Random Effects Models for Multivariate Survival Data: Hierarchical–Likelihood Approach

Il Do Ha[1], Youngjo Lee[2] and Jae–Kee Song[3]

Abstract

Modelling the dependence via random effects in censored multivariate survival data has recently received considerable attention in the biomedical literature. The random effects models model not only the conditional survival times but also the conditional hazard rate. Systematic likelihood inference for the models with random effects is possible using Lee and Nelder's (1996) hierarchical-likelihood (h–likelihood). The purpose of this presentation is to introduce Ha et al.'s (2000a,b) inferential methods for the random effects models via the h–likelihood, which provide a conceptually simple, numerically efficient and reliable inferential procedures.

Keywords: Frailty models; Hierarchical-likelihood; Marginal likelihood; Mixed linear models; Multivariate survival data; Random effects; Restricted maximum likelihood.

1. Introduction

Multivariate survival data are frequently encountered in biomedical research, and include multiple events experienced by an individual, matched pairs, or family members sharing the same genetic background; see, for example, data on kidney infection (McGilchrist and Aisbett, 1991), skin allografts (Andersen et al., 1997) and litter-matched tumorigenesis (Klein et al., 1999). Recently, a number of authors have proposed using the models with random effect, a common unobservable effect within the same individual, to account for the dependence between the survival times. In particular, they have widely used the random effects models such as mixed linear models (MLMs) and frailty models, an extension of linear models and Cox's (1972) proportional hazards models, respectively. For the former the random effect is modelled by acting linearly on each individual's survival time, whereas for the latter it (often, called by frailty) by acting multiplicatively on the individual's hazard rate.

Inferences for MLMs have been studied by many authors. Pettitt (1986) and Hughes (1999) proposed maximum-likelihood estimation procedures using respectively the EM algorithm and a Monte Carlo EM algorithm based on Gibbs sampling, both of which are computationally intensive. Klein et al. (1999) derived the Newton-Raphson method, but it is very complicated to obtain the marginal likelihood.

[1]Faculty of Information Science, Kyungsan University, Kyungpook, 712-240, Korea.
[2]Department of Statistics, Seoul National University, Seoul, 151-742, Korea.
[3]Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

Inferences for frailty models also have been studied by many authors. For the lognormal frailty model, McGilchrist and Aisbett (1991) and McGilchrist (1993) developed an estimation procedure using Cox's (1972, 1975) partial likelihood, but their method does not handle ties between survival times. For the mathematically convenient gamma frailty model, Klein (1992) and Nielsen et al. (1992) developed EM estimation procedures, which provide the same estimators; Nielsen et al. (1992) presented a counting process framework and Parner (1998) showed the consistency and asymptotic normality of Nielsen et al.'s estimator. However, with the EM method in general, the conditional expectation of frailty given the observed data would be computed via a numerical integration (except for the gamma frailty model, where an analytic solution exists) and also the variance estimates for the estimated parameters are not directly available; see for example Louis (1982) and Jamshidian and Jennrich (1997). For the gamma frailty model, Clayton (1991) and Aslanidou et al. (1998) developed Markov chain Monte Carlo procedures, which may be computationally intensive. For other frailty models see Hougaard (1987) and Costigan and Klein (1993).

Lee and Nelder (1996) proposed the use of h-likelihood for inferences from models with random effects, and they (2000) showed by numerical study that their procedure yields efficient parameter estimators in all the cases they studied. We (2000a,b) developed a new inferential method for the MLMs and frailty models via the h-likelihood, which provides a conceptually simple, numerically efficient and reliable inferential procedure. For the MLMs we (2000a) showed by simulation studies that our procedure is robust to violation of the normal assumption if the censoring rate is not too high. We (2000b) also showed that our procedure provides a simple unified framework for the frailty models with various frailty distributions including the lognormal and gamma. For the gamma frailty model, given the frailty parameter, our procedure is the same as that of Klein (1992) and Nielsen et al. (1992), and for the lognormal frailty model it becomes an extension of McGilchrist's (1993) restricted maximum likelihood (REML) method for data including ties. In Section 2 we introduce data structure and assumptions for both models. In Sections 3 and 4 we present a new estimation procedure for the Normal MLM and the frailty models, respectively.

2. Data Structure and Assumption

Let T_{ij} $(i = 1, \ldots, q,\ j = 1, \ldots, n_i,\ n = \sum_i n_i)$ be the survival time for the jth observation of the ith individual and C_{ij} be the corresponding censoring time. The observable random variables are $Y_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = I(T_{ij} \leq C_{ij})$, where $I(\cdot)$ is the indicator function. Denote by U_i the unobserved random effect on the ith individual.

We make the following two assumptions: see also Nielsen et al. (1992).

Assumption 1. Given $U_i = u_i$, $\{(T_{ij}, C_{ij}), j = 1, \ldots, n_i\}$ are conditionally independent and both T_{ij} and C_{ij} are also conditionally independent for $j = 1, \ldots, n_i$.
Assumption 2. Given $U_i = u_i$, $\{C_{ij}, j = 1, \ldots, n_i\}$ are noninformative about u_i.

3. Estimation of Mixed Linear Models

For T_{ij} we assume the Normal MLM as follows: for $i = 1, \ldots, q$ and $j = 1, \ldots, n_i$,

$$T_{ij} = x_{ij}^t \beta \ + \ U_i \ + \ \epsilon_{ij},$$

where $x_{ij} = (x_{ij1}, \ldots, x_{ijp})^t$ is a vector of fixed covariates, β is a $p \times 1$ vector of fixed effects and $U_i \sim N(0, \sigma_u^2)$ and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ are independent. Here, the dispersion or variance components σ_ϵ^2 and σ_u^2 stand for variability within and between individuals, respectively. The T_{ij} could be expressed on some suitably transformed scale, e.g., $\log(T_{ij})$. With the log-transformation, the Normal MLM becomes an accelerated failure-time model with random effects.

We shall first present a simple method for estimating the parameters in the Normal MLM. Because T_{ij}'s may be subject to censoring, only Y_{ij}'s are observed, but

$$E(Y_{ij}|U_i = u_i) \neq \mu_{ij},$$

where $\mu_{ij} = E(T_{ij}|U_i = u_i) = x_{ij}^t\beta + u_i$. Now, we consider an extended form of the pseudo-response variable Y_{ij}^* of Buckley and James (1979) for the linear model with censored data as follows:

$$Y_{ij}^* = Y_{ij}\delta_{ij} + E(T_{ij}|T_{ij} > Y_{ij}, U_i = u_i)(1 - \delta_{ij}). \tag{3.1}$$

Then from the conditional independence of T_{ij} and C_{ij} in Assumption 1, we have

$$E(Y_{ij}^*|U_i = u_i) = \mu_{ij} \; . \tag{3.2}$$

The proof of the expectation identity (3.2) is given in Appendix 1 of Ha et al. (2000a). Let y_{ij} and y_{ij}^* be the observed values for Y_{ij} and Y_{ij}^*, respectively. It can be easily shown that if $T_{ij}|(U_i = u_i) \sim N(\mu_{ij}, \sigma_\epsilon^2)$, then (3.1) becomes

$$y_{ij}^* = y_{ij}\delta_{ij} + \{\mu_{ij} + \sigma_\epsilon V(m_{ij})\}(1 - \delta_{ij}), \tag{3.3}$$

where $V(\cdot) = \phi(\cdot)/\bar{\Phi}(\cdot)$ is the hazard function for $N(0,1)$, ϕ and $\Phi(= 1 - \bar{\Phi})$ are the density and cumulative distribution functions for $N(0,1)$, respectively, and $m_{ij} = (y_{ij} - \mu_{ij})/\sigma_\epsilon$.

Under Assumptions 1 and 2 and following Ha et al. (2000b), the h-likelihood for the Normal MLM, denoted by h, is defined by

$$h = h(\beta, \sigma_\epsilon^2, \sigma_u^2) = \sum_{ij} \ell_{1ij} + \sum_i \ell_{2i}, \tag{3.4}$$

where $\ell_{1ij} = \ell_{1ij}(\beta, \sigma_\epsilon^2; y_{ij}, \delta_{ij}|u_i) = -\,\delta_{ij}\{\log(2\pi\sigma_\epsilon^2) + (m_{ij})^2\}/2 + (1-\delta_{ij})\log\{\bar{\Phi}(m_{ij})\}$ is the logarithm of the conditional density function for Y_{ij} and δ_{ij} given $U_i = u_i$, and $\ell_{2i} = \ell_{2i}(\sigma_u^2; u_i) = -\,\{\log(2\pi\sigma_u^2) + (u_i^2/\sigma_u^2)\}/2$ is the logarithm of the density function for U_i.

Given the dispersion components $\theta = (\sigma_\epsilon^2, \sigma_u^2)$, the maximum h-likelihood estimators (MHLEs) of $\tau = (\beta, u)$ with $u = (u_1, \ldots, u_q)^t$ are obtained by solving

$$\frac{\partial h}{\partial \beta_k} = \frac{1}{\sigma_\epsilon} \sum_{ij} \Big\{\delta_{ij} m_{ij} + (1 - \delta_{ij}) V(m_{ij})\Big\} x_{ijk} = 0 \;\; (k = 1, \ldots, p), \tag{3.5}$$

$$\frac{\partial h}{\partial u_i} = \frac{1}{\sigma_\epsilon} \sum_{j} \Big\{\delta_{ij} m_{ij} + (1 - \delta_{ij}) V(m_{ij})\Big\} \; - \; \frac{1}{\sigma_u^2} u_i = 0 \;\; (i = 1, \ldots, q). \tag{3.6}$$

Plugging (3.3) into the two MHL equations (3.5) and (3.6) reduces them, respectively, to

$$\frac{1}{\sigma_\epsilon^2}\sum_{ij}\Big(y_{ij}^* - \mu_{ij}\Big)x_{ijk} = 0 \ (k = 1,\ldots,p), \tag{3.7}$$

$$\frac{1}{\sigma_\epsilon^2}\sum_{j}\Big(y_{ij}^* - \mu_{ij}\Big) - \frac{1}{\sigma_u^2}u_i = 0 \ (i = 1,\ldots,q). \tag{3.8}$$

Because we cannot observe all the y_{ij}^*'s, we replace them by their estimates

$$\widehat{y_{ij}^*} = y_{ij}\delta_{ij} + \{\widehat{\mu}_{ij} + \widehat{\sigma}_\epsilon V(\widehat{m}_{ij})\}(1-\delta_{ij}),$$

where $\widehat{m}_{ij} = (y_{ij} - \widehat{\mu}_{ij})/\widehat{\sigma}_\epsilon$ and $\widehat{\mu}_{ij} = x_{ij}^t\widehat{\beta} + \widehat{u}_i$. When there is no censoring the equations (3.7) and (3.8) become Henderson's (1975) mixed-model equations using the data y_{ij}. These two estimating equations are also extensions to those of Wolynetz (1979), Schmee and Hahn (1979) and Aitkin (1981) for normal linear models with no random effects. From (3.7) and (3.8), given θ and y^*, the MHLEs $\hat{\tau} = (\hat{\beta}, \hat{u})$ are obtained by solving iteratively Henderson's (1975) mixed-model equations with pseudo-response variables y^*

$$\begin{pmatrix} X^tX & X^tZ \\ Z^tX & Z^tZ + \phi I_q \end{pmatrix}\begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X^ty^* \\ Z^ty^* \end{pmatrix}, \tag{3.9}$$

where X is the $n\times p$ matrix whose ijth row vector is x_{ij}^t, Z is the $n\times q$ group indicator matrix whose ijkth element z_{ijk} is $\partial\mu_{ij}/\partial u_k$, I_q is the $q\times q$ identity matrix, y^* is the $n\times 1$ vector with the ijth element y_{ij}^* and $\phi = \sigma_\epsilon^2/\sigma_u^2$. The asymptotic covariance matrix (Lee and Nelder, 1996) for $\hat{\tau} - \tau$ is given by H^{*-1} with

$$H^* = -\frac{\partial^2 h}{\partial \tau^2} = \frac{1}{\sigma_\epsilon^2}H, \tag{3.10}$$

where

$$H = \begin{pmatrix} X^tWX & X^tWZ \\ Z^tWX & Z^tWZ + \phi I_q \end{pmatrix}.$$

Here, $W = \mathrm{diag}(w_{ij})$ is the $n\times n$ diagonal matrix with the ijth element $w_{ij} = \delta_{ij} + (1-\delta_{ij})\xi(m_{ij})$ and $\xi(m_{ij}) = V(m_{ij})\{V(m_{ij}) - m_{ij}\}$. So, the upper left-hand corner of H^{*-1} in (3.10) gives the variance matrix of $\hat{\beta}$;

$$\mathrm{var}(\hat{\beta}) = \sigma_\epsilon^2(\mathrm{X}^t\Sigma^{-1}\mathrm{X})^{-1}, \tag{3.11}$$

where $\Sigma = W^{-1} + \phi^{-1}ZZ^t$. Note that both y^* in (3.9) and W in (3.10) depend on censoring patterns and that W is the weight matrix which takes into account the loss of information due to censoring; if the ijth observation is uncensored $w_{ij} = 1$.

For the estimation of the dispersion parameters θ given the estimates of τ, we use Lee and Nelder's (1996) adjusted profile h-likelihood, defined by

$$h_P = h_A|_{\tau=\hat{\tau}},$$

where $h_A = h + \frac{1}{2}\log\{\det(2\pi \mathrm{H}^{*-1})\}$. The maximum adjusted profile h-likelihood estimators (MAPHLEs) for θ are obtained by solving iteratively

$$\frac{\partial h_A}{\partial \theta}\Big|_{\tau=\hat{\tau}} = 0 \ , \tag{3.12}$$

where $\hat{\tau}$ is re-evaluated at each iteration. Though not reported here, we found via simulation studies that Lee and Nelder's (2000) first- and second-order REML methods show no improvement over their 1996 MAPHLEs for the Normal MLM. We (2000a) thus use the original MAPHLEs for σ_ϵ^2 and σ_u^2, given by

$$\widehat{\sigma_\epsilon^2} = \frac{\sum_{ij}(\widehat{y_{ij}^*} - \widehat{\mu}_{ij})^2}{n_1 - (p+q-\gamma_1)} \quad \text{and} \quad \widehat{\sigma_u^2} = \frac{\sum_i \widehat{u_i}^2}{q-\gamma_2}, \tag{3.13}$$

where n_1, γ_1 and γ_2 are given in Appendix 2 of Ha et al. (2000a). McGilchrist (1993) showed by simulation that the REML method gives a good standard-error estimator of $\widehat{\beta}$ for log-normal frailty model. When there is no censoring, $W = I$, so that the MAPHLEs become the REML estimators of Patterson and Thompson (1971). By simulation we (2000a) showed that the estimator of $\mathrm{var}(\widehat{\beta})$ in (3.11) using the MAPHLEs is reasonably good.

4. Estimation of Frailty Models

Given $U_i = u_i$ the conditional hazard function of T_{ij} is of the form

$$\lambda(t_{ij}|u_i) = \lambda_0(t_{ij})\exp(x_{ij}^t\beta)u_i, \tag{4.1}$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function. The frailties U_i are assumed to be i.i.d. random variables with a density function having frailty parameter α; the gamma and lognormal frailty models assume gamma and lognormal distributions for U_i, respectively.

Under Assumptions 1 and 2 and following Ha et al. (2000b), the h-likelihood for the frailty models, denoted by h, is defined by

$$h = h(\beta, \Lambda_0, \alpha) = \sum_{ij}\ell_{1ij} + \sum_i \ell_{2i}, \tag{4.2}$$

where $\ell_{1ij} = \ell_{1ij}(\beta, \Lambda_0; y_{ij}, \delta_{ij}|u_i) = \delta_{ij}\{\log\lambda_0(y_{ij}) + \eta_{ij}'\} - \{\Lambda_0(y_{ij})\exp(\eta_{ij}')\}$ is the logarithm of the conditional density function for Y_{ij} and δ_{ij} given $U_i = u_i$, $\ell_{2i} = \ell_{2i}(\alpha; v_i)$ is the logarithm of the density function for $V_i = \log(U_i)$, $\Lambda_0(\cdot)$ is the baseline cumulative hazard function, and $\eta_{ij}' = \eta_{ij} + v_i$ with $\eta_{ij} = x_{ij}^t\beta$ and $v_i = \log(u_i)$.

Let $\Lambda_0(t)$ be a step function with jumps at the observed death times. That is, $\Lambda_0(t) = \sum_{k:y_{(k)}\le t}\lambda_0(y_{(k)})$, where $y_{(k)}$ is the kth smallest distinct death time among the y_{ij}'s. Let $v = (v_1, \ldots, v_q)^t$. Then, by arguments similar to those in Johansen (1983), given $\tau = (\beta, v)$ the score equations $\partial h/\partial\lambda_0(y_{(k)}) = 0$ provide the nonparametric MHLE of $\Lambda_0(t)$:

$$\widehat{\Lambda}_0(t) = \sum_{k:y_{(k)}\le t}\left\{\frac{d_{(k)}}{\sum_{ij\in R(y_{(k)})}\exp(\eta_{ij}')}\right\}, \tag{4.3}$$

where $R(y_{(k)}) = \{ij : y_{ij} \geq y_{(k)}\}$ is the risk set at time $y_{(k)}$ and $d_{(k)}$ is the number of deaths at $y_{(k)}$. Thus the MHLE for $\tau = (\beta, v)$ can be obtained by maximising the profile h-likelihood h^* after eliminating $\Lambda_0(t)$:

$$h^* = h|_{\Lambda_0=\widehat{\Lambda}_0} \,. \tag{4.4}$$

Let $\eta'_{ij} = x^t_{ij}\beta + v_i$ be $\eta'_{ij} = x^t_{ij}\beta + z^t_{ij}v$, where $z_{ij} = (z_{ij1}, \ldots, z_{ijq})^t$ is the $q \times 1$ group indicator vector whose rth element is $\partial\eta'_{ij}/\partial v_r$. The kernel of h^* in (4.4) becomes

$$\sum_k \Big[s^t_{1(k)}\beta + s^t_{2(k)}v - d_{(k)} \log\Big\{ \sum_{ij \in R(y_{(k)})} \exp(\eta'_{ij}) \Big\}\Big] \; + \; \sum_i \ell_{2i}(\alpha; v_i), \tag{4.5}$$

where $s^t_{1(k)} = \sum_{ij \in D_{(k)}} x^t_{ij}$ and $s^t_{2(k)} = \sum_{ij \in D_{(k)}} z^t_{ij}$ are the sums of the vectors x^t_{ij} and z^t_{ij} over the set $D_{(k)}$ of individuals who die at $y_{(k)}$, respectively. Note that the estimator (4.3) and the profile h-likelihood h^* defined by (4.5) are respectively extensions of Breslow's (1974) estimator of the baseline cumulative hazard function and Breslow's (1974) partial likelihood for the Cox model to the frailty models, and also that the h^* is an extension of the partial likelihood of McGilchrist and Aisbett (1991) and McGilchrist (1993) for the lognormal frailty model.

Remark 4.1. Marginal likelihood, denoted by m, has been often used for inference; see for example Nielsen et al. (1992). Under Assumptions 1 and 2, m can be obtained by integrating out the frailty random variables from the h-likelihood:

$$m = m(\beta, \Lambda_0, \alpha) = \sum_i \log\Big\{ \int \exp(h_i)\, dv_i \Big\}, \tag{4.6}$$

where $h_i = \sum_j \ell_{1ij} \; + \; \ell_{2i}$ is the contribution of the ith individual to h in (4.2). We (2000b) showed that, in the gamma frailty model, the MHLE for β given α is the same as the maximum marginal likelihood estimator (MMLE), which is obtained by maximising the profile marginal likelihood after eliminating $\Lambda_0(t)$.

We use the Newton-Raphson method to solve the score equations $\partial h^*/\partial\tau = 0$. That is, given α the MHLE of τ is obtained by solving iteratively

$$\begin{pmatrix} \widehat{\beta}^{(i+1)} \\ \widehat{v}^{(i+1)} \end{pmatrix} = \begin{pmatrix} \widehat{\beta}^{(i)} \\ \widehat{v}^{(i)} \end{pmatrix} \; + \; J^{-1} \begin{pmatrix} \partial h^*/\partial\beta \\ \partial h^*/\partial v \end{pmatrix}_{(\beta,v)=(\hat{\beta}^{(i)},\hat{v}^{(i)})}, \tag{4.7}$$

where $J = -\;\partial^2 h^*/\partial\tau^2$ is the $(p+q) \times (p+q)$ observed information matrix whose inverse is the asymptotic covariance matrix of $\widehat{\beta}$ and $\widehat{v} - v$; see Lee and Nelder (1996). Even though a number of authors have suggested ways to obtain valid standard error estimates from the EM algorithm and also to accelerate its convergence, our procedures are faster and provide a direct estimate of $\text{var}(\hat{\beta})$ from the observed information matrix required for the Newton–Raphson method. For the gamma frailty model we can show that, given α, the standard error estimate of $\hat{\beta}$ calculated from the derivatives of h^* agrees algebraically with that of Andersen et al. (1997), which have showed how to obtain valid standard error estimates using the marginal likelihood. For the Cox model without the frailty the score equations (4.7) become those of Breslow (1974), and for the lognormal frailty model without ties they become those of McGilchrist and Aisbett (1991) and McGilchrist (1993). Next, for the estimation of the frailty parameter α given estimates of τ, we use Lee & Nelder's (1996) adjusted profile h-likelihood h_P^*, defined by

$$h_P^* = h_A^*|_{\beta=\hat{\beta},v=\hat{v}} \ ,$$

where $h_A^* = h^* + \frac{1}{2}\log\{\det(2\pi J^{-1})\}$. The MAPHLE for α is obtained by solving iteratively

$$\partial h_A^*/\partial\alpha|_{\beta=\hat{\beta},v=\hat{v}} = 0,$$

where $\hat{\beta}$ and $\hat{v}$ are re-evaluated in each iteration. In the lognormal frailty model where the V_i's are normal with mean 0 and variance α, it can be easily shown that

$$\partial h_A^*/\partial\alpha|_{\beta=\hat{\beta},v=\hat{v}} = (q-\gamma) - \sum_i \hat{v}_i^2/\alpha,$$

where $\gamma = \mathrm{tr}(K)/\alpha$ and K is the matrix given by the bottom right-hand corner of J^{-1}; this yields McGilchrist's (1993) REML estimator for α. Furthermore, for the gamma frailty model with $E(U_i) = 1$ and $\mathrm{var}(U_i) = \alpha$ the MAPHLE provides an extended REML.

References

Aitkin, M. A. (1981). A note on the regression analysis of censored data. *Technometrics*, 23 161-163.

Andersen, P. K., Klein, J. P., Knudsen, K. and Palacios, R. T. (1997). Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics*, 53, 1475-1484.

Aslanidou, H., Dey, D. K. and Sinha, D. (1998). Bayesian analysis of multivariate survival data using Monte Carlo methods. *The Canadian Journal of Statistics*, 26, 33-48.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics*, 30, 89-99.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66, 429-436.

Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, 47, 467-485.

Costigan, T. M. and Klein, J. P. (1993). Multivariate survival analysis based on frailty models. *Advances In Reliability*, Ed. A. P. Basu, 43-58. New York: North-Holland.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* B, 74, 187-220.

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

Ha, I. D., Lee, Y. and Song, J.-K. (2000a). Hierarchical likelihood approach for frailty models. To appear in *Biometrika*.

Ha, I. D., Lee, Y. and Song, J.-K. (2000b). Hierarchical likelihood approach for mixed linear models with censored data. To appear in *Lifetime Data Analysis*.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31, 423-447.

Hougaard, P. (1987). Modelling multivariate survival. *Scandinavian Journal of Statistics*, 14, 291-304.

Hughes, J. P. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics*, 55, 625-629.

Jamshidian, M. and Jennrich, R. I. (1997). Acceleration of the EM algorithm by using quasi-Newton methods. *Journal of the Royal Statistical Society* B, 59, 569-587.

Johansen, S. (1983). An extension of Cox's regression model. *International. Statistical Review*, 51, 165-174.

Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, 48, 795-806.

Klein, J. P., Pelz, C. and Zhang, M. (1999). Modelling random effects for censored data by a multivariate normal regression model. *Biometrics*, 55, 497-506.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society* B, 58, 619-678.

Lee, Y. and Nelder, J. A. (2000). Hierarchical generalized linear models: a synthesis of generalized linear models, random-effect models and structured dispersions," Technical report, Department of Mathematics, Imperial College, London.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society* B 44, 226-233.

McGilchrist, C. A. (1993). REML estimation for survival models with frailty. *Biometrics*, 49, 221-225.

McGilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics*, 47, 461-466.

Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sørensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 19, 25-44.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Annals of Statistics*, 26, 181-214.

Patterson, H. D. and Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika*, 58, 545-554.

Pettitt, A. N. (1986). Censored observations, repeated measures and mixed effects models: An approach using the EM algorithm and normal errors. *Biometrika*, 73, 635-643.

Schmee, J. and Hahn, G. J. (1979). A simple method for regression analysis with censored data. *Technometrics*, 21, 417-423.

Wolynetz, M. S. (1979). Maximum likelihood estimation in a linear model from confined and censored normal data (Algorithm AS139). *Applied Statistics*, 28, 195-206.