Model Selection for Frailty Structures

Gilbert MacKenzie[1], Il Do Ha [2] and Youngjo Lee[3]

[1] Centre of Biostatistics, University of Limerick, Ireland
[2] Daegu Haany University, South Korea
[3] Seoul National University, South Korea

Abstract: Frailty models are now widely used for analyzing multivariate survival data. An open question is how best to determine how to select the most appropriate frailty structure supported by the data. Herein, we develop a procedure for selecting the optimal frailty structure from a set of (possibly) non-nested frailty models. Our focus is on the dispersion parameters which define the frailty structure. We propose two new AIC criteria: one based on the deviance for goodness of fit and the other on the extended restricted likelihood (ERL) of Lee and Nelder (1996). A simulation study shows that the AIC based on the extended restricted likelihood is better when attention is focussed on selecting the frailty structure.

Keywords: Model Selection; AIC; DIC; ERL; Frailty Models; h-likelihood.

1 Introduction

Various multivariate survival models based of different frailty structures have now been developed; for example, time dependent models (Yau & McGilchrist, 1998) and nested models (Sastry, 1997; Yau, 2001). Typically these are based on Cox's (1972) proportional hazards (PH) model which has been generalized by including frailty components reflecting the complexity of the study design. Whilst all of these frailty components may be formally included in any model, not all may be supported by the data. Accordingly, it is important to develop a model selection procedure in which inference is focussed on the frailty variance components.

In this article the notion of focussing is intimately connected with nuisance parameter elimination, so that a focussed vehicle for inference, eg a likelihood, depends only on the parameters of interest - in this context - the variance components. Later we shall formalize this notion statistically in the context of h-likelihood inference.

Hierarchical generalized linear models (HGLMs), Lee & Nelder, (1996, 2001, 2005) extend, considerably, the class of random-effect models in the Exponential family. In a series of papers Lee & Nelder have developed inference for the parameters in these models based on their h-likelihood approach. In particular, for model selection, the deviance of goodness of

fit, D, in HGLMs can be used to form an information criterion, the Akaike information criterion (AIC, Akaike, 1973). Spiegelhalter et al. (2002) studied this type of deviance information criterion (DIC) in a Bayesian setting. Moreover, the extended restricted likelihood (ERL) of Lee and Nelder (1996, 2003) can also be used as a model selection criterion for dispersion parameters in the HGLM setting.
In this article we generalize the HGLM information criteria, D and ERL, to PH frailty models. Compared with classical random-effect models such as HGLMs, inference in semi-parametric frailty models is complicated by censoring and by the presence of a nonparametric baseline hazard function.

2 PH Frailty Models

Let T_{ij} $(i = 1, \ldots, q,\ j = 1, \ldots, n_i,\ n = \sum_i n_i)$ be the survival time for jth observation on the ith subject and C_{ij} be the corresponding censoring time. Then the observable data become $s_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = I(T_{ij} \leq C_{ij})$, where $I(\cdot)$ is the indicator function. Denote by U_i the unobserved frailty random variable for the ith subject. Let $V_i = \log U_i$ and $v_i = \log u_i$. A frailty model is described as follows. Given $V_i = v_i$, the conditional hazard function of T_{ij} is of the form:

$$\lambda_{ij}(t_{ij}|v_i) = \lambda_0(t_{ij}) \exp(x_{ij}^T \beta + v_i) \tag{1}$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard and β is a $p \times 1$ vector of fixed effects associated with fixed covariates $x_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$. The log-frailties, V_i $i = 1. \ldots q$, are assumed to be independent and identically distributed random variables having a frailty parameter α. Even though the results of this paper can be applied to non-normal frailties, for simplicity of argument we employ the normal distribution for V_i, which is useful for modelling correlated frailties.
To illustrate, we consider one-component model $X\beta + Zv$, which can be easily extended to a multi-component model as follows:

$$X\beta + Z_1 v^{(1)} + Z_2 v^{(2)} + \cdots + Z_k v^{(k)} \tag{2}$$

X is the $n \times p$ model matrix, Z_r $(r = 1, 2, \ldots, k)$ are the $n \times q_r$ model matrices corresponding to the $q_r \times 1$ frailties $v^{(r)}$, and $v^{(r)}$ and $v^{(l)}$ are independent for $r \neq l$. Let $Z = (Z_1, Z_2, \ldots, Z_k)$, $v = (v^{(1)T}, v^{(2)T}, \ldots, v^{(k)T})^T$, $\alpha = (\alpha_1, \ldots, \alpha_k)^T$, and $q = \sum_r q_r$. We use α and ρ to represent dispersion parameters in the frailty distribution. Then the multi-component model can be substituted directly into (1) leading to a straightforward extension.

3 *h*-likelihood Inference & ERL

Let $\ell = \ell(\beta, \theta)$ be a likelihood, either an h-likelihood, h, or a marginal likelihood, $m = \log\left\{\int \exp(h)\ dv\right\}$, with nuisance parameters θ. Lee and

Nelder (2001) considered a function $p_\theta(\ell)$, defined by

$$p_\theta(\ell) = [\ell - \frac{1}{2}\log\det\{A(\ell,\theta)/(2\pi)\}]|_{\theta=\widehat{\theta}} \tag{3}$$

where $A(\ell,\theta) = -\partial^2\ell/\partial\theta^2$ and $\widehat{\theta}$ solves $\partial\ell/\partial\theta = 0$. The function $p_\theta(\cdot)$ produces an adjusted profile likelihood, eliminating nuisance effects θ, which can be fixed effects β or random effects v or both.
In general, $p_\beta(m) \simeq r$, the restricted likelihood (REML) to the first order (Cox & Reid, 1987), $p_v(h)$ is the first-order Laplace approximation to m (i.e. $p_v(h) \simeq m$) and $p_{\beta,v}(h) \simeq p_\beta(m)$ (Lee & Nelder, 2001). In principle, one should use the h-likelihood, h, for inferences about v; the marginal-likelihood, m, for β; and the restricted likelihood, $p_\beta(m)$, for the dispersion parameters. When m is numerically difficult to obtain, we may use $p_v(h)$ and $p_{\beta,v}(h)$ as approximations to m and $p_\beta(m)$, respectively.

Lee & Nelder (2003) called $p_{\beta,v}(h)$ ($\simeq r$) the extended restricted likelihood (ERL) - the extension being to the (REML-based) elimination of the random effects, v.

4 Deviances, Extensions & Information Criteria

Nelder and Lee (1996) suggested the deviance of goodness of fit, D, as a model selection criteria, namely

$$D = D(y,\widehat{\mu}\,) \quad = \quad -2\{\ell_1(\,\widehat{\mu}\,;y|v) - \ell_1(y;y|v)\} \tag{4}$$

in the HGLM class of random effect models. It may be shown (Ha et al, 2005) that in general PH survival models with parametric frailty give rise to the same likelihood as Poisson HGLMs, so that (4) remains a vehicle for model selection in the frailty setting.
The deviance T_d, based upon the ERL, is given by

$$T_d = -2\{p_{\beta,v}(h)\} \tag{5}$$

is also a candidate model selection criterion (Lee & Nelder, 2001) and, notably, it is focussed solely on the variance components, the fixed effects and random effects having been eliminated by the REML adjustment.
As they stand these are not information criteria since we have not made any formal adjustment for nuisance parameter elimination. Moreover, we have not yet discussed the elimination of the nuisance function $\lambda_0(t)$. Nevertheless, the idea of comparing the performance of these unfocussed and focussed model selection criteria begins to intrude.
It may be shown that the two deviances can be extended to PH survival frailty models via the following information criteria:

$$\mathrm{AIC}(D^*) = D + 2p_D \tag{6}$$

where the estimated degrees of freedom, d.f. $= N - p_D$, $N = \sum_k \sum_{ij} I\{(i,j) \in R(y_{(k)})\}$ is the number of observations $y_{ij,k}$ for the equivalent Poisson HGLM, and $p_D = \text{trace}(H^{-1}H^*)$ where: $H = A(h, \theta)$, $H^* = A(\ell_1, \theta)$ and $\theta = (\lambda_0^T, \beta^T, v^T)^T$.
Similarly, the focussed model selection criterion for the dispersion parameters can be extended first to a new deviance:

$$T_d^* = -2\{p_{\beta,v}(h^*)\} \tag{7}$$

and then to the focussed information criterion:

$$\text{AIC}(T_d^*) = T_d^* + 2p_T \tag{8}$$

Notice that T_d^* is the ERL based on $p_{\beta,v}(h^*)$ which eliminates β, v from h^*, the profile h-likelihood from which the nuisance function $\lambda_0(t)$ has already been eliminated - a strategy proposed by Ha, Lee & Song (2001) and Ha & Lee (2005). Since T_d^* is by construction the ERL for the dispersion parameters, p_T turns out to be the number of dispersion parameters or variance components in the model.

5 Results

5.1 Mammary Tumour Data

We re-analyze the data of Gail et al. (1980) on multiple occurrences of mammary tumours for 48 female rats. The observations are the times to the development of a mammary tumour for 23 female rats in the treatment group and 25 female rats in the control group. Initially, 76 rats were injected with a carcinogen, and each rat was treated with retinyl acetate for the next 60 days. Some 48 rats were tumour-free after 60 days. These rats were randomly assigned to continued retinoid prophylaxis or to the control group, where they received no treatment. The main objective of the study was to evaluate treatment.
The survival time from the initial carcinogen injection, T_{ij} $(j = 1, \ldots, n_i)$ is then calculated as $t_{i,j} - t_{i,j-1}$, where $t_{i,j}$ with $t_{i,0} = 0$ is the jth tumour occurrence time of the ith rat, i.e., the gap time between tumour recurrences. Survival times on the same rat may be correlated due to shared genetic or environmental effects and this correlation can be modelled by a shared frailty. Moreover, the frailty of each rat may not be constant, but may vary stochastically over the gap times. Following Yau and McGilchrist (1998), we consider AR(1) frailty models for such dependency. Here we model a single fixed covariate x_{ij} (= 1 for treatment and = 0 for control) in the following five models, $\lambda_{ij}(t_{ij}|v) = \lambda_0(t_{ij}) \exp(\eta_{ij})$ where η_{ij} allows for the covariate and/or the frailty structures in models below:

TABLE 1. Deviance results for the mammary tumor data.

Model	T_d^*	p_T	AIC(T_d^*)	D	p_D	AIC(D)
M1 (Cox)	1946.8	0	7.0	1204.8	58	28.0
M2 (R)	1939.1	1	1.3	1147.4	78.8	12.2
M3 (AR(1)*)	1946.7	1	8.9	1204.8	58.0	28.0
M4 (AR(1))	1935.8	2	0	1080.4	106.2	0
M5 (R+AR(1))	1935.8	3	2.0	1080.4	106.2	0.0

R, rat frailty; AR(1), AR(1) frailty; AR(1)*, AR(1) frailty with $\rho = 0$; $T_d^* = -2\{p_{\beta,v}(h^*)\}$; p_T, the number of frailty parameters; D, deviance; $p_D = \text{trace}(H^{-1}H^*)$; AIC, Akaike information criterion differences where the smallest AIC is adjusted to be zero.

M1 (Cox): $\eta_{ij} = x_{ij}\beta$,

M2 (R): $\eta_{ij} = x_{ij}\beta + v_i$ with $v_i \sim N(0, \alpha_1)$,

M3 (AR(1)*): $\eta_{ij} = x_{ij}\beta + e_{ij}$ with $e_{ij} \sim N(0, \alpha_2)$,

M4 (AR(1)): $\eta_{ij} = x_{ij}\beta + v_{ij}$ with $v_{ij} \sim$ AR(1),

M5 (R+AR(1)): $\eta_{ij} = x_{ij}\beta + v_i + v_{ij}$ with $v_i \sim N(0, \alpha_1)$ and $v_{ij} \sim$ AR(1).

Here $v_{ij} \sim AR(1)$ means that $v_{ij} = \rho v_{ij-1} + e_{ij}$, $e_{ij} \sim N(0, \alpha_2)$ and $|\rho| < 1$.

In this paper we select the model which has the smallest AIC value among these models. For ease of comparison and ranking of candidate models, we have set the smallest value to be zero. In Table 1 we report the AIC differences, not the AIC values themselves.
Overall, the results in Table 1 suggest that p_D may not reflect model complexity properly when the variance of the frailties is near zero. Thus, for model selection related to (α, ρ) we should prefer AIC(T_d^*). If the AIC difference is larger than 1$\sim$2 it is considered to be significant, and if the difference is less than 1 it is not. Using this criterion, AIC(T_d^*) selects M4 as the final model, while the AIC(D) cannot select between M4 and M5.

TABLE 2. Simulation Results: percentage correct selection.

Simulation	Criterion	True model SM1	SM2	Average
Case 1: $(q, n_i) = (80, 5)$	AIC(T_d^*)	94	59	76.5
	AIC(D)	10	99	54.5
Case 2: $(q, n_i) = (20, 20)$	AIC(T_d^*)	93	98	95.5
	AIC(D)	46	100	73.0

5.2 Simulation Study

Here we report the results of a small simulation study based on 100 replications to evaluate the performance of the two AICs proposed above. We consider two non-nested models:

$$\text{SM1: } \eta_{ij} = x_{ij}\beta + v_i \text{ with } v_i \sim N(0, \alpha_1),$$
$$\text{SM2: } \eta_{ij} = x_{ij}\beta + v_{ij} \text{ with } v_{ij} \sim \text{AR}(1).$$

In order to generate the data from SM1 and SM2, we used exponential distributions for both survival and censoring times, with a censoring rate of 20%. Here we set $\lambda_0(t) = 1.0$ and $\beta = -1.0$, x_{ij} to 0 for the first $q/2$ subjects, to form the control group, and x_{ij} to 1 for the remaining $q/2$, to form the treatment group. We also set $\alpha_1 = \alpha_2 = 0.5$ and $\rho = 0.7$. We anticipated that the distinction between SM1 and SM2 might be difficult to detect with small n_i. Accordingly, we used a sample size of $n = \sum_{i=1}^{q} n_i = 400$ in two scenarios $(q, n_i) = (80, 5)$ and $(q, n_i) = (20, 20)$. From the 100 replications we computed the two AICs. Table 2 shows the rate of selection of the true model among 100 replications. The results clearly favour the focussed information criterion.

6 Final Remarks

Two new information criteria for model selection in non-nested frailty models have been defined. One criterion is based on a general deviance approach while the other is based on a focussed deviance approach; where the focus is on the variance components which define the frailty model structure. The latter relies on ERL inference in which REML-like arguments are used eliminate the nuisance parameters - fixed effects, random effects and arbitrary baseline hazard components - from the h-likelihood. The use of the criteria have been illustrated in the analysis of a well known multivariate data set

and by means of a small simulation study. Overall the results favour the use of the focussed criterion based on ERL inference.
It will be recognized that we are only in the early stages of developing ERL-based information criteria for frailty model selection. The idea of nuisance parameter elimination is, of course, not new; but the novelty of this paper lies in its sustained use of ERL arguments to produce, for the first time, a focussed information criterion applicable to a wide class of statistical models.

References

Akaike H. (1973). Information theory and an extension of maximum likelihood principle, In *Proceedings 2nd International Symposium on Information Theory*, B. N. Petrov and F. Csáki (eds). Akadémiai Kiadó: Budapest, pp. 267-281.

Cox DR. and Reid N. (1987). Parameter orthogonality and approximate conditional inference (with discussion), *Journal of the Royal Statistical Society B* vol. 49 pp. 1-39.

Gail M., Santner TJ. and Brown CC. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor, *Biometrics* vol. 36 pp. 255-266.

Ha, I.D., Lee, Y. and Song, J-K. (2001). Hierarchical likelihood approach for frailty models, *Biometrika* vol. 88 pp. 233-243.

Ha ID. and Lee Y. (2003). Estimating frailty models via Poisson hierarchical generalized linear models, *Journal of Computational Graphical Statistics*, vol. 12 pp. 663-681.

Ha ID. and Lee Y. (2005). Comparison of hierarchical likelihood versus orthodox best linear unbiased predictor approaches for frailty models, *Biometrika*, vol. 92 pp. 717-723.

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

Lee Y. and Nelder, J.A. (2001). Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions, *Biometrika* vol. 88 pp. 987-1006.

Lee Y. and Nelder JA. (2003). Extended-REML estimators, *Journal of Applied Statistics* vol. 30 pp. 845-856.

Sastry N. (1997). A nested frailty model for survival data, with application to study of child survival in Northeast Brazil. *Journal of the American Statistical Association* vol. 92 pp. 426-435.

Spiegelhalter DJ., Best N., Carlin BP. and van der Linde A. (2002). Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society B* vol. 64 pp. 583-639.

Yau KKW. and McGilchrist CA. (1998). ML and REML estimation in survival analysis with time dependent correlated frailty, *Statistics in Medicine* vol. 17 pp. 1201-1213.

Yau KKW. (2001). Multilevel models for survival analysis with random effects *Biometrics* vol. 57 pp. 96-102.