Non-PH Multivariate Survival Models Based on the GTDL

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

[1] Centre for Medical Statistics, Keele University, Keele, Staffordshire ST5 5BG, UK. E-mail: g.mackenzie@keele.ac.uk

[2] Faculty of Information Science, Kyungsan University, Kyungsan, 712-240, South Korea. E-mail: idha@kyungsan.ac.kr

[3] Department of Statistics, Seoul National University, Seoul, 151-742, South Korea. Email: youngjo@plaza.snu.ac.kr

Abstract

Correlated survival times may be modelled by introducing a random effect, or frailty, component into the hazard function. For multivariate survival data we extend a non-PH model, the generalized time-dependent logistic (GTDL) survival model (MacKenzie, 1996, 1997), to include random effects. The extension leads to two different, but related, non-PH models according to the method of incorporating the random effects. The h-likelihood procedures of Ha, Lee and Song (2001) and Ha and Lee (2003), which obviate the need for marginalization (over the random effect distribution), are derived for these extended models and their properties discussed. The new models are used to analyze two practical examples in the survival literature and the results are compared with those obtained from fitting the PH and PH frailty models.

Keywords: Frailty models; Generalized time-dependent logistic; Hierarchical-likelihood; Non-PH model; Random effect

1 Introduction

Proportion hazards (PH) frailty models which extend the standard PH model (Cox 1972) to allow frailty are frequently to analyze multivariate survival data which may arise, for example, when recurrent or multiple event times on the same subject. However, the assumption of proportionality can sometimes be inappropriate.

In this paper we introduce a flexible non-PH random-effect model based on the generalized time-dependent logistic (GTDL) survival model (MacKenzie, 1996). The GTDL generalizes the relative risk (RR) in Cox's PH model to time-dependent form. The model, a wholly parametric competitor for the PH model, has several interesting properties including a frailty interpretation. In particular, by retaining Cox's constant of proportionality as the leading term in the RR, the model is not only capable of representing data which conform the PH assumption, but can also accommodate a wider class of survival data in which the assumption of proportionality is untenable.
We, extend the GTDL to the multivariate survival data setting in two ways, adopt the hierarchical likelihood (h-likelihood) approach of Ha, Lee and Song (2001) and Ha and Lee (2003) for inference, use the new models to analyze two well known practical data sets from the literature and compare the results with the PH and PH frailty models.

2 The GTDL Model

A non-PH model, the GTDL regression model (MacKenzie, 1996), is defined by the hazard function:

$$\lambda(t;x) = \lambda_0 p(t;x), \tag{1}$$

where $\lambda_0 > 0$ is a scalar, $p(t;x) = \exp(t\alpha + x^T\beta)/\{1 + \exp(t\alpha + x^T\beta)\}$ is a linear logistic function in time, α is a scalar measuring the effect of time and β is a $p \times 1$ vector of regression parameters associated with fixed covariates $x = (x_1, \ldots, x_p)^T$. The relative risk (RR), the ratio of hazard rates for two subjects with different values of covariates, $x^{(1)}$ and $x^{(2)}$, is given by

$$\gamma(t; x^{(1)}, x^{(2)}) = \lambda(t; x^{(1)})/\lambda(t; x^{(2)}) = \exp\{(x^{(1)} - x^{(2)})^T\beta\}\psi(t, x^{(1)}, x^{(2)}), \tag{2}$$

where $\psi(t, x^{(1)}, x^{(2)}) = \{1+\exp(t\alpha+x^{(2)T}\beta)\}/\{1+\exp(t\alpha+x^{(1)T}\beta)\}$. The leading term on the right hand side of (2), Cox's constant RR over time, is thus moderated by $\psi(\cdot)$, a function of both time and covariates. That is, the model (1) is a non-PH, but when $\alpha = 0$ resulting RR is time invariant and model is then PH. The cumulative hazard function is given by

$$\Lambda(t;x) = \int_0^t \lambda(s;x)\ ds = \frac{\lambda_0}{\alpha}\log\left\{\frac{1+\exp(t\alpha + x^T\beta)}{1+\exp(x^T\beta)}\right\}. \tag{3}$$

Under non-informative censoring the ordinary censored-data likelihood, which depends on (1) and (3), is easily constructed.

3 Extended GTDL models

The multivariate data structures are as follows. Let T_{ij} $(i = 1, \ldots, q,\ j = 1, \ldots, n_i,\ n = \sum_i n_i)$ be the survival time for jth observation of the ith subject. Denote by U_i the unobserved frailty (or random effect) for the ith subject.

We extend the model (1) to include a frailty term acting multiplicatively on the individual hazard rate. Given $U_i = u_i$, the conditional hazard function of T_{ij} takes the form

$$\lambda_{1ij}(t|u_i) = \lambda_{ij}(t)u_i, \tag{4}$$

The frailties U_i are assumed to be independent and identically distributed random variables with a density function depending on the frailty parameter θ, say $g(.|\theta)$. Alternatively, we may consider another natural extension of model (1), by including a random component in the linear predictor, $t\alpha + x^T\beta$, of (1). Given $U_i = u_i$, the conditional hazard function of T_{ij} is then of the form

$$\lambda_{2ij}(t|u_i) = \lambda_0 \frac{\exp(t_{ij}\alpha + x_{ij}^T\beta + u_i)}{1 + \exp(t_{ij}\alpha + x_{ij}^T\beta + u_i)}\ . \tag{5}$$

where the U_i have been defined above.

Models (4) and (5) are similar, but (4) assumes that the random effects act multiplicatively on the hazard function while (5) assumes they are additive on a generalized $\log_e$-odds scale, which is the usual $\log_e$-odds scale when $\lambda_o = 1$. While (4) is a conventional frailty model, (5) is not, although it is nevertheless of interest, since then the random effects and the fixed effects act linearly on the same scale.

The choice of $g(.|\theta)$ may be important. For h-likelihood inference, the choice of parametric form is wide (and testable), since marginalization is not required. In this paper we shall adopt the log-Normal distribution for h-likelihood inference - a choice to which inference on β is robust (Ha et al., 2001; Ha and Lee, 2003). Perhaps a more natural choice for Model (4) is the Gamma distribution, see Blagojevic, MacKenzie & Ha (2003) for a marginal approach. Alternatively, we may adopt a non-parametric mixture model.

4 h-Likelihood Estimation & Inference

Let the observable random variables be $Y_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = I(T_{ij} \leq C_{ij})$, where C_{ij} is the censoring time corresponding to T_{ij} and $I(\cdot)$ is the indicator function.

Following Ha, Lee and Song (2001), the h-likelihood for the model (4), denoted by h, is defined by

$$h = h(\alpha, \beta, \theta) = \sum_{ij} \ell_{1ij} + \sum_i \ell_{2i}, \tag{6}$$

where

$$\ell_{1ij} = \ell_{1ij}(\alpha, \beta; y_{ij}, \delta_{ij}|u_i) = \delta_{ij} \log \lambda_1(y_{ij}|u_i) - \Lambda_1(y_{ij}|u_i)$$

is the logarithm of the conditional density function for Y_{ij} and δ_{ij} given $U_i = u_i$, and $\ell_{2i} = \ell_{2i}(\theta; v_i)$ is the logarithm of the density function for $V_i = v(U_i) = \log(U_i)$ with parameter θ. Here v is scale on which the random effects influence the linear predictor and $v_i = v(u_i) = \log u_i$: see also Lee and Nelder (1996). The maximum h-likelihood (MHL) estimating equations of $\tau = (\alpha, \beta^T, v^T)^T$ with $v = (v_1, \ldots, v_q)^T$ are given by

$$\partial h/\partial \tau = 0. \tag{7}$$

Note that the asymptotic covariance matrix (Ha et al., 2001) for $\widehat{\tau} - \tau$ is given by the inverse of $H = -\partial^2 h/\partial \tau^2$. For the estimation of the frailty parameter θ, we use Lee and Nelder's (1996) APHL (adjusted profile h-likelihood) h_P of θ after eliminating τ, defined by

$$h_P = h_A|_{\tau=\widehat{\tau}}\,, \tag{8}$$

where $h_A = h + \frac{1}{2}\log\{\det(2\pi H^{-1})\}$. Given estimates of τ, Lee and Nelder's (2001) REML (restricted maximum likelihood) estimating equation for θ, maximizing h_P, is given by

$$\partial h_A/\partial \alpha|_{\tau=\widehat{\tau}} = 0. \tag{9}$$

5 Results

We illustrate the use of models (4) and (5) and also their conditional forms (without frailty or random effects) and include Cox's PH and PH frailty models as comparators.

We analyze two sets of well-known multivariate survival data which have appeared in the literature. Firstly, the kidney infection data of McGilchrist and Aisbett (1991), comprising times to the first and second recurrences of infection in 38 kidney patients and consider a single fixed covariate, sex of the patients, coded 1 for female and 0 for male. Secondly, the placebo-controlled randomized trial of gamma interferon (γ–IFN) in chronic granulomatous disease (CGD) (Fleming and Harrington, 1991) in which scientific interest is focused on the effect of treatment on the (possibly multiple) recurrence times. In all, we analyze ten covariates including treatment. The results of the analyzes are shown in Tables 1 and 2 respectively (omitted). For the kidney data, the finding that femaleness is protective of recurrence is confirmed in all of the models fitted. The standard error is elevated in all frailty models suggesting that Cox Model and the non-PH models without frailty fail to account properly for the (positive) correlation between recurrence times. The α parameter in the non-PH models is not significant,

suggesting that there is no serious departure from the PH assumption in these data.
The results for the CGD data are broadly similar in that the treatment effect is correctly identified by all models fitted. However, in these data, there is clear evidence of non-proportionality $\alpha \neq 0$, but the size of this effect is small. On the other hand, there is some difference in interpretation of the longitudinal covariate, which is identified by all of the non-PH models.
More details will appear in the main paper.

References

Blagojevic M., MacKenzie G. and Ha I.D. (2003) A Comparison of non-PH & PH - Gamma frailty models. *IWSM 2003, submitted*.

Cox, D. R. (1972) Regression models and life-tables (with discussion). *J. R. Statist. Soc.* B, **34**, 187-220.

Fleming, T. R. and Harrington, D. P. (1991) *Counting processes and survival analysis*. New York: Wiley.

Ha, I. D. and Lee, Y. (2003) Estimating frailty models via Poisson hierarchical generalized linear models. *Journal of Computational and Graphical Statistics* (in press).

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

Lee, Y. and Nelder, J. A. (1996) Hierarchical generalized linear models (with discussion). *J. R. Statist. Soc.* B, 58, 619-678.

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

MacKenzie, G. (1996) Regression models for survival data: the generalised time dependent logistic family. *J. R. Statist. Soc.* D, 45, 21-34.

MacKenzie, G. (1997) On a non-proportional hazards regression model for repeated medical random counts. *Statist. Med.*, 16, 1831-1843.

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