inference for cumulative incidence functions in competing risks studies. *Statist. in Medicine* **16**, 901-910.

Maller, R. A. (1993). Quadratic negligibility and the asymptotic normality of operator normed sums. *J. Multivariate Anal.* **44**, 191-219.

Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors.* Wiley, Chichester.

Pepe, M. S. and Mori, M. (1993). Kaplan-Meier, marginal or conditional probability curves in summarising competing risks failure time data? *Statist. in Medicine* **12**, 737-751.

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

Self, S. G. and Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Amer. Statist. Assoc.* **82**, 605-610.

Tai, B-C., Machin, D., White, I. and Gebski, V. (2001). Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches. *Statist. in Medicine* **20**, 661-684.

Vu, H. T. V., Maller, R. A. and Klass, M. J. (1996). On the studentisation of random vectors. *J. Multivariate Anal.* **57**, 142-155.

Vu, H. T. V., Maller, R. A. and Zhou, X. (1998). Mixture models for failure data. *Ann. Inst. Statist. Math.* **50**, 627-653.

Vu, H. T. V. and Zhou, X. (1997). Generalisation of likelihood ratio tests under nonstandard conditions. *Ann. Statist.* **25**, 897-916.

Zhou, X. and Maller, R. A. (1995). The likelihood ratio test for the presence of immunes in a censored sample. *Statistics* **27**, 181-201.

Department of Accounting and Finance, University of Western Australia, 35 Stirling Highway, Crawley 6009, Western Australia.

E-mail: maller@maths.uwa.edu.au

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.

E-mail: maxzhou@polyu.edu.hk

(Received June 2000; accepted January 2002)