Additive Survival Models with Shared Frailty

Giovani L. Silva
CEAUL & DMIST - Universidade Técnica de Lisboa
gsilva@math.ist.utl.pt

M. Antónia Amaral-Turkman
CEAUL & DEIO - Universidade de Lisboa
maturkman@fc.ul.pt

Frailty models have been proposed in order to investigate other sources of variation when the observed covariates do not fully explain the dissimilarities of the individuals in study. The frailty term can be partitioned into two or more terms in order to assess various types of frailty within the same individual. For instance, the frailty associated with a person may be divided into two random effects describing separately genetic and environmental factors, which are actually shared with other people such as mother, father, etc. The aim is to present a Bayesian analysis of additive survival models with shared or correlated frailty terms. An analysis of the adoption data described by Sørensen et al. (1988) motivates and illustrates the frailty models developed, using Markov Chain Monte Carlo methods for estimating quantities of interest.

Keywords: Shared frailty model, Additive hazards model, Survival analysis. Bayesian analysis, MCMC methods.

1 Introduction

In regression analysis for survival data, as the observed covariates are not fully explain the variation from individual to individual, a random effect (frailty) is included into the hazard function to take account that unobserved heterogeneity, e.g., genetic predisposition within families. In addition, the frailty can be partitioned into two or more terms in order to assess various types of frailty within the same individual. For example, the frailty of a person may be divided into two random effects describing separately genetic and environmental factors, which are shared with other people such as mother, father, etc.

Shared or correlated frailty models are herein analyzed for additive survival models (Aalen, 1980) from a Bayesian perspective (Silva and Amaral-Turkman, 2004). The additive hazards models have been presented both as a diagnostic tool and as a useful alternative to multiplicative hazards models, especially when the hazard functions are not proportional.

This work is organized as follows. Section 2 describes Aalen's additive model based on counting processes, as well as an additive frailty model with shared and correlated frailty terms. Section 3 deals with the Bayesian analysis of the additive frailty model by using Markov chain Monte Carlo (MCMC) methods for estimating quantities of interest. In section 4, we illustrate the methodology introduced here through the analysis of the adoption data described by Sørensen et al. (1988).

2 A SHARED FRAILTY ADDITIVE MODEL

Aalen (1980) introduced an additive survival model defining the intensity of a counting process $N(t)$ - number of occurrences of a particular event up to time t - as

$$I(t|\mathbf{z}) = Y(t)\left(\alpha_0(t) + \sum_{q=1}^{p} \alpha_q(t) z_q\right), \tag{1}$$

where $Y(t)$ indicates whether the individual is in risk at time t, $\alpha_0(t)$ is the baseline intensity for individuals, and $\alpha_q(t)$ is the regression function that may reveal changes in the influence of the covariate z_q over time, $q = 1, \ldots, p$.

In order to account for the unobserved heterogeneity, a random effect (w) is introduced into the intensity (1) additively (Rocha, 1996). Silva and Amaral-Turkman (2004) proposed a Bayesian approach for that new model that is therein so-called *additive frailty model*. Note that $\alpha_0(t)$ in that new intensity is interpreted as the baseline intensity for individuals with "null" frailty ($w = 0$).

The frailty term w for each individual may be partitioned into two or more terms, e.g., $w = w_1 + \cdots + w_k$, where w_j are (correlated) frailty terms shared with other individuals, $j = 1, \ldots, k$. For genetic setting, the frailty of a child may be associated with genes shared with mother and father (see Figure 1).

Assuming a multivariate counting process $\mathbf{N}(t) = (N_1(t), \ldots, N_n(t))$ for n right-censored individuals (under a history $\mathcal{F}_{t-}$), shared frailty additive models are here defined by intensity function of $N_i(t)$, i.e.,

$$I_i(t|\mathbf{z}_i, \mathbf{w}) = Y_i(t)\left(\alpha_0(t) + \sum_{q=1}^{p} \alpha_q(t) z_{iq} + \mathbf{a}_i' \mathbf{w}\right), \tag{2}$$

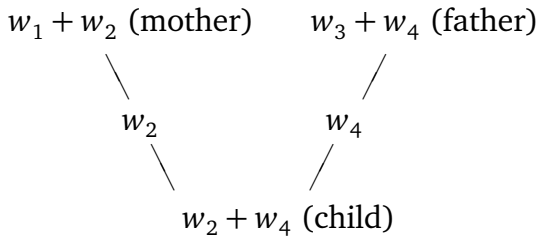

Figure 1: Graph of a frailty model for genetic data.

where $\mathbf{w} = (w_1, \ldots, w_k)'$ is the frailty vector and $\mathbf{a}_i = (a_{i1}, \ldots, a_{ik})'$ is the vector of frailty indicator functions for the i-th individual, $i = 1, \ldots, n$ (Silva and Amaral-Turkman, 2004). Petersen (1998) also showed a version of the model (2) for multiplicative frailty intensities.

3 AN BAYESIAN APPROACH OF THE CURRENT MODEL

Partitioning the time axis into m disjoint intervals $B_j = [t_{j-1}, t_j)$, $j = 1, \ldots, m$, independent gamma prior processes are assumed for the increments of the cumulative functions $\Omega_q(t) = \int_0^\infty \alpha_q(u)du$, i.e., the increment $\Omega_{qj} \equiv d\Omega_q(t)$ in B_j has gamma distribution with shape and scale parameters $c_q\Omega^*_{qj}$ and c_q, $j = 1, \ldots, m$, $q = 0, \ldots, p$. Notice that Ω^*_{qj} is interpreted as a prior guess of Ω_{qj} with degree of precision c_q.

Let $\mathscr{D} = \{(N_i(t), Y_i(t), \mathbf{z}_i)\}$ be the survival data with n right-censored individuals. Assigning independent gamma priors for $\Omega_q(t)$, the posterior of the frailty model (2), denoted by $\pi(\Omega, \mathbf{w}, \delta|\mathscr{D})$, is proportional to

$$\prod_{j=1}^{m}\left[\prod_{i=1}^{n}\left(I_{ij}^{N_{ij}}\, e^{-I_{ij}}\right)\prod_{q=0}^{p}\left(\Omega_{qj}^{c_q\Omega^*_{qj}-1} e^{-c_q\Omega_{qj}}\right)\right]\tau(\mathbf{w}|\delta)\,\tau(\delta), \qquad (3)$$

where $I_{ij} \equiv I_i(t_j)dt_j = Y_{ij}(t)(\mathbf{z}_i'\Omega_j + \mathbf{a}_i'\mathbf{w}\,dt_j)$, $N_{ij} \equiv dN_i(t_j)$, $\Omega = (\Omega_{11}, \ldots, \Omega_{pm})'$, $\Omega^*_{qj} = r_q\,dt_j$, r_q is a proposed value for $\alpha_q(t)$, $dt_j = t_j - t_{j-1}$, $\tau(\mathbf{w}|\delta)$ is the frailty distribution and $\tau(\delta)$ is a prior for hyperparameter δ.

The frailty distribution is traditionally gamma with hyperparameter δ, which measures the degree of unobserved heterogeneity through, e.g., via its standard deviation (σ_W). The posterior (3) is awkward to work with, since the marginal posterior distributions of Ω and δ are not easy to obtain explicitly. Nevertheless, these posteriors can be evaluated using Markov chain Monte Carlo (MCMC) methods (Spiegelhalter *et al.*, 2007).

4 ILLUSTRATION

Using the model (3) for the adoption data (Sørensen et al., 1988) with 125 families 1924-1987, the intensities of death by infection (e.g., pneumonia) for biological mother, son and adoptive mother are, respectively,

$$\begin{aligned} I_{i1}(t|\mathbf{w}) &= Y_{i1}(t)[\alpha_{01}(t) + w_{i1} + w_{i2}] \\ I_{i2}(t|\mathbf{w}) &= Y_{i2}(t)[\alpha_{02}(t) + w_{i1} + w_{i3} + w_{i4}] \\ I_{i3}(t|\mathbf{w}) &= Y_{i3}(t)[\alpha_{03}(t) + w_{i3} + w_{i5}]. \end{aligned} \tag{4}$$

For simplicity, the posterior (3) is here associated with gamma frailties $(1, \delta_l)$, non-informative priors for δ_l, 65 intervals B_j's, $c_q = 0.001$ and $r_q = 0.1$, $q = 1, 2, 3$, $l = 1, \ldots, 5$. After 6000 iterations simulated, including 1000 for burn-in period, some quantities of interest were estimated for the shared additive frailty model (4).

parameter	mean	s.d.	CI(2.5%)	CI (97.5%)
σ_{G_s}	0.022	0.0021	0.0178	0.0261
σ_{E_s}	0.023	0.0023	0.0191	0.0281
$\sigma^2_{E_s}/\sigma^2_{G_s}$	1.203	0.3337	0.6772	1.9830
$\sigma^2_{EG_{ns}}/\sigma^2_{G_s}$	0.895	0.2633	0.4796	1.4970
$\sigma^2_{EG_{ns}}/(\sigma^2_{G_s} + \sigma^2_{E_s})$	0.407	0.1058	0.2397	0.6564

Table 1: Estimates of variance components for frailties.

The estimates in Table 1 indicate little unobserved heterogeneity both shared genetic ($\widehat{\sigma}_{G_s}$) and environment ($\widehat{\sigma}_{E_s}$) factors, shared environment factors explain 20.3% more of the variability than the shared genes, while non-shared effects have 10.5% less importance than genes.

ACKNOWLEDGEMENTS

Special Thanks to Thorkild Sørensen for supplying the adoption data. This paper was partially supported by Pest-OE/MAT/UI0006/2011.

REFERENCES

[1] Aalen, O.O. (1980). A Model for Nonparametric Regression Analysis of Counting Processes. *Lectures Notes in Statistics*, 2, 1-12.

[2] Petersen, J.H. (1998). An additive frailty model for correlated life times. *Biometrics*, 54, 646-661.

[3] Rocha, C.S. (1996). Survival Models for Heterogeneity Using the Non-Central Chi-Squared Distribution with Zero Degrees of Freedom. In: *Lifetime Data: Models in Reliability and Survival Analysis* edited by Jewell, N.P.; Kimber, A.C.; Lee, M.T., and Whitmore, G.A.. 275-279, Kluwer Academic Publishers.

[4] Silva, G.L., and Amaral-Turkman, M.A. (2004). Bayesian analysis of an additive survival model with frailty. *Communications in Statistics – Theory and Methods*, 33, 2517-2533.

[5] Sørensen, T.I.A., Nielsen, G.G., Andersen, P.K. and Teasdale, T.W. (1988). Genetic and environmental influences on premature death in adult adoptees. *New England Journal of Medicine*, 312, 727-732.

[6] Spiegelhalter D., Thomas A., Best N.G. and Lunn D. (2007) Bayesian inference using Gibbs sampling for Windows (WinBUGS), Version 1.4.3, MRC Biostatistics Unit, Cambrigde, http://www.mrc-bsu.cam.ac.uk/bugs/.