Ann. Inst. Statist. Math.
Vol. 50, No. 4, 627–653 (1998)

ASYMPTOTIC PROPERTIES OF A CLASS OF MIXTURE MODELS FOR FAILURE DATA: THE INTERIOR AND BOUNDARY CASES

H. T. V. VU[1], R. A. MALLER[2] AND X. ZHOU[3]

[1] *Department of Public Health, University of Western Australia, Nedlands, WA 6907, Australia*
[2] *Department of Mathematics, University of Western Australia, Nedlands, WA 6907, Australia*
[3] *Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China*

(Received February 17, 1997; revised September 30, 1997)

Abstract. We analyse an exponential family of distributions which generalises the exponential distribution for censored failure time data, analogous to the way in which the class of generalised linear models generalises the normal distribution. The parameter of the distribution depends on a linear combination of covariates via a possibly nonlinear link function, and we allow another level of heterogeneity: the data may contain "immune" individuals who are not subject to failure. Thus the data is modelled by a mixture of a distribution from the exponential family and a "mass at infinity" representing individuals who never fail. Our results include large sample distributions for parameter estimators and for hypothesis test statistics obtained by maximising the likelihood of a sample. The asymptotic distribution of the likelihood ratio test statistic for the hypothesis that there are no immunes present in the population is shown to be "non-standard"; it is a 50-50 mixture of a chi-squared distribution on 1 degree of freedom and a point mass at 0. Our analysis clearly shows how "negligibility" of individual covariate values and "sufficient followup" conditions are required for the asymptotic properties.

Key words and phrases: Censored survival data, immune proportion, covariates, mixture models, failure time data, exponential family, boundary hypothesis tests.

1. Introduction

The use of "generalised linear models" ("glms"), proposed by Nelder and Wedderburn (1972), revolutionised statistics by extending the class of normal distributions to a wide ranging exponential family of distributions with (possibly discrete) density of the form

$$f(y) = e^{(y\theta - g(\theta))/a(\phi) + h(y,\phi)}. \tag{1.1}$$

Here θ and ϕ are parameters to be estimated. In a similar way we propose and analyse in this paper a generalisation of the class of exponential distributions to an exponential family of positive random variables with (possibly discrete) density

$$f(t) = e^{-\lambda q(t) + g(\lambda) + h(t)} \qquad (t \geq 0), \tag{1.2}$$

where $q(t)$ is a nonnegative function and λ is a parameter. These distributions provide a variety of models for the analysis of censored non-negative survival time data. We allow a further generalisation; the population may contain an "immune or cured proportion" of individuals who never fail. We cater for this by modelling the distribution from which the data is drawn as a mixture of the distribution given by (1.2) and a "point mass at infinity". There has been a great deal of recent interest in such "cure" models in medical statistics; see for example the literature reviewed in Maller and Zhou (1996).

One of the major strengths of the generalised linear model approach is its ability to deal with a covariate vector of information x_i associated with a response variable Y_i having density (1.1) corresponding to parameter θ_i. Covariate information is linked to the parameter θ_i via the relation $\Phi(\mathbb{E}(Y_i)) = \beta^T x_i$, where Φ is a known "link" function and β is a vector of parameters. In exactly the same way we link the parameter λ in (1.2) to a linear combination of the corresponding covariate vector, and allow the "immune proportion", if any, to depend on (possibly different) covariates, again via a generalised linear setup. This formulation allows us to test for the presence of immunes. This is a test of a boundary value hypothesis, leading to a "non-standard" asymptotic result.

Our analysis is based on the likelihood theory, and our methods of proof are similar in spirit to those of Fahrmeir and Kaufmann (1985), who dealt with glms. Like them, we strive for minimal conditions on the covariates. These take the form of "uniform asymptotic negligibility" conditions and are much weaker than requiring the covariates to be uniformly bounded, for example. However, there are two respects in which our analysis is quite different to that for glms. In the first place we allow for (right) censoring of the failure times, so the question of "sufficient followup" arises. This is, essentially, the question of how "heavy" the tail of a censoring distribution may be, relative to the survival distribution, while still permitting consistent estimation. It is manifested in our analysis by an integral condition relating to those tails. Secondly, allowance for immunes in the population also raises considerations of the interplay between the censoring and survival distributions, and these in turn interact with issues concerning the covariates. Our large sample analysis, by drawing out these relationships, highlights their importance for data analysis. We discuss these matters further in Sections 3 and 4, where we also relate them to examples and to some previous work on special cases.

2. Model specification and assumptions

We will work with an exponential family whose densities take the form

$$f_\lambda(t) = a(t)e^{-\lambda q(t)+g(\lambda)}, \qquad t \geq 0, \tag{2.1}$$

with respect to a σ-finite measure μ whose support $\boldsymbol{T} = \{t > 0 : a(t) > 0\}$ does not depend on λ. The function q is nonnegative on $[0, \infty)$. The natural parameter space of the exponential family is

$$\left\{\lambda \in \mathbb{R} : \int_{[0,\infty)} f_\lambda(y)\mu(dy) < \infty\right\}.$$

We will take the "parameter space" Λ of interest to us to be the interior of this space. Clearly, Λ is an open interval of the form $(-\infty, \infty)$ or (a, ∞), $a > -\infty$. For each λ in Λ, we assume that

$$F_\lambda(t) = \int_{[0,t]} f_\lambda(y)\mu(dy), \qquad t \geq 0, \tag{2.2}$$

defines a proper non-degenerate cumulative distribution function (c.d.f.). In practice, the main situations of interest are when $\mu(dy) = dy$ is Lebesgue measure, or when μ is a counting measure on a discrete set of mass points $\{a_j\}_{j\geq 1}$. Let

$$\tau_- = \inf\{t > 0 : F_\lambda(t) > 0\} \qquad \text{and} \qquad \tau_+ = \sup\{t > 0 : F_\lambda(t) < 1\}$$

denote the left and right extremes of F_λ, so that $0 \leq \tau_- < \tau_+ \leq \infty$. These do not depend on λ since $\boldsymbol{T}$ does not depend on λ. Also define the tail of F_λ on $[0, \infty)$ by

$$\bar{F}_\lambda(t) = \int_{(t,\infty)} f_\lambda(y)\mu(dy). \tag{2.3}$$

By applying Theorem 7.2 in Barndorff-Nielsen (1978) with the positive σ-finite measure ν defined by $d\nu = a(t)d\mu$, we obtain

(i) the function g is infinitely differentiable on Λ, and

(ii) $\mathbb{E}(q^j(\tilde{t}_\lambda))$ is finite for all $j = 1, 2, \ldots$ and all $\lambda \in \Lambda$, where $\tilde{t}_\lambda$ is a random variable with c.d.f. F_λ and the expectation is with respect to F_λ.

The data we wish to analyse consist of censored survival times t_i with associated covariates x_i and y_i, $1 \leq i \leq n$, which may be the same or more generally may be subvectors of an overall covariate vector associated with t_i. Let t_i^*, $i = 1, 2, \ldots$, be independent random variables with possibly improper c.d.f.'s $F_i^* = p_i F_{\lambda_i}$, where the F_{λ_i} are proper c.d.f.'s as in (2.2), and $0 < p_i \leq 1$. Assume that the parameters are "linked" to linear combinations of the covariates by

$$\lambda_i = \eta(\kappa_i), \qquad \text{where} \qquad \kappa_i = \beta^T x_i, \tag{2.4}$$

and

$$p_i = \zeta(\rho_i), \qquad \text{where} \qquad \rho_i = \gamma^T y_i. \tag{2.5}$$

Here η is a function from $\mathbb{R}$ to Λ and ζ is a real function. The covariates x_i and y_i are fixed (non-stochastic) vectors in $\mathbb{R}^{k_1}$ and $\mathbb{R}^{k_2}$ which may vary over individuals i, and β and γ are k_1 and k_2-dimensional vectors of parameters to be estimated. Then (2.4)–(2.5) together with the vector $\theta = (\beta^T \gamma^T)^T$ of dimension $k = k_1 + k_2$ represent our "generalised linear" parameterisation of the model. The parameter space Θ to which θ belongs is supposed to be of the form $\Theta = \Theta_1 \times \cdots \times \Theta_k$ where the Θ_j, $1 \le j \le k$, are nonempty intervals.

The improper distributions $F_i^*(t) = p_i F_{\lambda_i}(t)$ can be thought of as arising from unobserved auxiliary variables B_i with $\mathbb{P}(B_i = 1) = p_i = 1 - \mathbb{P}(B_i = 0)$ such that, for all $t \ge 0$, $\mathbb{P}(t_i^* \le t \mid B_i = 1) = F_{\lambda_i}(t)$ and $\mathbb{P}(t_i^* \le t \mid B_i = 0) = 0$. Thus the survival function of t_i^* is

$$\begin{aligned} \mathbb{P}(t_i^* > t) &= (1 - p_i)\mathbb{P}(t_i^* > t \mid B_i = 0) + p_i \mathbb{P}(t_i^* > t \mid B_i = 1) \\ &= 1 - p_i + p_i \bar{F}_{\lambda_i}(t). \end{aligned} \tag{2.6}$$

The interpretation of this model is that, when $B_i = 1$, individual i is subject to failure at a time t_i^* drawn from the c.d.f. $F_{\lambda_i}(t)$, while when $B_i = 0$ the individual never fails, or, formally, $t_i^* = \infty$. Let u_i be censoring variables with proper c.d.f.'s G_i which are noninformative, i.e., do not depend on the parameter θ. The observed data consist of n observations on $t_i = t_i^* \wedge u_i$, where we take $t_i^* \wedge u_i = u_i$ when $t_i^* = \infty$. When $B_i = 1$ we observe a censored value $t_i = u_i$ if $t_i^* > u_i$ and an uncensored value $t_i = t_i^*$ otherwise. When $B_i = 0$ the observation on the individual i is censored at value u_i. We do not observe B_i, so we do not know whether individual i is immune or not, but we do observe censoring variables $c_i = 1_{\{t_i^* \le u_i\}}$. An important part of our analysis will be to decide whether or not immunes are in fact present in the population. We assume that the t_i^* and u_i are all conditionally independent, given the B_i. Our analysis then takes place conditional on the values of the B_i, which we can ignore from now on.

The "true" value of θ is $\theta_0 = (\beta_0^T \gamma_0^T)^T$ where β_0 and γ_0 are the "true" values of β and γ. Let $\lambda_{i0} = \eta(\kappa_{i0})$ where $\kappa_{i0} = \beta_0^T x_i$, and $p_{i0} = \zeta(\rho_{i0})$ where $\rho_{i0} = \gamma_0^T y_i$. Rather than $f_{\lambda_{i0}}$, $F_{\lambda_{i0}}$, and $\bar{F}_{\lambda_{i0}}$, we will write f_{i0}, F_{i0}, and $\bar{F}_{i0}$. Also for brevity, let f_{λ_i}, F_{λ_i} and $\bar{F}_{\lambda_i}$ be denoted by f_i, F_i and $\bar{F}_i$. To be precise in specifying our model, we should introduce random variables t_{i0}^* with distributions $F_{i0}^* = p_{i0} F_{i0}$, and let $t_{i0} = t_{i0}^* \wedge u_i$, but we will continue to denote those by t_i^* and t_i. Expectations (except in Lemma 5.1 below) will be with respect to the "true" distributions.

The likelihood of a sample of n of the t_i is obtained by factoring in p_i times the density $f_i(t_i)$ if t_i is uncensored, or the survival function (2.6), evaluated at t_i, if t_i is censored. Thus we can write the log-likelihood function as

$$\mathcal{L}_n(\theta) = \sum_{i=1}^{n} \{c_i(\log p_i + g(\lambda_i) - \lambda_i q(t_i) + \log(a(t_i))) + (1 - c_i)\log(1 - p_i F_i(t_i))\}. \tag{2.7}$$

We consider two cases:

The interior case: $0 < p_{i0} < 1$, $1 \le i \le n$.

So that the main points are not obscured by technicalities, we will only consider the case when $\zeta(\rho) = e^{\rho}/(1+e^{\rho})$ in (2.5). (This is one of the most useful cases in practise.) Then we have $p_i = \zeta(\rho_i) = e^{\gamma^T y_i}/(1+e^{\gamma^T y_i})$. Note that $\mathcal{L}_n(\theta)$ is then finite a.s. for each θ. Assume that $\Theta_j = (a_j, b_j)$ where a_j and b_j are constants such that $a_j < \theta_{j0} < b_j$, so θ_0 is an interior point of Θ.

The boundary case: $p_{i0} = 1$, $1 \le i \le n$.

For this case we take $k_2 = 1$ and $y_i = 1$ for $1 \le i \le n$. The function ζ is irrelevant here, but we shall take $\zeta(\rho) = \rho$ since then the interior and boundary cases can be formulated together. Hence $\gamma = p_i = p$, say, and $\gamma_0 = p_0 = 1$. As a consequence of assumption (F5) below, G_i attributes no mass to τ_+ or to points larger than it if $\tau_+ < \infty$. Thus the log-likelihood will be finite a.s. in this case too. For $1 \le j \le k_1$, assume that $\Theta_j = (a_j, b_j)$ where a_j and b_j are any constants such that $a_j < \theta_{j0} < b_j$, and let $\Theta_k = (a_k, 1]$ where $a_k < 1$.

To analyse the maximum likelihood estimates (MLEs), we need the derivatives of the log-likelihood. From (2.7), the derivative of $\mathcal{L}_n(\theta)$ with respect to θ is

$$S_n(\theta) = \frac{\partial}{\partial\theta}\mathcal{L}_n(\theta) = \sum_{i=1}^{n} \boldsymbol{X}_i s_i(\theta) \tag{2.8}$$

where the 2-vector $s_i(\theta)$ has components

$$s_{i1}(\theta) = \frac{\partial \mathcal{L}_n(\theta)}{\partial\kappa_i} = c_i\eta'(\kappa_i)(g'(\lambda_i) - q(t_i)) - \frac{(1-c_i)p_i\eta'(\kappa_i)}{1-p_iF_i(t_i)}\frac{\partial F_i(t_i)}{\partial\lambda_i} \tag{2.9}$$

and

$$s_{i2}(\theta) = \frac{\partial \mathcal{L}_n(\theta)}{\partial\rho_i} = \frac{c_i\zeta'(\rho_i)}{p_i} - \frac{(1-c_i)\zeta'(\rho_i)F_i(t_i)}{1-p_iF_i(t_i)}, \tag{2.10}$$

and the matrices

$$\boldsymbol{X}_i = \begin{bmatrix} x_i & 0 \\ 0 & y_i \end{bmatrix}$$

are $k \times 2$ non-stochastic matrices. The negative of the second derivative of $\mathcal{L}_n(\theta)$ is a $k \times k$ symmetric matrix

$$\boldsymbol{F}_n(\theta) = -\frac{\partial^2}{\partial\theta^2}\mathcal{L}_n(\theta) = \sum_{i=1}^{n} \boldsymbol{X}_i \mathcal{F}_i(\theta) \boldsymbol{X}_i^T \tag{2.11}$$

where the 2×2 symmetric matrices $\mathcal{F}_i(\theta)$ have elements

$$\begin{aligned} f_i^{11}(\theta) = {} & c_i[\eta''(\kappa_i)(q(t_i) - g'(\lambda_i)) - (\eta'(\kappa_i))^2 g''(\lambda_i)] \\ & + (1-c_i)\left\{\frac{p_i\eta''(\kappa_i)}{1-p_iF_i(t_i)}\frac{\partial F_i(t_i)}{\partial\lambda_i}\right. \\ & \qquad + \frac{p_i^2(\eta'(\kappa_i))^2}{(1-p_iF_i(t_i))^2}\left(\frac{\partial F_i(t_i)}{\partial\lambda_i}\right)^2 \end{aligned} \tag{2.12}$$

$$+\frac{p_i(\eta'(\kappa_i))^2}{1-p_iF_i(t_i)}\frac{\partial^2 F_i(t_i)}{\partial\lambda_i^2}\Bigg\},$$

$$f_i^{22}(\theta) = c_i\left\{\frac{(\zeta'(\rho_i))^2}{p_i^2} - \frac{\zeta''(\rho_i)}{p_i}\right\} + (1-c_i)\left\{\frac{(\zeta'(\rho_i))^2F_i^2(t_i)}{(1-p_iF_i(t_i))^2} + \frac{\zeta''(\rho_i)F_i(t_i)}{1-p_iF_i(t_i)}\right\} \tag{2.13}$$

and

$$f_i^{12}(\theta) = \frac{(1-c_i)\eta'(\kappa_i)\zeta'(\rho_i)}{(1-p_iF_i(t_i))^2}\frac{\partial F_i(t_i)}{\partial\lambda_i}. \tag{2.14}$$

The $k\times k$ expected information matrix $\boldsymbol{D}_n$ has the form

$$\boldsymbol{D}_n = \mathbb{E}(\boldsymbol{F}_n(\theta_0)) = \sum_{i=1}^{n}\boldsymbol{X}_i\mathcal{D}_i\boldsymbol{X}_i^T \tag{2.15}$$

(recall that $\theta_0 = (\beta_0^T\gamma_0^T)^T$ denotes the true parameter) where

$$\mathcal{D}_i = \begin{bmatrix} d_i^{11} & d_i^{12} \\ d_i^{12} & d_i^{22}\end{bmatrix} \tag{2.16}$$

are 2×2 symmetric matrices with elements $d_i^{rs} = \mathbb{E}(f_i^{rs}(\theta_0))$, $r,s=1,2$. Formulae for d_i^{rs}, showing in particular that they are finite (provided $\mathbb{E}(1/\bar{F}_{i0}(u_i)) < \infty$ for the boundary case, see (F5) below) are given in Lemma 5.1 below. Thus the expectations in (2.16) will be finite under our assumptions. General likelihood theory (Cox and Hinkley (1974), pp. 107–108) suggests that

$$\mathbb{E}(S_n(\theta_0)) = 0 \quad\text{and}\quad \mathbb{E}(S_n(\theta_0)S_n^T(\theta_0)) = \mathbb{E}(\boldsymbol{F}_n(\theta_0)) = \boldsymbol{D}_n. \tag{2.17}$$

These hold (see Lemma 5.1) under assumptions which we now list.

Our assumptions place restrictions on the covariates and on the relation between the censoring and survival distributions. For $A>0$, define

$$N_n(A) = \{\theta\in\Theta : (\theta-\theta_0)^T\boldsymbol{D}_n(\theta-\theta_0)\le A^2\}. \tag{2.18}$$

Assumptions F1–F4.

(F1) $\eta(\kappa)\in\Lambda$ for all $\kappa\in\mathbb{R}$, η has continuous third derivatives in $\mathbb{R}$, and its first derivative is non-zero in $\mathbb{R}$.

(F2) For each $i\ge 1$ and $\lambda\in\Lambda$, $\mathbb{P}(u_i\ge\tilde{\tau}_\lambda>\tau_-)>0$.

(F3) The matrices $\sum_{i=1}^n x_ix_i^T$ and $\sum_{i=1}^n y_iy_i^T$ are positive definite for some $n\ge k$.

For the remainder of this section, conditions (F1)–(F3) will be in force. Lemma 5.2 shows that $\boldsymbol{D}_n$ is then invertible for large n.

(F4) $\lim_{n\to\infty}\sum_{i=1}^n(\mathcal{U}_i\,\mathrm{tr}(\boldsymbol{X}_i^T\boldsymbol{D}_n^{-1}\boldsymbol{X}_i))^{3/2}=0$ where $\mathcal{U}_i = M_i^{2/3}\Delta_i$, $i\ge1$, and M_i and Δ_i are defined as follows.

Let $\delta > 0$ be fixed. For each $i \geq 1$, $\Delta_i^{3/2}$ is the maximum value over any of the quantities

$$\begin{aligned}(2.19)\qquad &\{1, |\eta'''(\kappa[1])g'(\lambda[2])|, |\eta'''(\kappa[1])|h^{1/2}(\lambda[1]),\\ &|\eta''(\kappa[1])|^{3/2}h^{3/8}(\lambda[2]), |\eta'(\kappa[1])|^3 h^{3/4}(\lambda[2])|\},\end{aligned}$$

where $h(\lambda) = g^{(iv)}(\lambda)+3(g''(\lambda))^2$, $\kappa[j]$, $j = 1, 2$, vary over the intersection between $(\kappa_{i0} - \delta, \kappa_{i0} + \delta)$ and the parameter space of $\kappa[j]$, and $\lambda[j] = \eta(\kappa[j])$.

We define M_i separately for the interior and boundary cases.

The interior case: Let A be any fixed constant. Define

$$M_i = 1 \vee \mathbb{E}\left(\sup_{\theta \in N_n(A)} \frac{1 - p_{i0}F_{i0}(u_i)}{(1 - \zeta(\rho)F_\lambda(u_i))^{3/4}}\right).$$

More precisely, we should replace $\mathcal{U}_i$ and M_i by $\mathcal{U}_{in}(A)$ and $M_{in}(A)$ and (F4) is required to hold for each $A > 0$. But we continue to denote $\mathcal{U}_{in}(A)$ and $M_{in}(A)$ as $\mathcal{U}_i$ and M_i for convenience.

The boundary case: We assume further:

ASSUMPTION F5.

(F5) Let $\eta_{i0}(\delta) = \sup_{\kappa \in [\kappa_{i0} - \delta, \kappa_{i0} + \delta]} \eta(\kappa)$. There is a $\delta > 0$ such that for each $i \geq 1$

$$M_i = \sup_{\lambda \in \eta_{i0}(\delta)} \mathbb{E}\left(\frac{1}{\bar{F}_\lambda^2(u_i)}\right) < \infty.$$

Our results will be proved under (F1)–(F4) and in addition under (F5) for the boundary case. Condition (F1) is very mild and generally satisfied in practice (see the examples in Section 3 below). The condition $\mathbb{P}(u_i \geq \tilde{\tau}_\lambda > \tau_-) > 0$ of (F2) ensures that *uncensored* observations will be observed with positive probability, and represents a minimal requirement that "followup" be sufficient in the sample. Condition (F3) simply ensures that the covariates do not degenerate to a lower dimensional subspace for large enough samples. Condition (F4) is a "uniform asymptotic negligibility" type of requirement on the covariates which also incorporates some interplay between censoring and survival distributions. Condition (F5) implies $\mathbb{P}\{u_i \leq \tau_+\} = 1$, which is natural in the boundary case since no $u_i > \tau_+$ can be observed when $p_0 = 1$. Conditions (F1)–(F5), as is also discussed in the next section, reflect the relationships between the failure and censoring distributions under which consistent estimation is possible.

3. Main results and examples

We now state the major results of the paper. We say that a sequence of events $\{A_n\}$ occurs with probability approaching 1 (WPA1) if $\mathbb{P}\{A_n\} \to 1$ as $n \to \infty$.

THEOREM 3.1. *If conditions* (F1)–(F4) *for the interior case or conditions* (F1)–(F5) *for the boundary case are satisfied, then an MLE $\hat{\theta}_n$ of θ_0 exists, is locally unique WPA*1, *and is consistent in probability for θ_0.*

From now on, $\hat{\theta}_n$ will denote the estimator obtained in Theorem 3.1. We do not claim that $\hat{\theta}_n$ is uniquely defined on Θ, even WPA1, though it is uniquely defined WPA1 on the neighborhood $N_n(A)$ of θ_0 for each $A > 0$.

In the next theorem, we consider the *interior case*, i.e. the θ_{j0} are interior points of Θ_j, $1 \le j \le k$. Let $1 \le j_1 < j_2 < \cdots < j_m \le k$. We wish to test the hypothesis

$$H_0 : \theta_j = \theta_{j0}, \quad j = j_i, \quad 1 \le i \le m, \ m < k,$$

against an unrestricted alternative. Let $l_n(\theta)$ be the likelihood function and $\mathcal{L}_n(\theta) = \log(l_n(\theta))$. The likelihood ratio (LR) test statistic for H_0 is defined by $L_n = l_n(\tilde{\theta}_n)/l_n(\hat{\theta}_n)$ where $\tilde{\theta}_n$ is a local maximum of $\mathcal{L}_n(\theta)$ under H_0. Now define the "deviance" of the restricted model from the unrestricted model by

$$d_n = -2\log L_n = 2[\mathcal{L}_n(\hat{\theta}_n) - \mathcal{L}_n(\tilde{\theta}_n)]. \tag{3.1}$$

Small values of L_n or large values of d_n indicate that H_0 is unlikely to be true. Denote by χ^2_ν a chi-square random variable with ν degrees of freedom. Let $N(0, \boldsymbol{I}_k)$ be a normal random vector with mean 0 and identity covariance matrix $\boldsymbol{I}_k$, and let $\boldsymbol{D}_n^{1/2}$ and $\boldsymbol{D}_n^{T/2}$ be any left and right square roots of $\boldsymbol{D}_n$, i.e., any square matrices such that $\boldsymbol{D}_n^{1/2}\boldsymbol{D}_n^{T/2} = \boldsymbol{D}_n$.

THEOREM 3.2. (*Interior case*) *If conditions* (F1)–(F4) *are satisfied, then for every $y \ge 0$, as $n \to \infty$,*

$$\mathbb{P}\{d_n \le y\} \to \mathbb{P}\{\chi^2_{k-m} \le y\}, \tag{3.2}$$

and

$$\boldsymbol{D}_n^{T/2}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \boldsymbol{I}_k). \tag{3.3}$$

For the rest of this section, consider the *boundary case*. Recall that $k_2 = 1$ and $\Theta_k = (a_k, 1]$ in this case. We wish to test the hypothesis $H_0 : \gamma_0 = 1$ which corresponds to $p_{i0} = 1$, $1 \le i \le n$, against an unrestricted alternative. Thus no individuals are immune to failure under the null hypothesis H_0. Let $\dot{\theta}_n = (\dot{\beta}_n^T\ 1)^T$ be a local maximum of $\mathcal{L}_n(\theta)$ under H_0. Again define the "deviance" of the restricted model from the unrestricted model by

$$d_n = 2[\mathcal{L}_n(\hat{\theta}_n) - \mathcal{L}_n(\dot{\theta}_n)]. \tag{3.4}$$

Partition the expected information matrix $\boldsymbol{D}_n$ as

$$\boldsymbol{D}_n = \begin{bmatrix} \bar{\boldsymbol{D}}_n & g_n \\ g_n^T & c_n \end{bmatrix} \tag{3.5}$$

where $\bar{\boldsymbol{D}}_n$ is $k_1 \times k_1$ and g_n is $k_1 \times 1$. Denote by N a standard normal random variable. Also let $X \sim N(0, \boldsymbol{I}_{k_1})$, $Y \sim N(0,1)$ and $Z = (X^T\ Y)^T \sim N(0, \boldsymbol{I}_k)$. Our next theorem shows that d_n has a non-standard asymptotic distribution.

THEOREM 3.3. (*Boundary case*) *If conditions* (F1)–(F5) *are satisfied, then for every $x < 0$ and $y \geq 0$, as $n \to \infty$,*

$$\mathbb{P}\{d_n \leq y\} \to \frac{1}{2} + \frac{1}{2}\mathbb{P}\{\chi_1^2 \leq y\}, \tag{3.6}$$

$$\mathbb{P}\left\{\sqrt{c_n - g_n^T \bar{\boldsymbol{D}}_n^{-1} g_n}(\hat{p}_n - 1) \leq x\right\} \to \mathbb{P}\{Y \leq x\}, \tag{3.7}$$

and

$$\mathbb{P}\{\hat{p}_n < 1\} \to 1/2. \tag{3.8}$$

Also if $\boldsymbol{D}_n^{T/2}$ is the right Cholesky square root of $\boldsymbol{D}_n$, then for any Borel set $B \subseteq \mathbb{R}^k$, as $n \to \infty$,

$$\mathbb{P}\{\boldsymbol{D}_n^{T/2}(\hat{\theta}_n - \theta_0) \in B, \hat{p}_n < 1\} \to \mathbb{P}\{Z \in B, Y \leq 0\}. \tag{3.9}$$

Theorem 3.3 generalises the results of Zhou and Maller (1995) which deal with the exponential distribution in the case where there is no covariate information. For an example of the use of the "50-50" chi-squared distribution in (3.6), see Maller and Zhou ((1996), Chapter 5) and Zhou and Maller (1995). Ghitany *et al.* (1994) discuss how to use results like Theorems 3.1 and 3.2 to analyse exponentially distributed data classified into groups.

To be able to use the distributional results in (3.3), (3.7) and (3.9) in practice, we must be able to "Studentise" them: replace $\boldsymbol{D}_n^{T/2}$ by a sample estimator. Theorem 2.2 of Vu *et al.* (1996) together with (6.1) below show that (3.3), (3.7) and (3.9) remain true under the specified conditions if $\boldsymbol{D}_n^{T/2}$ is replaced by $\boldsymbol{F}_n^{T/2}(\hat{\theta}_n)$ and c_n and g_n are replaced by the corresponding components of $\boldsymbol{F}_n(\hat{\theta}_n)$. Vu *et al.* (1996) discuss the use of the Cholesky square root in this context.

Examples. Many commonly used survival distributions are special cases of our model and conditions (F1)–(F5) are easy to interpret in specific cases. We restrict ourselves here to a discussion of the geometric distribution. In this case

$$f_\lambda(t) = (1-\theta)\theta^{t-1} = e^{-\lambda t + g(\lambda)}, \qquad t = 1, 2, \ldots,\ 0 < \theta < 1. \tag{3.10}$$

Here $\lambda = -\log(\theta) > 0$ and $g(\lambda) = \log(e^\lambda - 1)$. Clearly (3.10) is of the form (2.1) with $\boldsymbol{T} = [1, 2, \ldots)$, μ as counting measure on T, $q(t) = t$, $1 = \tau_- < \tau_+ = \infty$ and

$\Lambda = (0, \infty)$. We consider the estimation of β via a loglinear link to the parameter λ, i.e., $\lambda_i = e^{\beta^T x_i}$. Condition (F1) is easily seen to be satisfied with $\eta(\kappa) = e^\kappa$. (F2) reduces here to the minimal requirement that not all the mass of the u_i is placed to the left of 2, which we assume throughout. Assume also (F3); this can always be satisfied by eliminating redundant (linearly dependent) covariates, if necessary.

For the *interior case*, a sufficient condition for (F4) is given in Lemma 5.5 below as

$$\lim_{n\to\infty} \sum_{i=1}^{n} (\Delta_i' \operatorname{tr}\{\boldsymbol{X}_i^T \boldsymbol{D}_n^{-1} \boldsymbol{X}_i\})^{3/2} = 0 \tag{3.11}$$

where Δ_i' is defined by

$$\Delta_i' = \max_{\kappa \in [\kappa_{i0} - \delta, \kappa_{i0} + \delta]} \{\Delta_i, [\eta'(\kappa) g'(\lambda)]^2, [\eta'(\kappa) g'(4\lambda_{i0} - 3\lambda)]^2\} \quad (\text{with} \quad \lambda = \eta(\kappa)). \tag{3.12}$$

The Δ_i' given by (3.12) are bounded for δ small enough. Thus, via (3.11), (F4) is implied in this case by

$$\lim_{n\to\infty} \sum_{i=1}^{n} (\operatorname{tr}(\boldsymbol{X}_i^T \boldsymbol{D}_n^{-1} \boldsymbol{X}_i))^{3/2} = 0. \tag{3.13}$$

Consequently, for the interior case, when (3.13) holds, Theorems 3.1–3.2 give the existence, consistency and asymptotic normality of the MLE, and an asymptotic chi-square distribution for the deviance. If in addition we assume that the u_i are i.i.d., not degenerate at 0, and

$$\sup_{i\geq 1} (|x_i| \vee |y_i|) < \infty, \tag{3.14}$$

then (3.13) and hence (F4) are implied by

$$\max_{1\leq i\leq n} \left(x_i^T \left(\sum_{j=1}^{n} x_j x_j^T \right)^{-1} x_i + y_i^T \left(\sum_{j=1}^{n} y_j y_j^T \right)^{-1} y_i \right) \to 0 \quad (n \to \infty). \tag{3.15}$$

This follows from Lemmas 5.3–5.4 below. (3.15) is an asymptotic negligibility condition on the covariates which is possibly minimal in this context: see the discussion in Maller (1993). Conditions (3.14) and (3.15) hold if, for example, the x_i are indicators of a finite number of groups or classes to which individuals may belong.

Next take the *boundary case*. Assume conditions (F1)–(F3), as discussed above. Since the Δ_i in (F4) are bounded for δ small enough, (F4) is implied by

$$\lim_{n\to\infty} \sum_{i=1}^{n} M_i (\operatorname{tr}(\boldsymbol{X}_i^T \boldsymbol{D}_n^{-1} \boldsymbol{X}_i))^{3/2} = 0 \tag{3.16}$$

provided $M_i < \infty$, $1 \le i \le n$. Thus the conclusions of Theorem 3.3 hold for the geometric model if not all the mass of the u_i is to the left of 2 and (F3) and (3.16) hold. Now condition $M_i < \infty$ in (F5) can be written as

$$\int_{[0,\tau_+]} \frac{G_i(du)}{\bar{F}^2_{\eta_{i0}(\delta)}(u)} < \infty, \tag{3.17}$$

where G_i is the c.d.f. of the censoring random variable u_i. (3.17) tells us that the tail of G_i must not be too *heavy* relative to that of $F_{\eta_{i0}(\delta)}(t)$. In other words, the censoring must not be too *light* relative to the failure times.

Suppose for example that for each i, $G_i(u)$ is the (very light-tailed) uniform distribution on $[0, c]$. Condition (F2) simply corresponds to $c > 1$. We have $F_\lambda(t) = 1 - e^{-\lambda([t]-1)}$, essentially the discrete exponential distribution with parameterisation $\lambda_i = e^{\beta^T x_i}$. Now $\eta_{i0}(\delta) = e^{\beta_0^T x_i + \delta} = b\lambda_{i0}$ with $b = e^\delta > 1$, and the integral in (3.17) is no larger than $[e^{2bc\lambda_{i0}} - 1]/[2bc\lambda_{i0}] \le e^{2bc\lambda_{i0}}$. In other words, (F5) always holds for these G_i and F_{λ_i}. If in addition

$$\sup_{i \ge 1} |x_i| < \infty, \tag{3.18}$$

then by Lemmas 5.3 5.4 below (F4) is implied by the uniform asymptotic negligibility condition

$$\max_{1 \le i \le n} x_i^T \left(\sum_{j=1}^n x_j x_j^T \right)^{-1} x_i \to 0 \qquad (n \to \infty). \tag{3.19}$$

Thus Theorems 3.1 and 3.2 hold under (3.18)–(3.19) in this case. When (3.18) holds, (3.19) is equivalent to requiring the minimum eigenvalue of $\sum_{j=1}^n x_j x_j^T$ to tend to ∞ as $n \to \infty$ (and similarly for (3.15)). Very similar conditions to (3.18)–(3.19) occur in Fahrmeir and Kaufmann's treatment of glms.

For a second illustration, suppose that G_i is exponential with parameter μ_i. Then (3.17) holds if $\mu_i > 2b\lambda_{i0}$ for some $b > 1$. Since $\mathbb{E}(u_i) = 1/\mu_i$ and $\mathbb{E}(\tilde{t}_{i0}) = 1/\lambda_{i0}$, then, this says that censoring times must be on average at most one half as large as failure times. So again censoring must be not too "light". It may seem curious at first that a certain "heaviness" of censoring is insisted on in the boundary case. Condition (3.17) here reflects the fact that the boundary test only makes sense if there is a positive probability of having survival times larger than the largest possible censoring time; otherwise, there would be no point in testing H_0. Nevertheless, (3.17) is not optimal in this regard; the power of 2 may be reduced to 1 for the i.i.d. model without covariates and with exponential distribution F_i, as shown in Zhou and Maller (1995). In practice, the u_i will be bounded, in which case (3.17) holds for any F_λ with $\tau_+ = \infty$.

To conclude this section we remark that the Poisson and binomial distributions can also be handled within our framework. For these, also, $q(t) = t$. An example of (2.1) with $q(t)$ not the identity function is given by the Pareto distribution, with

$$f_\lambda(t) = \frac{(\theta - 1)a^{\theta-1}}{t^\theta} = e^{-\lambda \log t + \log((\lambda-1)a^{\lambda-1})}, \qquad t \ge a > 0, \ \theta > 1.$$

Here $\lambda = \theta$ and $q(t) = \log(t)$. Of course, this distribution can be transformed to the exponential distribution on $[\log a, \infty)$, but our results apply to an analysis of the data carried out on the original scale, as is natural.

4. Discussion

Theorems 3.1–3.3 provide the theoretical foundations for the use of some kinds of exponential family models in the analysis of censored data with immune individuals, similar to the way in which Fahrmeir and Kaufmann's (1985) results justify the use of large-sample normal approximations for the distributions of estimators, and the chi-squared distribution for the distribution of the deviance, in glms. While those results carry over precisely for our "interior" models, we note that modifications are required for the "boundary" models used in testing for the presence of immunes. All of our results are directly applicable in practice as a result of our justification of the large-sample "Studentised" results mentioned in Section 3.

For maximum utility, the results given here could be generalised in a number of directions. Not all survival models in common usage are in the exponential family, just as not all nonlinear models of importance in the analysis of "ordinary" data (with covariates) are glms. For example, the Fahrmeir and Kaufmann theory does not apply to give the consistency and asymptotic normality of the MLE for the shape parameter of the Weibull distribution, properties which are much relied on in practice (see, e.g., Aitkin *et al.* (1989), Collett (1994), Maller and Zhou (1996)). The Weibull and similarly the Gamma distributions can be handled within our framework, if the shape parameter in the Weibull or the scale parameter in the Gamma is "known", or takes some hypothesised value. There are realistic situations in which this is plausible, but an extension of our methods to cover the general cases would certainly be valuable too.

5. Basic properties of the model and conditions (F2)–(F4)

In this section, we state and sketch the proofs of some basic properties concerning the model, and also analyse the imposed conditions. In Lemma 5.1, $\tilde{t}_\lambda$ is a random variable with c.d.f. F_λ, and expectations are with respect to F_λ.

LEMMA 5.1. (i) *For each* $t > 0$ *and* $\lambda \in \Lambda$, *we have*

$$\partial f_\lambda(t)/\partial\lambda = (g'(\lambda) - q(t))f_\lambda(t); \tag{5.1}$$

$$\mathbb{E}(q(\tilde{t}_\lambda)) = g'(\lambda); \tag{5.2}$$

$$\mathrm{Var}(q(\tilde{t}_\lambda)) = -g''(\lambda); \tag{5.3}$$

$$\mathbb{E}((q(\tilde{t}_\lambda) - g'(\lambda))^3) = g'''(\lambda); \tag{5.4}$$

$$\mathbb{E}((q(\tilde{t}_\lambda) - g'(\lambda))^4) = -g^{(iv)}(\lambda) + 3(g''(\lambda))^2; \tag{5.5}$$

$$-|g''(\lambda)|^{1/2} \le \frac{\partial F_\lambda(t)}{\partial\lambda} = \int_{[0,t]} (g'(\lambda) - q(y))F_\lambda(dy) \le |g''(\lambda)|^{1/2}; \tag{5.6}$$

$$g''(\lambda) \le \frac{\partial^2 F_\lambda(t)}{\partial\lambda^2} = g''(\lambda)F_\lambda(t) + \int_{[0,t]} (g'(\lambda) - q(y))^2 F_\lambda(dy) \le -g''(\lambda). \tag{5.7}$$

(ii) *Suppose in addition that* (F1) *holds. For* $1 \le i \le n$, *we have*

$$d_i^{11} = p_{i0}(\eta'(\kappa_{i0}))^2 \mathbb{E}\left(\int_{[0,u_i]} (g'(\lambda_{i0}) - q(y))^2 F_{i0}(dy) + \frac{p_{i0}}{1 - p_{i0}F_{i0}(u_i)} \cdot \left(\int_{[0,u_i]} (g'(\lambda_{i0}) - q(y)) F_{i0}(dy) \right)^2 \right), \tag{5.8}$$

$$d_i^{22} = \mathbb{E}\left(\frac{(\zeta'(\rho_{i0}))^2 F_{i0}(u_i)}{p_{i0}(1 - p_{i0}F_{i0}(u_i))} \right), \tag{5.9}$$

and

$$d_i^{12} = \mathbb{E}\left(\frac{\eta'(\kappa_{i0})\zeta'(\rho_{i0})}{1 - p_{i0}F_{i0}(u_i)} \int_{[0,u_i]} (g'(\lambda_{i0}) - q(y)) F_{i0}(dy) \right). \tag{5.10}$$

(iii) *Assume in addition that* $\mathbb{E}(1/\bar{F}_{i0}(u_i)) < \infty$ *for the boundary case. Then* (2.17) *holds.*

PROOF. (5.1)–(5.7) are verified by routine calculus which we omit. The derivatives in (5.8)–(5.10) exist by (F1). Lemma 5.1 then follows easily from (2.12)–(2.14) and Lemma 2 in Ghitany *et al.* (1994). □

(2.17) shows that the positive definiteness of $\boldsymbol{D}_n$ is essential. This in turn depends on the positive definiteness of the $\mathcal{D}_i$ and the lack of linear dependencies among the covariates, investigated in Lemma 5.2.

LEMMA 5.2. (i) *Suppose that condition* (F1) *holds. For* $1 \le i \le n$, *the matrix* $\mathcal{D}_i$ *defined by* (2.16) *is positive definite if and only if* (F2) *holds.*

(ii) *Suppose* $\lambda_{\min}(\mathcal{D}_i) > 0$, $i \ge 1$. *Then* (F3) *holds if and only if*

$$\lambda_{\min}(\boldsymbol{D}_n) > 0 \quad \text{for some} \quad n \ge k, \tag{5.11}$$

or equivalently, if $\lambda_{\min}(\boldsymbol{D}_n) > 0$ *for all large enough* n.

PROOF. (i) For κ and ρ in $\mathbb{R}$, let $p_\rho = \zeta(\rho)$ and $\lambda_\kappa = \eta(\kappa)$, and define

$$h_{\kappa,\rho}^{11}(u_i) = p_\rho(\eta'(\kappa))^2 \left[\int_{[0,u_i]} (g'(\lambda_\kappa) - q(y))^2 F_{\lambda_\kappa}(dy) + \frac{p_\rho}{1 - p_\rho F_{\lambda_\kappa}(u_i)} \cdot \left(\int_{[0,u_i]} (g'(\lambda_\kappa) - q(y)) F_{\lambda_\kappa}(dy) \right)^2 \right], \tag{5.12}$$

$$h_{\kappa,\rho}^{22}(u_i) = \frac{(\zeta'(\rho))^2 F_{\lambda_\kappa}(u_i)}{p_\rho(1 - p_\rho F_{\lambda_\kappa}(u_i))}, \tag{5.13}$$

and

$$(5.14)\qquad h^{12}_{\kappa,\rho}(u_i) = \frac{\eta'(\kappa)\zeta'(\rho)}{1-p_\rho F_{\lambda_\kappa}(u_i)} \int_{[0,u_i]} (g'(\lambda_\kappa) - q(y)) F_{\lambda_\kappa}(dy).$$

Let $h^{rs}_{i0}(u_i) = h^{rs}_{\kappa_{i0},\rho_{i0}}(u_i)$ and note from (5.8)–(5.10) that $\mathbb{E}(h^{rs}_{i0}(u_i)) = d^{rs}_i$.

Suppose that (F2) holds, i.e., $\mathbb{P}(u_i > \tau_-) > 0$. Then $d^{22}_i > 0$ by (5.13). Since $\mathcal{D}_i = (d^{rs}_i)$ is symmetric, it thus suffices to show that $\det(\mathcal{D}_i) > 0$. Apply the inequality $\mathbb{E}(\sqrt{UV}) \le \sqrt{\mathbb{E}(U)\mathbb{E}(V)}$ with $U = h^{11}_{i0}(u_i)$ and $V = h^{22}_{i0}(u_i)$ to see that

$$(5.15)\qquad \begin{aligned}\det(\mathcal{D}_i) &= \mathbb{E}(h^{11}_{i0})\mathbb{E}(h^{22}_{i0}) - (\mathbb{E}(h^{12}_{i0}))^2 \\ &\ge \left(\mathbb{E}\left(\sqrt{h^{11}_{i0}(u_i)h^{22}_{i0}(u_i)}\right)\right)^2 - (\mathbb{E}(h^{12}_{i0}(u_i)))^2.\end{aligned}$$

Therefore, it suffices to show that for κ and ρ in $\mathbb{R}$

$$(5.16)\qquad \mathbb{E}\left(\sqrt{h^{11}_{\kappa,\rho}(u_i)h^{22}_{\kappa,\rho}(u_i)} \pm h^{12}_{\kappa,\rho}(u_i)\right) > 0.$$

For $u > 0$, define functions

$$(5.17)\qquad h_{\kappa,\rho}(u) = h^{11}_{\kappa,\rho}(u)h^{22}_{\kappa,\rho}(u) - (h^{12}_{\kappa,\rho}(u))^2.$$

After some algebra, we can write, for $u > 0$,

$$(5.18)\qquad \begin{aligned}&\frac{h_{\kappa,\rho}(u)}{(\eta'(\kappa))^2(\zeta'(\rho))^2} \\ &\quad = \frac{F^2_{\lambda_\kappa}(u)}{1-p_\rho F_{\lambda_\kappa}(u)} \int_{[0,u]} \left(q(y) - \int_{[0,u]} q(z)\frac{F_{\lambda_\kappa}(dz)}{F_{\lambda_\kappa}(u)}\right)^2 \frac{F_{\lambda_\kappa}(dy)}{F_{\lambda_\kappa}(u)}\end{aligned}$$

where the integral is the conditional variance of $q(\tilde{t}_{\lambda_\kappa})$, given $\tilde{t}_{\lambda_\kappa} \le u$.

For any $u > \tau_-$, $F_\lambda(u) > 0$ and $\tilde{t}_\lambda$ is not constant a.s. on $[0, u]$ by the condition (F2). Therefore, for each pair $(\kappa, \rho) \in \mathbb{R}^2$ and for any large enough $u > \tau_-$,

$$(5.19)\qquad \frac{h_{\kappa,\rho}(u)}{(\eta'(\kappa))^2(\zeta'(\rho))^2} = \frac{F^2_{\lambda_\kappa}(u)}{1-p_\rho F_{\lambda_\kappa}(u)} \operatorname{Var}(q(\tilde{t}_{\lambda_\kappa}) \mid \tilde{t}_{\lambda_\kappa} \le u) > 0.$$

Obviously $h_{\kappa,\rho}(u) = 0$ when $u < \tau_-$. For $u = \tau_-$ we have

$$(5.20)\qquad \begin{aligned}\frac{h_{\kappa,\rho}(u)}{(\eta'(\kappa))^2(\zeta'(\rho))^2} &= \frac{F^2_{\lambda_\kappa}(\tau_-)(g'(\lambda_\kappa) - q(\tau_-))^2}{1-p_\rho F_{\lambda_\kappa}(\tau_-)} \\ &\quad - \frac{F^2_{\lambda_\kappa}(\tau_-)(g'(\lambda_\kappa) - q(\tau_-))^2}{1-p_\rho F_{\lambda_\kappa}(\tau_-)} = 0.\end{aligned}$$

So we have shown that for each pair $(\kappa, \rho) \in \mathbb{R}^2$ and any $u \ge 0$

$$(5.21)\qquad h_{\kappa,\rho}(u) \ge 0.$$

It follows from (5.17), (5.19) and (5.20) that

$$\mathcal{C}_{\kappa,\rho}(u_i) = \sqrt{h^{11}_{\kappa,\rho}(u_i)h^{22}_{\kappa,\rho}(u_i)} - h^{12}_{\kappa,\rho}(u_i) \geq 0 \quad \text{a.s.} \tag{5.22}$$

Now suppose $\mathbb{E}(\mathcal{C}_{\kappa,\rho}(u_i)) = 0$ for some $\kappa, \rho \in \mathbb{R}$. By (5.22), then, $\mathcal{C}_{\kappa,\rho}(u_i) = 0$ a.s. Therefore by (5.17), $h_{\kappa,\rho}(u_i) = 0$ a.s. which contradicts (5.19) by (F2). Thus $\mathbb{E}(\mathcal{C}_{\kappa,\rho}(u_i)) > 0$. Similarly, we have

$$\mathbb{E}\left(\sqrt{h^{11}_{\kappa,\rho}(u_i)h^{22}_{\kappa,\rho}(u_i)} + h^{12}_{\kappa,\rho}(u_i)\right) > 0. \tag{5.23}$$

Thus (5.16) holds, and $\det(\mathcal{D}_i) > 0$.

Conversely, suppose that (F2) fails, i.e., $\mathbb{P}(u_i \geq \tilde{\tau}_\lambda > \tau_-) = 0$. Since $h^{rs}_{i0}(u_i) = 0$ for $u_i < \tau_-$ (see (5.12)–(5.14)), $\det(\mathcal{D}_i) = h_{i0}(\tau_-)\mathbb{P}^2(u_i = \tau_-) = 0$ by (5.20).

(ii) Next suppose that $\lambda_{\min}(\mathcal{D}_i) > 0$, $i \geq 1$, and that (F3) holds. Define a matrix $\boldsymbol{C}_n$ by

$$\boldsymbol{C}_n = \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^T = \begin{bmatrix} \sum_{i=1}^n x_i x_i^T & 0 \\ 0 & \sum_{i=1}^n y_i y_i^T \end{bmatrix}. \tag{5.24}$$

Then (F3) is obviously equivalent to

$$\lambda_{\min}\{\boldsymbol{C}_{n_0}\} > 0 \quad \text{for some} \quad n_0 \geq k. \tag{5.25}$$

Let ξ be any k-dimensional unit vector. Suppose $\lambda_{\min}(\boldsymbol{D}_n) = 0$ for some $n \geq n_0$. Since

$$\xi^T \boldsymbol{D}_n \xi = \sum_{i=1}^n \xi^T \boldsymbol{X}_i \mathcal{D}_i \boldsymbol{X}_i^T \xi \geq \sum_{i=1}^n \lambda_{\min}(\mathcal{D}_i) \xi^T \boldsymbol{X}_i \boldsymbol{X}_i^T \xi$$

and since $\lambda_{\min}(\mathcal{D}_i) > 0$, for $i \geq 1$, this means $\xi^T \boldsymbol{X}_i \boldsymbol{X}_i^T \xi = 0$, for some $\xi = \xi(n)$ and $1 \leq i \leq n$. This contradicts (5.25) as $n_0 \leq n$. Hence (F3) in fact implies that (5.11) holds for all values of $n \geq n_0$, i.e., for all values of n large enough.

Conversely, suppose that for some $n \geq k$, $\lambda_{\min}(\boldsymbol{D}_n) > 0$, so that the matrix

$$\begin{bmatrix} x_1 & 0 & \cdots & x_n & 0 \\ 0 & y_1 & \cdots & 0 & y_n \end{bmatrix}_{k \times 2n}$$

is of full rank k. Then $\sum_{i=1}^n x_i x_i^T$ and $\sum_{i=1}^n y_i y_i^T$ are invertible, proving (F3). □

Now we come to consider (F4).

LEMMA 5.3. *Suppose that* (F3) *holds, that there is a positive constant* c_0 *such that* $\inf_{i\geq 1} \lambda_{\min}(\mathcal{D}_i) > c_0$, *and that*

$$\max_{1\leq i\leq n} \mathcal{U}_i^{3/2} \left(x_i^T \left(\sum_{j=1}^n x_j x_j^T \right)^{-1} x_i + y_i^T \left(\sum_{j=1}^n y_j y_j^T \right)^{-1} y_i \right)^{1/2} \to 0 \quad (n \to \infty). \tag{5.26}$$

Then (F4) *holds.*

PROOF. Suppose (F3) holds and $\lambda_{\min}(\mathcal{D}_i) \geq c_0$ for all $i \geq 1$. Then each $\mathcal{D}_i$ is invertible and $\boldsymbol{D}_n$ is invertible for n large enough by Lemma 5.2. Let $\boldsymbol{D}_n^{-1/2}$ be the symmetric positive definite square root of $\boldsymbol{D}_n^{-1}$, and similarly for $\mathcal{D}_i^{-1/2}$. Then

$$\begin{aligned}
(5.27)\qquad \sum_{i=1}^{n} \operatorname{tr}(\boldsymbol{X}_i^T \boldsymbol{D}_n^{-1} \boldsymbol{X}_i) &= \sum_{i=1}^{n} \operatorname{tr}(\boldsymbol{D}_n^{-1/2} \boldsymbol{X}_i \mathcal{D}_i^{1/2} \mathcal{D}_i^{-1} \mathcal{D}_i^{1/2} \boldsymbol{X}_i^T \boldsymbol{D}_n^{-1/2}) \\
&\leq \sum_{i=1}^{n} \frac{k}{\lambda_{\min}(\mathcal{D}_i)} \lambda_{\max}(\boldsymbol{D}_n^{-1/2} \boldsymbol{X}_i \mathcal{D}_i \boldsymbol{X}_i^T \boldsymbol{D}_n^{-1/2}) \\
&\leq \frac{k}{c_0} \sum_{i=1}^{n} \operatorname{tr}(\boldsymbol{D}_n^{-1/2} \boldsymbol{X}_i \mathcal{D}_i \boldsymbol{X}_i^T \boldsymbol{D}_n^{-1/2}) \\
&= \frac{k}{c_0} \operatorname{tr}(\boldsymbol{D}_n^{-1/2} \boldsymbol{D}_n \boldsymbol{D}_n^{-1/2}) = \frac{k^2}{c_0}.
\end{aligned}$$

Suppose in addition that (5.26) is satisfied. Let $\boldsymbol{C}_n$ be the matrix defined in (5.24). By (F3), this is invertible for all n large enough. Thus

$$\begin{aligned}
(5.28)\qquad \operatorname{tr}(\boldsymbol{X}_i^T \boldsymbol{D}_n^{-1} \boldsymbol{X}_i) &\leq 2\lambda_{\max}(\boldsymbol{X}_i^T \boldsymbol{C}_n^{-1/2} [\boldsymbol{C}_n^{1/2} \boldsymbol{D}_n^{-1} \boldsymbol{C}_n^{1/2}] \boldsymbol{C}_n^{-1/2} \boldsymbol{X}_i) \\
&\leq 2 \operatorname{tr}(\boldsymbol{C}_n^{1/2} \boldsymbol{D}_n^{-1} \boldsymbol{C}_n^{1/2}) \lambda_{\max}(\boldsymbol{X}_i^T \boldsymbol{C}_n^{-1} \boldsymbol{X}_i) \\
&= 2 \operatorname{tr}(\boldsymbol{D}_n^{-1/2} \boldsymbol{C}_n \boldsymbol{D}_n^{-1/2}) \\
&\quad \cdot \left(x_i^T \left(\sum_{j=1}^{n} x_j x_j^T \right)^{-1} x_i + y_i^T \left(\sum_{j=1}^{n} y_j y_j^T \right)^{-1} y_i \right).
\end{aligned}$$

Now it follows from (5.27) that

$$\begin{aligned}
\operatorname{tr}(\boldsymbol{D}_n^{-1/2} \boldsymbol{C}_n \boldsymbol{D}_n^{-1/2}) &= \operatorname{tr} \left(\sum_{i=1}^{n} \boldsymbol{D}_n^{-1/2} \boldsymbol{X}_i \boldsymbol{X}_i^T \boldsymbol{D}_n^{-1/2} \right) \\
&= \sum_{i=1}^{n} \operatorname{tr}(\boldsymbol{X}_i^T \boldsymbol{D}_n^{-1} \boldsymbol{X}_i) \leq k^2/c_0.
\end{aligned}$$

So by (5.26)–(5.28), we conclude that as $n \to \infty$,

$$\begin{aligned}
&\sum_{i=1}^{n} (\mathcal{U}_i \operatorname{tr}(\boldsymbol{X}_i^T \boldsymbol{D}_n^{-1} \boldsymbol{X}_i))^{3/2} \\
&\qquad \leq \max_{1 \leq i \leq n} \mathcal{U}_i^{3/2} \sqrt{\operatorname{tr}(\boldsymbol{X}_i^T \boldsymbol{D}_n^{-1} \boldsymbol{X}_i)} \frac{k^2}{c_0} \\
&\qquad \leq \sqrt{2} k^3 \max_{1 \leq i \leq n} \mathcal{U}_i^{3/2} \\
&\qquad\quad \cdot \left(x_i^T \left(\sum_{j=1}^{n} x_j x_j^T \right)^{-1} x_i + y_i^T \left(\sum_{j=1}^{n} y_j y_j^T \right)^{-1} y_i \right)^{1/2} \Bigg/ c_0 \to 0. \qquad \square
\end{aligned}$$

Next we show that $\inf_{i\geq 1}\lambda_{\min}(\mathcal{D}_i)>0$ for bounded covariates under (F1) and (F2), when the u_i are i.i.d.

LEMMA 5.4. *Assume that* (F1), (F2) *and* (3.14) *hold. Also assume for the boundary case that* $\sup_{i\geq 1}\mathbb{E}(1/\bar{F}_{i0}(u_i))<\infty$. *Then for some positive constant* l_1

$$\sup_{i\geq 1}(d_i^{11}+d_i^{22})\leq l_1. \tag{5.29}$$

Furthermore, if the u_i *are i.i.d., then for some positive constant* l_2,

$$l_2\leq \inf_{i\geq 1}\det(\mathcal{D}_i), \qquad \text{and} \qquad \inf_{i\geq 1}\lambda_{\min}(\mathcal{D}_i)>0. \tag{5.30}$$

PROOF. Let (F1) and (F2) hold and define

$$E_1=\Big[\inf_{i\geq 1}\kappa_{i0},\sup_{i\geq 1}\kappa_{i0}\Big],\qquad E_2=\Big[\inf_{i\geq 1}\rho_{i0},\sup_{i\geq 1}\rho_{i0}\Big]\qquad\text{and}\qquad E=E_1\times E_2.$$

Assume also (3.14). Then E_1 and E_2 are finite intervals containing κ_{i0} and ρ_{i0} for each $i\geq 1$. Since ζ is continuous on the compact set E_2, there exists $\bar{\rho}$ so that $\zeta(\bar{\rho})=\max_{\rho\in E_2}\zeta(\rho)$. Let $\bar{p}=\zeta(\bar{\rho})$ and, as before, $p_\rho=\zeta(\rho)$ and $\lambda_\kappa=\eta(\kappa)$.

For the interior case, we have $\sup_{\rho\in E_2}p_\rho=\bar{p}<1$ for $u\geq 0$. Thus it follows from (5.12), (5.2), (5.3) and (5.6) that

$$\sup_{(\kappa,\rho)\in E}h^{11}_{\kappa,\rho}(u)\leq \sup_{(\kappa,\rho)\in E}(\eta'(\kappa))^2(-g''(\lambda))\left(1+\frac{p_\rho}{1-p_\rho}\right)=\sup_{\kappa\in E_1}\frac{(\eta'(\kappa))^2(-g''(\lambda))}{1-\bar{p}}$$

which is finite since g' and g'' are continuous on the finite interval E_1. Similarly it follows from (5.13) that for all $u>0$

$$\sup_{(\kappa,\rho)\in E}h^{22}_{\kappa,\rho}(u)\leq \frac{\sup_{\rho\in E_2}(\zeta'(\rho))^2}{\inf_{\rho\in E_2}p_\rho(1-\bar{p})}<\infty. \tag{5.31}$$

Since $d_i^{rs}=\mathbb{E}(h^{rs}_{\kappa_{i0},\rho_{i0}}(u_i))$, these imply (5.29).

Now consider the boundary case. Assume further that $\sup_{i\geq 1}\mathbb{E}(1/\bar{F}_{i0}(u_i))<\infty$. Then (5.29) holds by (5.12)–(5.13).

Now assume in addition that the u_i are i.i.d. We show that

$$\inf_{(\kappa,\rho)\in E}\mathbb{E}(h_{\kappa,\rho}(u_1))>0. \tag{5.32}$$

Suppose to the contrary that there exist $(\kappa_i,\rho_i)\in E$ such that $\mathbb{E}(h_{\kappa_i,\rho_i}(u_1))\to 0$. We can further assume without loss of generality that $\kappa_i\to\kappa_0$, $\rho_i\to\rho_0$, and $(\kappa_0,\rho_0)\in E$ since E_1 and E_2 are closed. Let $\lambda_0=\eta(\kappa_0)$. Of course $\lambda_0\in\Lambda$ if $\Lambda=\mathbb{R}$. If $\Lambda=(a,\infty)$, we have by (F1) that $\eta(\kappa)>a$ since E_1 is bounded. Thus $\lambda_0=\eta(\kappa_0)\in\Lambda$. Since the $h^{rs}_{\kappa,\rho}(u)$ are continuous in (κ,ρ) for each $u>0$ (see (5.12)–(5.14)) we have $h_{\kappa_i,\rho_i}(u)\to h_{\kappa_0,\rho_0}(u)$ for each $u>0$. Then by Fatou's lemma $\mathbb{E}(h_{\kappa_0,\rho_0}(u_1)) = 0$ (recall that $h_{\kappa,\rho}(u) \geq 0$ for all $\kappa, \rho, u \geq 0$) so $h_{\kappa_0,\rho_0}(u_1) = 0$ a.s. By (5.19), then, $\operatorname{Var}(q(\tilde{t}_{\lambda_0}) \mid \tilde{t}_{\lambda_0} \leq u_1) = 0$ a.s. Thus $q(\tilde{t}_{\lambda_0})$ is constant on $[0, u_1]$. But under (F2), i.e., $\mathbb{P}(u_1 \geq \tilde{\tau}_\lambda > \tau_-) > 0$, this is impossible. Hence (5.32) holds and then the first inequality in (5.30) follows from the argument used in (5.15)–(5.16).

Finally, we have $\inf_{i\geq 1} \lambda_{\min}(\mathcal{D}_i) \geq l_2/l_1 > 0$, since for $i \geq 1$

$$\frac{1}{\lambda_{\min}(\mathcal{D}_i)} = \lambda_{\max}(\mathcal{D}_i^{-1}) \leq \operatorname{tr}(\mathcal{D}_i^{-1}) = \frac{d_i^{11} + d_i^{22}}{d_i^{11}d_i^{22} - (d_i^{12})^2} = \frac{\operatorname{tr}(\mathcal{D}_i)}{\det(\mathcal{D}_i)} \leq \frac{l_1}{l_2}. \qquad \square$$

The next lemma gives sufficient and necessary conditions for (F4).

LEMMA 5.5. (i) *Let conditions* (F1)–(F3) *be satisfied. Then for each* $A > 0$ *and* n *large enough we have*

$$\sup_{\theta \in N_n(A)} |\boldsymbol{X}_i^T(\theta - \theta_0)|^2 \leq A^2 \operatorname{tr}(\boldsymbol{X}_i^T \boldsymbol{D}_n^{-1} \boldsymbol{X}_i), \tag{5.33}$$

and for some constant $B > 0$ *not depending on* A *or* n,

$$\sup_{\theta \in N_n(A)} |\theta - \theta_0|^2 \leq BA^2 \max_{1\leq i\leq n} \operatorname{tr}(\boldsymbol{X}_i^T \boldsymbol{D}_n^{-1} \boldsymbol{X}_i). \tag{5.34}$$

Furthermore, $\lambda_{\min}(\boldsymbol{D}_n) \to \infty$ *as* $n \to \infty$ *if in addition* (F4) *holds.*

(ii) *If* $\Lambda = (0, \infty)$ *with* $\eta(\kappa) = e^\kappa$ *or* $\Lambda = \mathbb{R}$, *then* (F4) *is implied by* (3.11).

PROOF. (i) (5.33) is (4.14) of Ghitany *et al.* (1994), and (5.34) follows just as in the working after (4.24) of Ghitany *et al.* (1994). Furthermore, let v_n be a k-dimensional eigenvector of $\boldsymbol{D}_n$ associated with $\lambda_{\min}(\boldsymbol{D}_n)$, i.e. $\boldsymbol{D}_n v_n = \lambda_{\min}(\boldsymbol{D}_n) v_n$, and define the k-vector $\theta_n = \theta_0 + \sqrt{A^2/v_n^T \boldsymbol{D}_n v_n} v_n$. Since $\theta_n \in N_n(A)$, it follows from (5.34) that

$$A^2 = \lambda_{\min}(\boldsymbol{D}_n)|\theta_n - \theta_0|^2 \leq \lambda_{\min}(\boldsymbol{D}_n) BA^2 \max_{1\leq i\leq n} \operatorname{tr}(\boldsymbol{X}_i^T \boldsymbol{D}_n^{-1} \boldsymbol{X}_i).$$

Thus (F4) implies that $\lambda_{\min}(\boldsymbol{D}_n) \to \infty$ as $n \to \infty$.

(ii) Let $\lambda = \eta(\kappa)$ and $p_\rho = \zeta(\rho)$. Let $A > 0$ be fixed. It suffices to show that (3.11) implies that for all n large enough

$$\max_{1\leq i\leq n} \mathbb{E}\left(\sup_{\theta \in N_n(A)} \frac{1 - p_{i0} F_{i0}(u_i)}{(1 - p_\rho F_\lambda(u_i))^{3/4}} \right) < \infty. \tag{5.35}$$

For $\eta(\kappa) = e^\kappa = e^{\beta^T x_i}$, it follows from (5.33) that $4\lambda_{i0} - 3\lambda = e^{\kappa_{i0}}(4 - 3e^{(\kappa - \kappa_{i0})}) > 0$ for all n large enough and $\theta \in N_n(A)$. Hence, from the Hölder inequality,

$$\begin{aligned}
&\int_{(u_i,\infty)} a(y) \exp(-\lambda_{i0} q(y)) \mu(dy) \\
&\quad = \int_{(u_i,\infty)} a(y) \exp\left\{ -\left(\lambda_{i0} - \frac{3}{4}\lambda\right) q(y) \right\} \exp\left\{ -\frac{3}{4}\lambda q(y) \right\} \mu(dy) \\
&\quad \leq \left(\int_{(u_i,\infty)} a(y) \exp(-[4\lambda_{i0} - 3\lambda] q(y)) \mu(dy) \right)^{1/4} \\
&\qquad \cdot \left(\int_{(u_i,\infty)} a(y) \exp(-\lambda q(y)) \mu(dy) \right)^{3/4}
\end{aligned}$$

for θ in $N_n(A)$. Therefore, we have

$$\bar{F}_{i0}(u_i) \le \exp\left(\frac{3}{4}[g(\lambda_{i0}) - g(\lambda)] + \frac{1}{4}[g(\lambda_{i0}) - g(4\lambda_{i0} - 3\lambda)]\right)\bar{F}_\lambda^{3/4}(u_i).$$

By Taylor's expansion, we have

$$|g(\lambda_{i0}) - g(\lambda)| = |\kappa - \kappa_{i0}||\eta'(\tilde{\kappa}_i)|g'(\eta(\tilde{\kappa}_i)),$$

and

$$|g(\lambda_{i0}) - g(4\lambda_{i0} - 3\lambda)| = 3|\kappa - \kappa_{i0}||\eta'(\tilde{\kappa}_{i1})|g'(4\lambda_{i0} - 3\eta(\tilde{\kappa}_{i1}))$$

where $\tilde{\kappa}_i$ and $\tilde{\kappa}_{i1}$ are on the line segment between κ and κ_{i0}. So by (5.33) and (3.11), we have for all n large enough and $\theta \in N_n(A)$

$$\begin{aligned} &\bar{F}_{i0}(u_i) \le 2\bar{F}_\lambda^{3/4}(u_i) \quad \text{and} \\ &\frac{1-p_{i0}}{1-p_\rho} = \frac{1+e^{(\rho-\rho_{i0})}e^{\rho_{i0}}}{1+e^{\rho_{i0}}} \le \frac{1+2e^{\rho_{i0}}}{1+e^{\rho_{i0}}} \le 2. \end{aligned} \tag{5.36}$$

Therefore, (5.35) holds as (5.36) implies that for all n large enough and $\theta \in N_n(A)$

$$\frac{1-p_{i0}F_{i0}(u_i)}{(1-p_\rho F_\lambda(u_i))^{3/4}} = \frac{1-p_{i0}+p_{i0}\bar{F}_{i0}(u_i)}{(1-p_\rho F_\lambda(u_i))^{3/4}} \le \frac{(1-p_{i0})}{1-p_\rho} + \frac{\bar{F}_{i0}(u_i)}{\bar{F}_\lambda^{3/4}(u_i)} \le 4. \qquad \square$$

6. Proofs of Theorems 3.1–3.2 and (3.6) of Theorem 3.3

We apply the general results of Vu and Zhou (1997). In our case the $g(Y_i, \theta)$ of Vu and Zhou (1997) is the log-likelihood of the i-th observation with Y_i replaced by t_i. Obviously, conditions (A1)–(A3) of that paper are satisfied under (F1). The matrix $\boldsymbol{V}$ in (B4) of that paper is simply $\boldsymbol{I}_k$ in our case. Since $\lambda_{\min}(\boldsymbol{D}_n) \to \infty$ as $n \to \infty$ by Lemma 5.5, we can apply the results in that paper provided that the observed information matrix can be approximated by the expected information matrix in the sense that

$$\sup_{\theta^* \in [N_n(A)]^4} \|\boldsymbol{D}_n^{-1/2}\boldsymbol{F}_n^*(\theta^*)\boldsymbol{D}_n^{-T/2} - \boldsymbol{I}_k\|_1 \xrightarrow{P} 0 \quad \text{for each} \quad A > 0 \tag{6.1}$$

and that the score function is asymptotically normal in the sense that for any unit vector ξ_n, possibly depending on n, in $\mathbb{R}^k$,

$$\xi_n^T \boldsymbol{D}_n^{-1/2} S_n(\theta_0) \xrightarrow{D} N(0,1). \tag{6.2}$$

In (6.1), $\|\cdot\|_1$ denotes the sum of the absolute values of the elements of a matrix,

$$[N_n(A)]^4 = \{\theta^* = (\theta^{11}, \theta^{12}, \theta^{21}, \theta^{22}) : \theta^{rs} \in N_n(A), r, s = 1, 2\}, \tag{6.3}$$

and

$$\boldsymbol{F}_n^*(\theta^{11},\theta^{12},\theta^{21},\theta^{22}) = \sum_{i=1}^n \boldsymbol{X}_i \begin{bmatrix} f_i^{11}(\theta^{11}) & f_i^{12}(\theta^{12}) \\ f_i^{12}(\theta^{21}) & f_i^{22}(\theta^{22}) \end{bmatrix} \boldsymbol{X}_i^T. \tag{6.4}$$

Then Theorem 3.1 and the asymptotic distribution of d_n in Theorems 3.2–3.3 follow from Theorems 2.1 and 2.2 in Vu and Zhou (1997). For the asymptotic distribution of $\hat{\theta}_n$ given in (3.3), let u be a unit vector in $\mathbb{R}^k$. Recall that $\hat{\theta}_n$ maximises the log-likelihood on $N_n(A)$ WPA1. Thus $S_n(\hat{\theta}_n) = 0$ WPA1 as the log-likelihood is concave on $N_n(A)$ for all n large enough by (6.1). Hence there exists, by Taylor expansion, a $\bar{\theta}_n$ on the line segment between $\hat{\theta}_n$ and θ_0 such that

$$\begin{aligned} u^T\boldsymbol{D}_n^{-1/2}S_n(\theta_0) &= u^T\boldsymbol{D}_n^{-1/2}S_n(\theta_0) - u^T\boldsymbol{D}_n^{-1/2}S_n(\hat{\theta}_n) \\ &= u^T\boldsymbol{D}_n^{-1/2}\boldsymbol{F}_n(\bar{\theta}_n)(\hat{\theta}_n-\theta_0) + o_p(1) \\ &= u^T\{\boldsymbol{I}_k + \boldsymbol{D}_n^{-1/2}\{\boldsymbol{F}_n(\bar{\theta}) - \boldsymbol{D}_n\}\boldsymbol{D}_n^{-1/2}\}\boldsymbol{D}_n^{1/2}(\hat{\theta}_n-\theta_0) + o_p(1) \\ &= u^T\{\boldsymbol{I}_k + o_p(1)\}\boldsymbol{D}_n^{1/2}(\hat{\theta}_n-\theta_0) + o_p(1) \qquad \text{(by (6.1))}. \end{aligned}$$

Thus $\hat{\theta}_n$ is asymptotically normal by (6.2). It remains to prove (6.1)–(6.2) from a series of lemmas below. □

LEMMA 6.1. *Assume that* (F1)–(F4) *hold for the interior case or* (F1)–(F5) *hold for the boundary case. Then there exists a constant* K *such that for* $r,s=1,2$,

$$\mathbb{E}\left(\sup_{\theta\in N_n(A)} |f_i^{rs}(\theta) - f_i^{rs}(\theta_0)|\right) \le K\mathcal{U}_i^{3/2}\sqrt{\operatorname{tr}(\boldsymbol{X}_i^T\boldsymbol{D}_n^{-1}\boldsymbol{X}_i)} \tag{C1}$$

for all n *large enough*;

$$\mathbb{E}(|f_i^{rs}(\theta_0)|^{3/2}) \le K\mathcal{U}_i^{3/2}; \tag{C2}$$

and

$$\mathbb{E}(|s_i(\theta_0)|^3) \le K\mathcal{U}_i^{3/2}. \tag{C3}$$

PROOF. The details of the proof are fairly standard and are omitted here. □

Note that the constant K in (C1) may depend on A. Let $\boldsymbol{D}_n^{-1/2}\boldsymbol{X}_i = w_i = (w_{i1}\ w_{i2})$ where $w_{ir}\in\mathbb{R}^k$, $r=1,2$. Hence

$$\operatorname{tr}(\boldsymbol{X}_i^T\boldsymbol{D}_n^{-1}\boldsymbol{X}_i) = |w_{i1}|^2 + |w_{i2}|^2.$$

For any unit vector $u\in\mathbb{R}^k$ and $r,s=1,2$, let $a_{in}^{rs} = u^Tw_{ir}w_{is}^Tu$. Then for $r,s=1,2$,

$$|a_{in}^{rs}| = |u^Tw_{ir}w_{is}^Tu| \le |w_{ir}||w_{is}| \le \frac{|w_{ir}|^2+|w_{is}|^2}{2} \le \operatorname{tr}(\boldsymbol{X}_i^T\boldsymbol{D}_n^{-1}\boldsymbol{X}_i). \tag{6.5}$$

LEMMA 6.2. *If* (F4) *and* (C1)–(C2) *hold, then* (6.1) *holds.*

PROOF. Suppose that (F4) and (C1)–(C2) hold. We first prove that for $(r,s)=(1,2)$

$$\sum_{i=1}^{n} a_{in}^{rs}(f_i^{rs}(\theta_0)-d_i^{rs}) \xrightarrow{P} 0 \quad \text{as} \quad n\to\infty. \tag{6.6}$$

Let $\varepsilon>0$ be given. Define $\mathcal{A}_{in}=\{|a_{in}^{rs}f_i^{rs}(\theta_0)|>\varepsilon\}$ and $g_{in}^{rs}=1_{\mathcal{A}_{in}^c}f_i^{rs}(\theta_0)$. Then

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{n} a_{in}^{rs}(f_i^{rs}(\theta_0)-d_i^{rs})\right|>\varepsilon\right)\\
&\quad\le \mathbb{P}\left(\bigcup_{i=1}^{n}\mathcal{A}_{in}\right)\\
&\qquad+\mathbb{P}\left(\left|\sum_{i=1}^{n} a_{in}^{rs}(g_{in}^{rs}-\mathbb{E}(g_{in}^{rs}))\right|+\sum_{i=1}^{n}|a_{in}^{rs}|\mathbb{E}(1_{\mathcal{A}_{in}}|f_i^{rs}(\theta_0)|)>\varepsilon\right).
\end{aligned}\tag{6.7}$$

Since by the Markov inequality

$$\mathbb{P}\left(\bigcup_{i=1}^{n}\mathcal{A}_{in}\right)\le\sum_{i=1}^{n}\mathbb{P}(\mathcal{A}_{in})\le\frac{1}{\varepsilon^{3/2}}\sum_{i=1}^{n}|a_{in}^{rs}|^{3/2}\mathbb{E}(|f_i^{rs}(\theta_0)|^{3/2}),$$

while by the Chebychev inequality

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{n} a_{in}^{rs}(g_{in}^{rs}-\mathbb{E}(g_{in}^{rs}))\right|>\frac{\varepsilon}{2}\right)\\
&\quad\le\frac{4}{\varepsilon^2}\sum_{i=1}^{n}(a_{in}^{rs})^2\operatorname{Var}(g_{in}^{rs})\\
&\quad\le\frac{4}{\varepsilon^2}\sum_{i=1}^{n}(a_{in}^{rs})^2\mathbb{E}((g_{in}^{rs})^2)\le\frac{4}{\varepsilon^{3/2}}\sum_{i=1}^{n}|a_{in}^{rs}|^{3/2}\mathbb{E}(|f_i^{rs}(\theta_0)|^{3/2}),
\end{aligned}$$

and similarly

$$\sum_{i=1}^{n}|a_{in}^{rs}|\mathbb{E}(1_{\mathcal{A}_{in}}|f_i^{rs}(\theta_0)|)\le\frac{1}{\varepsilon^{1/2}}\sum_{i=1}^{n}|a_{in}^{rs}|^{3/2}\mathbb{E}(|f_i^{rs}(\theta_0)|^{3/2}),$$

(6.6) follows from (6.7), (6.5), (C2) and (F4).

Now we can prove (6.1). Let $\theta^*\in[N_n(A)]^4$ and write

$$\boldsymbol{D}_n^{-1/2}\boldsymbol{F}_n^*(\theta^*)\boldsymbol{D}_n^{-T/2}=\boldsymbol{I}_k+\mathcal{E}_n^{(1)}(\theta_0)+\mathcal{E}_n^{(2)}(\theta^*)$$

where

$$\mathcal{E}_n^{(1)}(\theta_0)=\boldsymbol{D}_n^{-1/2}\{\boldsymbol{F}_n(\theta_0)-\boldsymbol{D}_n\}\boldsymbol{D}_n^{-T/2}$$

and

$$\mathcal{E}_n^{(2)}(\theta^*) = \boldsymbol{D}_n^{-1/2}\{\boldsymbol{F}_n^*(\theta^*) - \boldsymbol{F}_n(\theta_0)\}\boldsymbol{D}_n^{-T/2}.$$

Let u be any unit vector $\mathbb{R}^k$. Observe that

$$u^T\mathcal{E}_n^{(1)}(\theta_0)u = \sum_{i=1}^n \sum_{1\le r,s\le 2} a_{in}^{rs}(f_i^{rs}(\theta_0) - d_i^{rs}) \tag{6.8}$$

and that

$$\sup_{\theta^*\in[N_n(A)]^4} |u^T\mathcal{E}_n^{(2)}(\theta^*)u| \le \sum_{i=1}^n \sum_{1\le r,s\le 2} |a_{in}^{rs}| \sup_{\theta^{rs}\in N_n(A)} |f_i^{rs}(\theta^{rs}) - f_i^{rs}(\theta_0)|. \tag{6.9}$$

Thus (6.8) tends to 0 in probability by (6.6), and (6.9) tends to 0 in probability by the Markov inequality, (6.5), (C1) and (F4). Hence (6.1) is proved. □

LEMMA 6.3. *If* (F4) *and* (C3) *hold, then* (6.2) *holds.*

PROOF. Let ξ_n be any unit vector in $\mathbb{R}^k$ and define $Y_{in} = \xi_n^T\boldsymbol{D}_n^{-1/2}\boldsymbol{X}_i s_i(\theta_0)$ for $1 \le i \le n$. Let $\sigma_{in}^2 = \mathrm{Var}(Y_{in}) = \xi_n^T\boldsymbol{D}_n^{-1/2}\boldsymbol{X}_i\mathcal{D}_i\boldsymbol{X}_i^T\boldsymbol{D}_n^{-T/2}\xi_n$. Thus Y_{in}, $1 \le i \le n$, are mutually independent for each n, $\mathbb{E}(Y_{in}) = 0$, and $\sigma_n^2 = \sigma_{1n}^2 + \cdots + \sigma_{nn}^2 = 1$. It follows from (C3) that

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}(|Y_{in}|^3) &\le \sum_{i=1}^n \mathbb{E}([\lambda_{\max}(s_i(\theta_0)s_i^T(\theta_0))\xi_n^T\boldsymbol{D}_n^{-1/2}\boldsymbol{X}_i\boldsymbol{X}_i^T\boldsymbol{D}_n^{-T/2}\xi_n]^{3/2})\\ &\le \sum_{i=1}^n \mathbb{E}(|s_i(\theta_0)|^3)(\mathrm{tr}(\boldsymbol{X}_i^T\boldsymbol{D}_n^{-1}\boldsymbol{X}_i))^{3/2} \le K\sum_{i=1}^n(\mathcal{U}_i\,\mathrm{tr}(\boldsymbol{X}_i^T\boldsymbol{D}_n^{-1}\boldsymbol{X}_i))^{3/2}.\end{aligned}$$

By (F4), the last expression tends to 0 as $n \to \infty$, hence (6.2) holds by the Lyapounov theorem in Billingsley ((1968), p. 44). □

7. Proof of (3.7)–(3.8) of Theorem 3.3

Recall that $\gamma_0 = p_0 = 1$. Also recall the partitioning of the information matrix $\boldsymbol{D}_n$ in (3.5) into components $\bar{\boldsymbol{D}}_n$, g_n, and c_n. Note (2.15) and (3.5) imply

$$\bar{\boldsymbol{D}}_n = \mathbb{E}(\bar{\boldsymbol{F}}_n(\theta_0)) \quad \text{with} \quad \bar{\boldsymbol{F}}_n(\theta_0) = \sum_{i=1}^n f_i^{11}(\theta_0)x_i x_i^T. \tag{7.1}$$

Partition a $k \times k$ real symmetric matrix $\boldsymbol{F}_n$ and a $k \times k$ real matrix $\tilde{\boldsymbol{F}}_n$ as

$$\boldsymbol{F}_n = \begin{bmatrix} \bar{\boldsymbol{F}}_n & f_n \\ f_n^T & a_n \end{bmatrix} \quad \text{and} \quad \tilde{\boldsymbol{F}}_n = \begin{bmatrix} \bar{\tilde{\boldsymbol{F}}}_n & \tilde{f}_n \\ \tilde{h}_n^T & \tilde{a}_n \end{bmatrix} \tag{7.2}$$

where $\bar{\boldsymbol{F}}_n$ and $\bar{\tilde{\boldsymbol{F}}}_n$ are $k_1 \times k_1$, and f_n, $\tilde{f}_n$, and $\tilde{h}_n$ are k_1-vectors, $k_1 = k - 1$. Denote by $\bar{S}_n(\theta_0)$ the k_1-vector $(S_{n1}(\theta_0) \dots S_{nk_1}(\theta_0))^T$.

LEMMA 7.1. *Suppose that* $\boldsymbol{D}_n^{-1/2}\boldsymbol{F}_n\boldsymbol{D}_n^{-T/2} \xrightarrow{P} \boldsymbol{I}_k$. *Then*

$$\bar{\boldsymbol{D}}_n^{-1/2}\bar{\boldsymbol{F}}_n\bar{\boldsymbol{D}}_n^{-T/2} \xrightarrow{P} \boldsymbol{I}_{k_1} \quad \textit{and} \quad [\boldsymbol{F}_n^{-1}]_{kk} = (1+o_p(1))[\boldsymbol{D}_n^{-1}]_{kk}.$$

Also if $\boldsymbol{D}_n^{-1/2}S_n(\theta_0) \xrightarrow{D} N(0,\boldsymbol{I}_k)$, *then* $\bar{\boldsymbol{D}}_n^{-1/2}\bar{S}_n(\theta_0) \xrightarrow{D} N(0,\boldsymbol{I}_{k_1})$.

PROOF. This is easily obtained from standard arguments. □

The next lemma is a key result since it specifies when an interior or boundary maximum of the likelihood occurs. For any fixed $A > 0$, define a subset of $\mathbb{R}^k$ by

$$N_n'(A) = \{\theta : (\theta-\theta_0)^T\boldsymbol{D}_n(\theta-\theta_0) \le A^2, \gamma=\gamma_0\}. = \{\theta = (\beta^T\ \gamma_0)^T : (\beta-\beta_0)^T\bar{\boldsymbol{D}}_n(\beta-\beta_0) \le A^2\}. \tag{7.3}$$

LEMMA 7.2. *Suppose conditions* (F1)–(F5) *hold.*

(i) *When* $\hat{\gamma}_n < \gamma_0$, $\hat{\theta}_n$ *is an interior stationary point of* $\mathcal{L}_n(\theta)$ *WPA*1, *i.e.*,

$$\mathbb{P}(S_n(\hat{\theta}_n) \ne 0, \hat{\gamma}_n < \gamma_0) \to 0 \quad (n\to\infty). \tag{7.4}$$

(ii) *Let* $\dot{\beta}_n$ *be the maximum of the log-likelihood function in* $N_n'(A)$. *We have*

$$\mathbb{P}(S_{nj}(\dot{\beta}_n,\gamma_0) = 0 \textit{ for all } j\in[1,k_1]) \to 1 \quad (n\to\infty), \tag{7.5}$$

and

$$\mathbb{P}(S_{nj}(\hat{\theta}_n) = 0 \textit{ for all } j\in[1,k_1]) \to 1 \quad (n\to\infty). \tag{7.6}$$

(iii) $\hat{\gamma}_n = \gamma_0$ *if and only if* $S_{nk}(\dot{\beta}_n,\gamma_0) \ge 0$ *WPA*1.

PROOF. (i) If an MLE $\hat{\theta}_n$ exists and $\hat{\gamma}_n < \gamma_0$, $\hat{\theta}_n$ is an interior maximum of $\mathcal{L}_n(\theta)$. Then (7.4) follows from (F1) and Theorem 3.1.

(ii) It is easy to check that $\lambda_{\min}(\bar{\boldsymbol{D}}_n) \ge \lambda_{\min}(\boldsymbol{D}_n)$. Using this and Lemma 7.1, (A1)–(A3) and (B1)–(B5) in Vu and Zhou (1997) hold for the parameter space $\bar{\Theta} = \mathbb{R}^{k_1}$, $\bar{\boldsymbol{D}}_n$, $\bar{\boldsymbol{F}}_n$, $\bar{S}_n$. Then (7.5) follows from (F1) and Theorem 3.1 as $\dot{\beta}_n$ is an interior maximum of $\mathcal{L}_n(\beta,\gamma_0)$. Since $\hat{\gamma}_n = \gamma_0$ implies that $\hat{\beta}_n = \dot{\beta}_n$, (7.6) follows from (7.4)–(7.5).

(iii) Suppose that $\hat{\theta}_n$ exists and $\hat{\gamma}_n = \gamma_0$. Then we have $\hat{\beta}_n = \dot{\beta}_n$. Let the function h_n on $\Theta_k = (a_k,\gamma_0]$ be defined by $h_n(x) = \mathcal{L}_n(\dot{\beta}_n, x)$. Then the function h_n has a maximum at $x=\gamma_0$. Thus $S_{nk}(\dot{\beta}_n,\gamma_0) \ge 0$. Since $\hat{\theta}_n$ exists WPA1 by Theorem 3.1, we have $\mathbb{P}\{S_{nk}(\dot{\beta}_n,\gamma_0) < 0, \hat{\gamma}_n = \gamma_0\} \to 0$.

Let $\dot{\theta}_n = (\dot{\beta}_n,\gamma_0)$. If $\varepsilon > 0$ is given, then there exists $A > 0$ such that

$$\mathbb{P}\{\dot{\theta}_n \in N_n'(A) \text{ and } \hat{\theta}_n \in N_n(A)\} > 1-\varepsilon. \tag{7.7}$$

Suppose that $\dot{\theta}_n \in N_n'(A)$ and $\hat{\theta}_n \in N_n(A)$. Then there exists $\tilde{\theta}_n \in N_n(A)$ between $\hat{\theta}_n$ and $\dot{\theta}_n$ such that

$$\mathcal{L}_n(\hat{\theta}_n) - \mathcal{L}_n(\dot{\theta}_n) = (\hat{\theta}_n - \dot{\theta}_n)S_n(\dot{\theta}_n) - \frac{1}{2}(\hat{\theta}_n-\dot{\theta}_n)^T\boldsymbol{F}_n(\tilde{\theta}_n)(\hat{\theta}_n-\dot{\theta}_n). \tag{7.8}$$

Since $S_{nj}(\dot\theta_n) = 0$, $j = 1, \ldots, k_1$, by (F1), it follows from (7.8) and (6.1) that

$$(\hat\theta_n - \dot\theta_n) S_n(\dot\theta_n) = (\hat\gamma_n - \gamma_0) S_{nk}(\dot\beta_n, \gamma_0) > 0 \tag{7.9}$$

if $\hat\gamma_n < \gamma_0$. Hence $\mathbb{P}\{S_{nk}(\dot\beta_n, \gamma_0) \geq 0, \hat\gamma_n < \gamma_0\} \leq \varepsilon$ for all n large enough. □

Let

$$b_n = \sqrt{c_n - g_n^T \boldsymbol{G}_n^{-1} g_n} \tag{7.10}$$

and

$$v_n^T = b_n [0 \ldots 0\ 1] \boldsymbol{D}_n^{-T/2} = \sqrt{c_n - g_n^T \boldsymbol{G}_n^{-1} g_n} [0 \ldots 0\ 1] \boldsymbol{D}_n^{-T/2}. \tag{7.11}$$

LEMMA 7.3. *v_n is a unit vector. Also*

$$[\tilde{\boldsymbol{F}}_n^{-1}]_{kk} = (\tilde a_n - \tilde h_n^T \bar{\bar{\boldsymbol{F}}}_n^{-1} \tilde f_n)^{-1}, \qquad [\boldsymbol{D}_n^{-1}]_{kk} = (c_n - g_n^T \bar{\boldsymbol{D}}_n^{-1} g_n)^{-1} = b_n^{-2}, \tag{7.12}$$

and

$$[0 \ldots 0\ 1] \tilde{\boldsymbol{F}}_n^{-1} S_n(\theta_0) = \frac{S_{nk}(\theta_0) - h_n^T \bar{\bar{\boldsymbol{F}}}_n^{-1} \bar S_n(\theta_0)}{\tilde a_n - \tilde h_n^T \bar{\bar{\boldsymbol{F}}}_n^{-1} \tilde f_n}. \tag{7.13}$$

If $\boldsymbol{D}_n^{-1/2} \tilde{\boldsymbol{F}}_n \boldsymbol{D}_n^{-T/2} \xrightarrow{P} \boldsymbol{I}_k$, then

$$b_n [0 \ldots 0\ 1] \tilde{\boldsymbol{F}}_n^{-1} S_n(\theta_0) = v_n^T \boldsymbol{D}_n^{-1} S_n(\theta_0) + o_p(1). \tag{7.14}$$

PROOF. It follows from a formula for the inverse of a partitioned matrix (Press (1982), p. 26) that the k-th row of $\tilde{\boldsymbol{F}}_n^{-1}$ is

$$\left[\frac{-\tilde h_n^T}{\tilde a_n} \left(\bar{\bar{\boldsymbol{F}}}_n - \frac{\tilde f_n \tilde h_n^T}{\tilde a_n} \right)^{-1} \quad \frac{1}{\tilde a_n - \tilde h_n^T \bar{\bar{\boldsymbol{F}}}_n^{-1} \tilde f_n} \right] = \frac{[-\tilde h_n^T \bar{\bar{\boldsymbol{F}}}_n^{-1}\ 1]}{\tilde a_n - \tilde h_n^T \bar{\bar{\boldsymbol{F}}}_n^{-1} \tilde f_n}. \tag{7.15}$$

The equality in (7.15) is obtained by observing that

$$\left(\bar{\bar{\boldsymbol{F}}}_n - \frac{\tilde f_n \tilde h_n^T}{\tilde a_n} \right)^{-1} = \bar{\bar{\boldsymbol{F}}}_n^{-1} + \frac{\bar{\bar{\boldsymbol{F}}}_n^{-1} \tilde f_n \tilde h_n^T \bar{\bar{\boldsymbol{F}}}_n^{-1}}{\tilde a_n - \tilde h_n^T \bar{\bar{\boldsymbol{F}}}_n^{-1} \tilde f_n}. \tag{7.16}$$

Similarly, the k-th row of $\boldsymbol{D}_n^{-1}$ is $b_n^{-2}[-g_n^T \bar{\boldsymbol{D}}_n^{-1}\ 1]$. Hence v_n is a unit vector and (7.12)–(7.13) hold.

Now suppose that $\boldsymbol{D}_n^{-1/2} \tilde{\boldsymbol{F}}_n \boldsymbol{D}_n^{-T/2} \xrightarrow{P} \boldsymbol{I}_k$. Let $\tilde{\boldsymbol{F}}_n^{-1} = \boldsymbol{D}_n^{-1} + \boldsymbol{D}_n^{-T/2} \boldsymbol{B}_n \boldsymbol{D}_n^{-1/2}$ where $\boldsymbol{B}_n = o_p(1)$ is a $k \times k$ real matrix. Then

$$b_n [0 \ldots 0\ 1] \tilde{\boldsymbol{F}}_n^{-1} S_n(\theta_0) = v_n^T \boldsymbol{D}_n^{-1/2} S_n(\theta_0) + v_n^T \boldsymbol{B}_n \boldsymbol{D}_n^{-1/2} S_n(\theta_0). \tag{7.17}$$

Since $|v_n^T \boldsymbol{B}_n| \xrightarrow{P} 0$ as $n \to \infty$, (7.14) follows from (7.17) and (6.2). □

Now we complete the proof of (3.7)–(3.8). We will work throughout on the event $\{\hat{\gamma}_n < \gamma_0\}$. By Lemma 7.2, $S_n(\hat{\theta}_n) = S_n(\hat{\beta}_n, \hat{\gamma}_n) = 0$ WPA1 on this event. By Taylor expansion, we have

$$(7.18)\qquad \begin{aligned} S_n(\theta_0) &= S_n(\theta_0) - S_n(\hat{\beta}_n, \gamma_0) + S_n(\hat{\beta}_n, \gamma_0) - S_n(\hat{\beta}_n, \hat{\gamma}_n) \\ &= \sum_{i=1}^n \begin{bmatrix} f_i^{11}(\tilde{\theta}_n^{11}) x_i x_i^T (\hat{\beta}_n - \beta_0) \\ f_i^{12}(\tilde{\theta}_n^{21}) x_i^T (\hat{\beta}_n - \beta_0) \end{bmatrix} + (\hat{\gamma}_n - \gamma_0) \sum_{i=1}^n \begin{bmatrix} f_i^{12}(\tilde{\theta}_n^{12}) x_i \\ f_i^{22}(\tilde{\theta}_n^{22}) \end{bmatrix} \end{aligned}$$

where $\tilde{\theta}_n^{11}$ and $\tilde{\theta}_n^{21}$ lie between θ_0 and $(\hat{\beta}_n, \gamma_0)$, and $\tilde{\theta}_n^{12}$ and $\tilde{\theta}_n^{22}$ lie between $(\hat{\beta}_n, \gamma_0)$ and $\hat{\theta}_n$. Define

$$(7.19)\qquad \tilde{\bar{\boldsymbol{F}}}_n = \sum_{i=1}^n f_i^{11}(\tilde{\theta}_n^{11}) x_i x_i^T, \qquad \tilde{h}_n = \sum_{i=1}^n f_i^{12}(\tilde{\theta}_n^{21}) x_i,$$

$$(7.20)\qquad \tilde{f}_n = \sum_{i=1}^n f_i^{12}(\tilde{\theta}_n^{12}) x_i, \qquad \tilde{a}_n = \sum_{i=1}^n f_i^{22}(\tilde{\theta}_n^{22})$$

and

$$(7.21)\qquad \tilde{\boldsymbol{F}}_n = \begin{bmatrix} \tilde{\bar{\boldsymbol{F}}}_n & \tilde{f}_n \\ \tilde{h}_n^T & \tilde{a}_n \end{bmatrix} = \sum_{i=1}^n \boldsymbol{X}_i \begin{bmatrix} f_i^{11}(\tilde{\theta}_n^{11}) & f_i^{12}(\tilde{\theta}_n^{12}) \\ f_i^{12}(\tilde{\theta}_n^{21}) & f_i^{22}(\tilde{\theta}_n^{22}) \end{bmatrix} \boldsymbol{X}_i^T.$$

Thus

$$(7.22)\qquad S_n(\theta_0) = \tilde{\boldsymbol{F}}_n(\hat{\theta}_n - \theta_0).$$

Since $\boldsymbol{D}_n^{-1/2} \tilde{\boldsymbol{F}}_n \boldsymbol{D}_n^{-T/2} \xrightarrow{P} \boldsymbol{I}_k$ by (6.1), it follows from (7.22) and (7.14) that

$$(7.23)\qquad \begin{aligned} \sqrt{c_n - g_n^T \bar{\boldsymbol{D}}_n^{-1} g_n}(\hat{\gamma}_n - \gamma_0) &= b_n [0 \ldots 0\ 1] \tilde{\boldsymbol{F}}_n^{-1} S_n(\theta_0) \\ &= v_n^T \boldsymbol{D}_n^{-1/2} S_n(\theta_0) + o_p(1) \end{aligned}$$

on $\{\hat{\gamma}_n < \gamma_0\}$, with v_n defined by (7.11).

Therefore, $\mathbb{P}\{\hat{\gamma}_n < \gamma_0, v_n^T \boldsymbol{D}_n^{-1/2} S_n(\theta_0) \ge 0\} \to 0$ as $n \to \infty$. Suppose that

$$(7.24)\qquad \mathbb{P}\{\hat{\gamma}_n = \gamma_0, v_n^T \boldsymbol{D}_n^{-1/2} S_n(\theta_0) < 0\} \to 0 \quad \text{as} \quad n \to \infty.$$

Then it follows from (6.2) that for $x < 0$

$$\lim_{n \to \infty} \mathbb{P}\left\{\sqrt{c_n - g_n^T \bar{\boldsymbol{D}}_n^{-1} g_n}(\hat{p}_n - 1) \le x\right\} = \mathbb{P}\{N \le x\}.$$

Thus (3.7)–(3.8) follow from (7.24). Now we prove (7.24). Since $\bar{S}_n(\dot{\theta}_n) = 0$ WPA1 by (7.5), there exist by Taylor's expansion $\tilde{\theta}_{n1}$ and $\tilde{\theta}_{n2}$ between $\dot{\theta}_n$ and θ_0 such that

$$\bar{S}_n(\theta_0) = \bar{S}_n(\theta_0) - \bar{S}_n(\dot{\theta}_n) = \sum_{i=1}^{n} x_i[s_{i1}(\theta_0) - s_{i1}(\dot{\theta}_n)] \tag{7.25}$$
$$= \sum_{i=1}^{n} f_i^{11}(\tilde{\theta}_{n1})x_i x_i^T(\dot{\beta}_n - \beta_0) = \bar{\dot{\boldsymbol{F}}}_n(\dot{\beta}_n - \beta_0)$$

and

$$S_{nk}(\dot{\theta}_n) = S_{nk}(\theta_0) - \sum_{i=1}^{n} f_i^{22}(\tilde{\theta}_{n2})x_i^T(\dot{\beta}_n - \beta_0) = S_{nk}(\theta_0) - \dot{h}_n^T(\dot{\beta}_n - \beta_0) \tag{7.26}$$

where

$$\bar{\dot{\boldsymbol{F}}}_n = \sum_{i=1}^{n} f_i^{11}(\tilde{\theta}_{n1})x_i x_i^T, \qquad \dot{h}_n = \sum_{i=1}^{n} f_i^{22}(\tilde{\theta}_{n2})x_i, \qquad \text{and} \qquad \dot{\boldsymbol{F}}_n = \begin{bmatrix} \bar{\dot{\boldsymbol{F}}}_n & \dot{h}_n \\ \dot{h}_n^T & c_n \end{bmatrix}.$$

Then (6.1) implies that $\boldsymbol{D}_n^{-1/2}\dot{\boldsymbol{F}}_n\boldsymbol{D}_n^{-T/2} \xrightarrow{P} \boldsymbol{I}_k$. By substituting $\dot{\boldsymbol{F}}_n$, $\bar{\dot{\boldsymbol{F}}}_n$, $\dot{h}_n$ and c_n for $\tilde{\boldsymbol{F}}_n$, $\bar{\tilde{\boldsymbol{F}}}_n$, $\tilde{h}_n$, $\tilde{f}_n$ and $\tilde{a}_n$ into (7.13)–(7.14), it follows from (7.25)–(7.26) that

$$\frac{b_n S_{nk}(\dot{\theta}_n)}{c_n - \dot{h}_n^T \bar{\dot{\boldsymbol{F}}}_n^{-1}\dot{h}_n} = \frac{b_n(S_{nk}(\theta_0) - \dot{h}_n^T \bar{\dot{\boldsymbol{F}}}_n^{-1}\bar{S}_n(\theta_0))}{c_n - \dot{h}_n^T \bar{\dot{\boldsymbol{F}}}_n^{-1}\dot{h}_n} \tag{7.27}$$
$$= v_n^T\boldsymbol{D}_n^{-1/2}S_n(\theta_0) + o_p(1).$$

By Lemma 7.2, the event $\{\hat{\gamma}_n = \gamma_0\}$ occurs if and only if $\{S_{nk}(\dot{\theta}_n) \geq 0\}$ occurs WPA1. Thus (7.24) follows from (7.27). This completes the proof of (3.7)–(3.8). □

8. Proof of (3.9) of Theorem 3.3

Suppose that $\boldsymbol{D}_n^{1/2}$ and $\bar{\boldsymbol{D}}_n^{1/2}$ are the left Cholesky square roots of $\boldsymbol{D}_n$ and $\bar{\boldsymbol{D}}_n$ and define

$$\boldsymbol{Q}_n^T = \begin{bmatrix} \bar{\boldsymbol{D}}_n^{T/2} & \bar{\boldsymbol{D}}_n^{-1/2}g_n \\ 0 & b_n \end{bmatrix} \boldsymbol{D}_n^{-T/2} \tag{8.1}$$

where b_n is given by (7.10). It is easy to check that $\boldsymbol{Q}_n\boldsymbol{Q}_n^T = \boldsymbol{I}_k$. Since the inverses and products of lower (upper) triangular matrices are also lower (upper) triangular matrices, $\boldsymbol{Q}_n^{-1}$ is a lower triangular matrix while $\boldsymbol{Q}_n^T$ is an upper triangular matrix. Thus $\boldsymbol{Q}_n$ must be diagonal because $\boldsymbol{Q}_n^T = \boldsymbol{Q}_n^{-1}$. Moreover, the diagonal elements of $\boldsymbol{Q}_n$ can only be 1 since they are obviously positive and since $\boldsymbol{Q}_n$ is orthogonal. Consequently, $\boldsymbol{Q}_n = \boldsymbol{I}_k$.

Assume from now on that (F1)–(F5) are satisfied.

Consider the distribution of $\hat{\beta}_n$ when $\hat{\gamma}_n = \gamma_0 = 1$. Let $\boldsymbol{D}_n^{-1}S_n(\theta_0) = (X_n^T\ Y_n)^T$ where X_n is a k_1-vector and Y_n is real-valued. Then

$$S_n(\theta_0) = \begin{bmatrix} \bar{S}_n(\theta_0) \\ S_{nk}(\theta_0) \end{bmatrix} = \boldsymbol{D}_n \begin{bmatrix} X_n \\ Y_n \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{D}}_n X_n + Y_n g_n \\ g_n^T X_n + c_n Y_n \end{bmatrix}.$$

Recall that b_n and v_n are defined by (7.10)–(7.11). Since it is shown in the proof of Lemma 7.3 that the last row of $\boldsymbol{D}_n^{-1}$ is $b_n^{-2}[-g_n^T\bar{\boldsymbol{D}}_n^{-1}\ 1]$, we have

$$\begin{aligned}(8.2)\qquad v_n^T\boldsymbol{D}_n^{-1/2}S_n(\theta_0) &= b_n[0\ldots0\ 1]\boldsymbol{D}_n^{-1}S_n(\theta_0) = b_n^{-1}(-g_n^T\bar{\boldsymbol{D}}_n^{-1}\ \gamma_0)S_n(\theta_0)\\ &= b_n^{-1}(-g_n^TX_n - g_n^T\bar{\boldsymbol{D}}_n^{-1}g_nY_n + g_n^TX_n + c_nY_n) = b_nY_n.\end{aligned}$$

Let $Z_n = ((\bar{\boldsymbol{D}}_n^{T/2}X_n + Y_n\bar{\boldsymbol{D}}_n^{-1/2}g_n)^T\ b_nY_n)^T = \boldsymbol{Q}_n^T\boldsymbol{D}_n^{T/2}[X_n^T\ Y]^T$ where $\boldsymbol{Q}_n$ is defined by (8.1). But then $\boldsymbol{Q}_n^T = \boldsymbol{I}_k$ if $\boldsymbol{D}_n^{T/2}$ and $\bar{\boldsymbol{D}}_n^{T/2}$ are the right Cholesky square roots of $\boldsymbol{D}_n$ and $\bar{\boldsymbol{D}}_n$, as we showed above. It follows from (6.2) that

$$(8.3)\qquad Z_n = \boldsymbol{D}_n^{T/2}(X_n^T\ Y_n)^T = \boldsymbol{D}_n^{-1/2}S_n(\theta_0) \xrightarrow{D} N(0,\boldsymbol{I}_k).$$

Recall that $Z = (X^T\ Y)^T \sim N(0,\boldsymbol{I}_k)$. Let B be a Borel set in $\mathbb{R}^k$. Since the event $\{\hat{\gamma}_n < \gamma_0\}$ occurs if and only if the event $\{Y_n < 0\}$ occurs WPA1 by (8.2), as shown in the proof of (3.7) and (3.8), (3.9) follows from (8.1)–(8.2) and (6.2) via

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{P}(\boldsymbol{D}_n^{T/2}(\hat{\theta}_n-\theta_0)\in B, \hat{\gamma}_n < \gamma_0)\\ &\quad= \lim_{n\to\infty}\mathbb{P}(\boldsymbol{D}_n^{-1/2}S_n(\theta_0)\in B, Y_n < 0)\\ &\quad= \lim_{n\to\infty}\mathbb{P}(\boldsymbol{D}_n^{-1/2}S_n(\theta_0)\in B, [\boldsymbol{D}_n^{-1/2}S_n(\theta_0)]_k < 0) = \mathbb{P}(Z\in B, Y\le 0).\end{aligned}$$

References

Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1989). *Statistical Modelling in GLIM*, Oxford Science Publications, New York.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*, Wiley, Chichester.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Collett, D. (1994). *Modelling Survival Data in Medical Research*, Chapman and Hall, London.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models, *Ann. Statist.*, **13**, 342–368.

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

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.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models, *J. Roy. Statist. Soc. Ser. A*, **135**, 370–384.

Press, S. J. (1982). *Applied Multivariate Analysis*, 2nd ed., Krieger, Malabar, Florida.

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

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

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