PH and Non-PH Frailty Models for Multivariate Survival Data

M. Blagojevic[1] and G. MacKenzie [1]

[1] Keele, University, Centre for Medical Statistics, Keele, Staffordshire ST5 5BG, UK

Abstract: We generalize the previously developed Non-PH CTDL-Gamma and the PH Weibull-Gamma frailty models to correlated survival data. In particular, we seek analytical results using the marginal approach, to determine whether the univariate results generalize to the multivariate context. We consider both the shared and correlated frailty cases. We also develop non-parametric frailty models which enable us to check the appropriateness of the assumed distributional form of the random effect.

Keywords: Frailty Models; Non-PH; EM; Finite Mixtures; Correlated Survival Model.

1 Introduction

In our earlier work, we have extended the Weibull proportional hazards (PH) regression survival model to a Gamma frailty model by means of a multiplicative random effect acting on the hazard function (Hougaard 1984). However, not all survival data are PH and it is therefore useful to have alternative non-PH models. This is relevant as, increasingly, random effect models are being used to analyze multivariate survival data (Ha 2001).

A flexible non-PH model is the Canonical Time-Dependent Logistic (CTDL) described by MacKenzie (1996) and later by MacKenzie(1997). We have already generalized this model by including a multiplicative Gamma frailty term in the hazard function. The resulting frailty model was obtained in closed form and we compared its properties with the Weibull frailty model, noting the connection with a general class of frailty models described by Aalen (1988). The performance of the four models, Weibull and CTDL with and without frailty, was investigated using data from the N. Ireland lung cancer study and it was shown that the CTDL-Gamma model provided the best fit. In addition, non-parametric frailty models are developed to check whether Gamma distribution is appropriate for the random effect. We now extend the models to multivariate case and special consideration is given to correlated frailty scenario.

2 Univariate Survival Models

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

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

where $\lambda > 0$ is a scalar, $p(t|x) = \exp(t\alpha + x'\beta)/\{1 + \exp(t\alpha + x'\beta)\}$ is a linear logistic function in time, α is a scalar measuring the effect of time, β is a $p \times 1$ vector of regression parameters associated with fixed covariates $x' = (x_1, \ldots, x_p)$ and $\theta' =(\lambda, \alpha, \beta)$.
When developing the CTDL-gamma mixture model, we assumed that the random component has a multiplicative effect on the hazard, such that $\lambda(t; x, u) = u\lambda(t; x)$. U follows a Gamma distribution with $E(U) = 1$ and $V(U) = \sigma^2$. We then used the marginalization approach to obtain the pdf for the resulting marginal frailty distribution:

$$f_f(t|x) = \frac{\lambda p_i}{\left\{1 - \frac{\lambda\sigma^2}{\alpha}\log_{\mathrm{e}}(g_i q_i)\right\}^{1+\frac{1}{\sigma^2}}} \tag{2}$$

where,

$$\begin{aligned} p_i &= \exp(t_i\alpha + x_i'\beta)/\{1 + \exp(t_i\alpha + x_i'\beta)\} \\ q_i &= 1/\{1 + \exp(t_i\alpha + x_i'\beta)\} \\ g_i &= 1 + \exp(x_i'\beta) \end{aligned} \tag{3}$$

and where, for notational convenience, we have suppressed the dependence on time and the covariates on the LHS of (3).

Similarly for Weibull-gamma model, we found that:

$$f_f(t|x) = \frac{\lambda^\rho \rho e^{x'\beta} t^{\rho-1}}{\left\{1 + \sigma^2(\lambda t)^\rho e^{x'\beta}\right\}^{1+\frac{1}{\sigma^2}}} \tag{4}$$

3 Non-Parametric Frailty

The estimated effect of covariates may be influenced (to a varying degree in different sets of data) by the choice of the distributional form of the frailty density. In order to minimize the impact of frailty distribution assumption, we fit a non-parametric (NP) frailty component based on a finite mixture. We use the EM algorithm for implementation. We are interested in estimating the NP frailty component simultaneously with the mixing proportions. These estimated values will typically suggest the mixture from which the data were generated and hence will provide a useful check on any parametric assumptions made.

The resulting CTDL and Weibull log-likelihoods respectively are:

$$\ell_{ctdl}(\pi,\theta) = \sum_{j=1}^{c}\sum_{i=1}^{n}\left\{ z_{ij}\log_{\mathrm{e}}\pi_j + z_{ij}[\delta_i \log_{\mathrm{e}}(u_j p_i) + \frac{u_j\lambda}{\alpha}\log_{\mathrm{e}}(q_i g_i)]\right\} \quad (5)$$

where, π_j is the jth component of the mixture, c is the dimension of the mixture, $u_j = e^{c^{ij}\gamma}$ is the non-parametric frailty component, θ is the vector of parameters to be estimated, p_i, q_i, g_i are as before.

and

$$\ell_w(\pi,\theta) = \sum_{j=1}^{c}\sum_{i=1}^{n}\left\{ z_{ij}\log_{\mathrm{e}}\pi_j + z_{ij}[\delta_i \log_{\mathrm{e}}(\lambda^\rho \rho t_i^{\rho-1} e^{x_i^{'}\beta} u_j) - (\lambda t_i)^\rho e^{x_i^{'}\beta} u_j]\right\} \quad (6)$$

An algorithm was written in S-Plus (V4.5) to maximize (5) and (6).

4 Multivariate Survival Data

We turn now to the idea of generalizing the parametric frailty models introduced earlier to correlated survival data. In particular, we seek analytical results using the marginal approach, in order to determine whether the univariate results generalize to the multivariate context.

Suppose we have $f(t_i|u_i,\theta)$ and $g(u_i|\sigma^2)$ where $t_i = (t_{i1}, t_{i2}, \cdots, t_{i_{m_i}})$ is the vector of survival times on the ith subject. m_i is the number of measurements on the ith subject, whence t_{ij}, $i = 1, \cdots, n$; $j = 1, \cdots, m_i$, become our data. The joint likelihood of t and u is then:

$$L(\theta,\sigma^2) = \prod_{i=1}^{n} f(t_i|u_i,\theta) g(u_i|\sigma^2) \quad (7)$$

However, under the h-likelihood assumption that the survival times within a subject are independent given the random effect:

$$f(t_i|u_i,\theta) = \prod_{j=1}^{m_i} f(t_{ij}|u_i,\theta) \quad (8)$$

then, after marginalizing over u and assuming non-informative censoring (7) becomes:

$$L(\theta,\sigma^2) = \prod_{i=1}^{n}\int_0^\infty g(u_i|\sigma^2)\prod_{j=1}^{m_i}[\lambda(t_{ij}|u_i,\theta)]^{\delta_i} S(t_{ij}|u_i,\theta)du_i \quad (9)$$

4.1 Bivariate Models

Let us consider a bivariate case with $m_i = 2$ so that there are two survival times measured on each subject. After some algebra, we obtain our two models:

Bivariate Weibull-Gamma model:

$$f_f(t_{i1}, t_{i2}|\theta) = \frac{(1+\sigma^2)(\lambda^\rho \rho e^{x'\beta})^2 \lambda^2 (t_{i1}t_{i2})^{\rho-1}}{(1+\sigma^2\lambda^\rho[t_{i1}^\rho + t_{i2}^\rho]e^{x'\beta})^{2+\frac{1}{\sigma^2}}} \tag{10}$$

and Bivariate CTDL-Gamma model:

$$f_f(t_{i1}, t_{i2}|\theta) = \frac{(1+\sigma^2)\lambda^2 p_{i1} p_{i2}}{\left\{1 - \frac{\lambda\sigma^2}{\alpha}\log_e(g_i^2 q_{i1} q_{i2})\right\}^{2+\frac{1}{\sigma^2}}} \tag{11}$$

Therefore, contrary to some claims that have been made previously, we have been able to use marginalization approach to obtain the bivariate CTDL-Gamma model. We should note that models (10) and (11) are proportional to their corresponding univariate forms, and can easily be extended to higher dimensional data.

4.2 Correlated Gamma frailty

In the previous section, we assumed shared frailty when dealing with bivariate survival data. However, this assumption may not always be plausible, and hence we should perhaps like each of the two survival times measured on an individual to have its own frailty component associated with it. A case that is of a particular interest is when the two frailty components follow Gamma distributions which are correlated. A substantial amount of research has been done in this field, mainly by Yashin, e.g. (Yashin 1995), especially when dealing with twin data, but the thrust of this work is wholly in relation to PH models. No attention has been given to the case where a non-PH hazard is assumed.

Let the two frailties be constructed as $U_1 = Y_0 + Y_1$ and $U_2 = Y_0 + Y_2$, where Y_i are independent Gamma random variables with parameters (k_i, θ_i), $i = 0, 1, 2$. Let us further suppose that $V[U_1] = \sigma_1^2$, $V[U_2] = \sigma_2^2$ and $corr[U_1, U_2] = \rho_u$. We force U_1 and U_2 to be Gamma distributed by assuming $\theta_0 = \theta_1 = \theta_2$. We also retain the earlier assumption of conditional independence of survival times.

For the CTDL model, we have, after some algebra:

$$S(t_1, t_2) = \int_0^\infty \int_0^\infty \int_0^\infty (g_i q_{i1})^{\frac{\lambda}{\alpha}(y_0+y_1)} (g_i q_{i2})^{\frac{\lambda}{\alpha}(y_0+y_2)} g(y_0) g(y_1) g(y_2) dy_0 dy_1 dy_2 \tag{12}$$

$$= [1 + \theta\Lambda(t_1)]^{-k_1}[1 + \theta\Lambda(t_2)]^{-k_2}[1 + \theta\Lambda(t_1) + \theta\Lambda(t_2)]^{-k_0} \quad (13)$$

where $g(y_i)$ are probability density functions of the random variables Y_i, $i = 0, 1, 2$, respectively. A similar form is obtained for the Weibull distribution. A simulation study was performed, the results will appear elsewhere.

5 Final Remarks

In this paper we have extended the non-PH based Gamma frailty and its standard PH-based Gamma frailty competitor to bivariate case. The models we obtained when the frailties are correlated are of a more general form than those standardly used, since we do not assume identical distribution of Y_i, $i = 0, 1, 2$.

Our development of multivariate parametric versions of the non-PH frailty model to deal with correlated survival data and our investigation of correlated frailties opens up further interesting avenues of research. The development of this class of models and various non-parametric, finite mixture, competitors is also being pursued.

Key References

Aalen O.O. (1988) Heterogeneity in Survival Analysis. *Statistics in Medicine*, 7, 1121-1137 .

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

Hougaard, P. (1984). Life table methods for heterogeneous populations : Distributions describing the heterogeneity. *Biometrika*, 71, 75-83.

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

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

Yashin, A. I. et al (1995) Correlated Individual Frailty: An Advantageous Approach to Survival Analysis of Bivariate Data. *Mathematical Population Studies*, 5, 145-159.