

CTDL-Positive Stable Frailty Model

M. Blagojevic¹, G. MacKenzie²

¹ Department of Mathematics, Keele University, Staffordshire ST5 5BG, UK and
² Centre of Biostatistics, University of Limerick, Ireland

Abstract: The non-PH Canonical Time Dependent Logistic (CTDL) survival regression model is extended by incorporating a positive stable frailty component into the hazard function within the Bayesian framework. The resulting model is compared numerically with the Weibull-positive stable frailty model, using data from a placebo controlled randomized trial of gamma interferon in chronic granulomatous disease (CGD). Moreover, supremum bounds of the ratio-of-uniforms (ROU) algorithm, used for sampling from complete conditional distributions, are obtained analytically thus yielding a more efficient form of the algorithm.

Keywords: Canonical Logistic, Frailty Models, Positive Stable, Non-PH, Ratio-of-Uniforms.

1 Introduction

A flexible non-PH model is the Canonical Time-Dependent Logistic (CTDL) model described by MacKenzie (1996, 1997). In our earlier work, the CTDL model was extended to univariate and multivariate gamma frailty models within the frequentist framework using a marginal Likelihood approach and its properties compared with the Weibull-Gamma models analytically and numerically. It was revealed via an extensive simulation study that the Weibull and Weibull-gamma models gave more precise results, in terms of standard errors, than the CTDL and CTDL-gamma models. However, analysis of real data revealed that the CTDL based models provided superior fits to the data.

It has been shown (Qiou et al., 1999) that allowance is made for a higher degree of heterogeneity among subjects by using infinite variance frailty distributions than would be possible using finite variance frailty, such as gamma. A positive stable frailty term has previously been mixed with PH models, but never with non-PH basic models. Therefore, it was timely to develop a such a new model for the multivariate shared frailty setting using the CTDL hazard as the basic model and comparing its performance with that of the Weibull-positive stable frailty model. Inference was carried out in a Bayesian framework, and bounds of components necessary for the ratio-of-uniforms (ROU) algorithm for sampling from complete conditional
distributions were derived successfully, thus improving the efficiency of the algorithm.

2 Positive stable frailty models

Let the random variable \( U \) denote unobservable individual frailties. Buckle (1995) gives the joint density of \( n \) iid 4-parameter stable distributed rv's by using the joint pdf of \( U_i \) and \( Y_i \), \( f(u_i, y|\omega) \), from which the marginal density of \( U_i \) turns out to be the stable pdf.

\[
f(u_i|\omega) = \frac{\omega|u_i|^{1/(\omega-1)}}{|\omega| - 1} \int_{-1/2}^{1/2} \exp \left( -\left| \frac{u_i}{\tau_\omega(y)} \right|^{\omega/(\omega-1)} \right) \frac{1}{\tau_\omega(y)}^{\omega/(\omega-1)} dy \quad (1)
\]

where

\[
\tau_\omega(y) = \sin(\pi \omega y + \psi_\omega) \cos(\pi/2 + \delta_\omega \omega \cos(y)) / \cos(\pi y)
\]

\( \omega \in (0, 1) \), \( y \in (-1/2, 1/2) \) and \( \psi_\omega = \min(\omega, 2 - \omega)\pi/2 \).

The observed data for the \( j \)th time observation for the \( i \)th individual (or alternatively for the \( j \)th individual in the \( i \)th group) is \((t_{ij}, \delta_{ij}, x_{ij})\). Let \( D_{\text{obs}} \) denote all such triplets. The unobserved data are the frailties \( u = (u_1, ..., u_n) \). So the complete data is \( D = (D_{\text{obs}}, u) \). Note that \( u \) is based on a vector of auxiliary variables \( y = (y_1, ..., y_n) \) in equation (1). So given the data \( D_{\text{obs}} \) and the parameters of interest, a likelihood and prior for parameters are needed so that a posterior density may be obtained.

2.1 CTDL-positive stable frailty model

A non-PH CTDL regression model is defined by the hazard function

\[
\lambda(t|x) = \lambda \exp(t\alpha + x'\beta) / \{1 + \exp(t\alpha + x'\beta)\} \quad (2)
\]

where \( \lambda > 0 \) is a scalar, \( \alpha \) is a scalar measuring the effect of time and \( \beta \) is a \( p \times 1 \) vector of regression parameters associated with fixed covariates \( x' = (x_1, ..., x_p) \).

The corresponding survival function is

\[
S(t|x) = \{(1 + \exp(t\alpha + x'\beta)) / (1 + \exp(x'\beta))\}^{-1/\beta} \quad (3)
\]

So the complete data likelihood is given by:

\[
L(\lambda, \alpha, \beta, \omega|D) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} [\lambda(t_{ij}|\lambda, \alpha, \beta, u_i)]^{\delta_{ij}} S(t_{ij}|\lambda, \alpha, \beta, u_i)
\]
\begin{align}
\prod_{i=1}^{m} \prod_{j=1}^{m_i} \left\{ u_i \lambda \exp(t_{ij} \alpha + x'_{ij} \beta) \right\}^{\delta_{ij}} \left\{ \frac{1 + \exp(t_{ij} \alpha + x'_{ij} \beta)}{1 + \exp(x'_{ij} \beta)} \right\}^{-\frac{u_i \lambda}{\alpha}}
\end{align}

The observed data likelihood, which is simply the marginal model once the frailty components have been integrated out, is:

\begin{align}
L(\lambda, \alpha, \beta, \omega | D_{obs}) &= \prod_{i=1}^{n} \int_{m_i} \prod_{j=1}^{m_i} \left\{ u_i \lambda \exp(t_{ij} \alpha + x'_{ij} \beta) \right\}^{\delta_{ij}} \left\{ \frac{1 + \exp(t_{ij} \alpha + x'_{ij} \beta)}{1 + \exp(x'_{ij} \beta)} \right\}^{-\frac{u_i \lambda}{\alpha}} \\
&\times \frac{\omega |u_i|^{1/(\omega-1)}}{|\omega - 1|} \int_{-1/2}^{1/2} \exp \left[ -\frac{u_i}{\tau_{\omega}(y_i)} |\omega/(\omega-1)| \right] \\
&\times \frac{1}{\tau_{\omega}(y_i)} |\omega/(\omega-1)| dy_i du_i
\end{align}

The posterior density, expressed in terms of the observed data likelihood and the joint prior for the parameters is:

\begin{align}
\pi(\lambda, \alpha, \beta, \omega | D_{obs}) \propto L(\lambda, \alpha, \beta, \omega | D_{obs}) p(\lambda)p(\alpha)p(\beta)p(\omega)
\end{align}

where \(\pi(\cdot)\) denotes the posterior and \(p(\cdot)\) the prior distribution. Note that independence among all parameters is assumed.

The integrals in equation (5) do not have a closed form, so instead the unknown parameter vector \((\lambda, \alpha, \beta, \omega)\) is augmented with vectors \(u\) and \(y\) and MCMC methods are used to obtain samples for \((\lambda, \alpha, \beta, \omega, u, y)\).

Complete conditional distributions are needed for \(\lambda, \alpha, \beta, \omega, u\) and \(y\) from which the corresponding samples are to be drawn. They are derived as being proportional to (6) and only the components involving the parameter of interest are retained.

The choice of priors is given in table 1. All resulting complete conditional distributions are of non-standard forms and we use rejection algorithm for generation of \(y\). Metropolis-Hastings algorithm with a beta proposal density in the case of \(\omega\) and ratio of uniforms (ROU) algorithm in all other instances. We have developed a more efficient way of implementing the ROU algorithm, namely deriving its components analytically instead of performing numerical bisection. In order to demonstrate the new approach clearly, we outline briefly the ROU algorithm, exemplified by generation of \(\alpha\).

Suppose \(V\) and \(W\) are real variables that are uniform on \(A = \{(v, w) : 0 < v < \sqrt{f(w)}\}\), where \(f(\cdot)\) is the complete conditional distribution of interest, then the variable \(\alpha = \frac{W}{V}\) has the pdf that is proportional to \(f(\alpha)\). The result is most useful when the set \(A\) is contained in a rectangle \([0, a] \times [b_1, b_2]\), in which case a type of rejection sampling can be used:

Step 1: Generate \(V_1, V_2 \sim U(0, 1)\).

Step 2: Let \(V = aV_1\) and \(W = b_1 + (b_2 - b_1)V_2\), then \((V, W)\) is a point
randomly chosen in the rectangle \([0, a] \times [b_1, b_2]\).

Step 3: If \((V, W) \in A\), then accept \(\alpha = W V\), otherwise repeat from step 1.

\(A \in [0, a] \times [b_1, b_2]\) where

\[
a = \sup_{\alpha} \sqrt{f(\alpha)}
\]

\[
b_1 = -\sqrt{\sup_{\alpha \leq 0}(\alpha^2 f(\alpha))}
\]

\[
b_2 = \sqrt{\sup_{\alpha \geq 0}(\alpha^2 f(\alpha))}
\]

Hence, step 3 above may be re-written as:

Step 3: Accept \(\alpha = W V\) if and only if \(V^2 \leq f(W V)\).

Derivation of quantities \(a, b_1\) and \(b_2\) is necessary, and it has been claimed by various authors that this cannot be done analytically. However, we show that it is in fact possible to obtain their bounds, details will appear elsewhere.

2.2 Weibull-positive stable frailty model

The familiar Weibull regression distribution has the following hazard and survival function

\[
\lambda(t|x) = \lambda \rho (t \rho)^{\rho-1} \exp(x' \beta)
\]

\[
S(t|x) = \exp(- (t \rho)^{\rho} e^{x' \beta})
\]

So the complete data likelihood is given by:

\[
L(\lambda, \rho, \beta, \omega|D) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} \left\{ u_i \lambda \rho (\lambda t_{ij})^{\rho-1} e^{x_{ij}' \beta} \right\}^{\delta_{ij}} \exp(- u_i(\lambda t_{ij})^{\rho} e^{x_{ij}' \beta})
\]

The observed data likelihood is then:

\[
L(\lambda, \rho, \beta, \omega|D_{obs}) = \prod_{i=1}^{n} \int \prod_{j=1}^{m_i} \left\{ u_i \lambda \rho (\lambda t_{ij})^{\rho-1} e^{x_{ij}' \beta} \right\}^{\delta_{ij}} \exp(- u_i(\lambda t_{ij})^{\rho} e^{x_{ij}' \beta})
\]

\[
\times \frac{\omega|u_i|^{1/(\omega-1)}}{|\omega - 1|} \int_{-1/2}^{1/2} \exp \left[- \frac{u_i}{\tau_\omega(y_i)} \right]^{\omega/(\omega-1)}
\]

\[
\times \frac{1}{\tau_\omega(y_i)} |\omega/(\omega-1)| dy_i du_i
\]

The posterior density is:

\[
\pi(\lambda, \rho, \beta, \omega|D_{obs}) \propto L(\lambda, \rho, \beta, \omega|D_{obs}) p(\lambda)p(\rho)p(\beta)p(\omega)
\]

Again, the integrals in (10) do not have a closed form, so MCMC methods are used to obtain samples for \((\lambda, \rho, \beta, \omega, u, y)\) as in the previous section. The choice of priors is given in table 1. It should be mentioned that only
the prior for $\lambda$ is conjugate with the likelihood, giving a standard gamma complete conditional distribution. All other complete conditional distributions are of non-standard forms and ROU sampling algorithm is employed except in the instances for $y$ and $\omega$ as in the previous section. Last section only mentioned $\omega$ as a clear exception.

3 Example Data Analysis

Consider now an application of the models outlined above to a data set from a placebo controlled randomized trial of gamma interferon in chronic granulotomous disease (CGD). The CGD study is described in detail in a report by the International CGD Cooperative Study Group (1991). The treatment was given to each of the 128 patients at the first scheduled visit for that patient. The data for each patient give the time to first and any recurrent serious infections, from study entry until the first scheduled visit of the patient. There is a minimum of 1 record per patient, with an additional record for each serious infection occurring up to the study completion date. For bivariate data, only information pertaining to record(s) 1, 2 and 3 is needed from which the gap times as well as censoring indicators are calculated. Only one factor, gender, is included in the analysis.

Prior and hyperparameter specifications for the two models are given in table 1. Gibbs sampler is used to generate samples from the derived complete conditional distributions; S-Plus (V4.5) was used for programming. 5000 iterations were taken as "burn-in" and a further 5000 iterations taken for inference purposes. Table 2 gives Gelman and Rubin convergence statistic for the parameters of the two models; successful convergence is indicated for all parameters.

Posterior distributions are summarized in terms of means and standard errors of each parameter in the CTDL and Weibull models with positive stable frailty in table 3.

Gender is statistically significant in both models, and its negative estimated effect means that females are at a lower risk of being infected. Note that all other parameters are statistically significant too.

Values $\omega = 0.751$ in the CTDL case and $\omega = 0.622$ in the Weibull case correspond to a reasonable degree of dependence between times of each patient (note that the value of 1(0) implies maximum dependence(independence)), more so in the CTDL case.

4 Remarks and Future work

So we have shown that the CTDL-positive stable frailty model confirmed a higher degree of dependence between individual observations than the Weibull-positive stable frailty model. Sensitivity to prior specifications for models considered in this paper will be the subject of future work. We have
already developed the CTDL and Weibull models with gamma distributed frailty in Bayesian framework and a comparison of these with their corresponding "frequentist" counterparts will be the subject of future work.

Key References


TABLE 1. Prior and Hyperparameter specifications

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior</th>
<th>Hyperparameter specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_c$</td>
<td>Gamma($\gamma$, $\gamma$)</td>
<td>$\gamma = 0.001$</td>
</tr>
<tr>
<td>$\lambda_w$</td>
<td>Gamma($\mu$, $\mu$)</td>
<td>$\mu = 0.001$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Normal($\xi$, $\nu$)</td>
<td>$\xi = 0, \nu = 1000$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Normal($\epsilon$, $m$)</td>
<td>$\epsilon = 0, m = 1000$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$p(\omega) = 1$</td>
<td>$0 &lt; \omega &lt; 1$</td>
</tr>
</tbody>
</table>

NB: $\hat{\lambda}_c = \lambda$ in the CTDL, $\hat{\lambda}_w = \lambda$ in the Weibull.

TABLE 2. Gelman and Rubin statistics of Model Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CTDL + Positive Stable Frailty</th>
<th>Gelman and Rubin (97.5% quantile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_c$</td>
<td>1.00(1.00)</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1.02(1.02)</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.00(1.00)</td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>1.00(1.00)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Weibull + Positive Stable Frailty</th>
<th>Gelman and Rubin (97.5% quantile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_w$</td>
<td>1.00(1.00)</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>1.06(1.10)</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.00(1.00)</td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>1.00(1.00)</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 3. Posterior Summary of Model Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CTDL + Positive Stable Frailty</th>
<th>Posterior Mean</th>
<th>Posterior S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_c$</td>
<td>0.236</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-0.097</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.098</td>
<td>0.039</td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.751</td>
<td>0.082</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Weibull + Positive Stable Frailty</th>
<th>Posterior Mean</th>
<th>Posterior S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_w$</td>
<td>0.058</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>1.342</td>
<td>0.321</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.073</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.622</td>
<td>0.066</td>
<td></td>
</tr>
</tbody>
</table>