Some new models for multivariate survival data

Sujit Sahu [a]

University of Southampton

http://www.maths.soton.ac.uk/staff/Sahu/

October 2001: University of Kent

[a] Joint work with Dipak Dey

Motivation

Want to

- have more flexible models.
- exploit advanced MCMC methods.
- learn more from the data.

Outline

- What are multivariate survival data?
- The general model and setup
  - Models for frailty
  - Modelling the baseline hazard
- Computations using the reversible jump
- Examples
  - Example 1: Rat litter data
  - Example 2: Kidney infection data
- Discussion

Multivariate survival data

- Survival times of family members.
- Recurrence times of diseases.
- Failure times of paired human organs.

Right censored data with covariates.

- Three rats from each of 50 litters were observed to develop tumor: A random rat was given a drug.
- Data on first and second recurrence of 38 kidney patients.
- Onset of blindness for diabetic retinopathy patients.
- Plenty of more.

The general model

Suppose that data

$$T_{ij}, j = 1, \ldots, m, \;\; i = 1, \ldots, n$$

is survival time of the jth subject in the ith group. Assume the Cox proportional hazard model conditional on the frailty.

$$h(t_{ij}|\mathbf{z}_{ij}, w_i) \;=\; w_i \; h_o(t_{ij}) \; \exp(\boldsymbol{\beta}^T \mathbf{z}_{ij})$$

w_i: the individual level effect is called *frailty*.

Hazard: Probability of instantaneous death given that it has survived upto that point.

Density = Hazard $\times$ Exp(– Cumulative Hazard)

i.e.

$$f(t) = h(t) \; \exp(-\int_0^t h(u)du)$$

Frailty Distributions

Gamma frailty:

$$f(w|\eta) = \frac{\eta^{\eta}}{\Gamma(\eta)} w^{\eta-1} \exp(-w\eta),\ w > 0.$$

Has mean 1, variance $1/\eta$; provides a conjugate prior; easy to use.

Has the Laplace transform

$$E\left[\exp(-sW)\right] = \left(\frac{\eta}{\eta + s}\right)^{\eta}.$$

Does not give proportional hazard in the marginal model.

Positive stable frailty

One specifies the Laplace transform.

$$E\left[\exp(-sW)\right] = \exp(-s^{\alpha}).$$

- It maintains the proportional hazard assumption in the marginal model.
- It is heavy tailed.
- Its density is not unique.

There are many possible solutions for the density of W. One version due to Buckle (1995) is $f(w|\alpha) =$

$$\frac{\alpha\, w^{1/(\alpha-1)}}{|\alpha-1|}\int_{-1/2}^{1/2}\exp\left\{-\left|\frac{w}{d_\alpha(y)}\right|^{\alpha/(\alpha-1)}\right\}$$

$$\left|\frac{1}{d_\alpha(y)}\right|^{\alpha/(\alpha-1)} dy$$

where

$$d_\alpha(y) = \frac{\sin(\pi\alpha y + s_\alpha)}{\cos \pi y}\times$$

$$\left[\frac{\cos \pi y}{\cos\{\pi(\alpha-1)y + s_\alpha\}}\right]^{(\alpha-1)/\alpha},$$

and $s_\alpha = \min(\alpha, 2-\alpha)\pi/2$.

Log-Skew-t frailty

We specify the density $f_b(b|\rho, \delta, \nu)$ of $b = \log(w)$.

$$2\left(\rho^2 + \delta^2\right)^{-1/2} f_{t;\nu}\left(\frac{b}{\sqrt{\rho^2 + \delta^2}}\right) \times$$

$$F_{t;\nu+1}\left(\sqrt{q(b)}\,\frac{\delta}{\rho}\,\frac{b}{\sqrt{\rho^2 + \delta^2}}\right)$$

where

$$q(b) = \frac{\nu + 1}{\nu + b^2/(\rho^2 + \delta^2)}$$

and f_t and F_t are the pdf and cdf of the standard t distribution with m df.

- With $\delta = 0$ the density reduces to the standard t density with ν df.
- If $\nu \to \infty$ then it approaches the normal distribution.
- It is Cauchy if $\delta = 0$ and $\nu = 1$.
- Non-zero values of δ provide skewness.
  - Positively skewed if $\delta > 0$.
  - Negatively skewed if $\delta < 0$.

Let us compare

Assume the bivariate setup, i.e. $m = 2$.

Tools for comparison:

- Local dependence measure, e.g. $\frac{S(t_1,t_2)}{S(t_1)S(t_2)}$
- Kendall's
  $\tau = E\text{sign}\{(T_{11} - T_{21})(T_{12} - T_{22})\}$
- corr(T_1, T_2)
- corr$(\log T_1, \log T_2)$

We use the correlation between $\log(T_1)$ and $\log(T_2)$. Following Hougaard (2000)

$$\text{corr}(\log T_1, \log T_2) = \frac{\text{Var}(b)}{\text{Var}(b) + \pi^2/6}.$$

- For gamma frailty: $E(b) = \psi(\eta) - \log(\eta)$ and Var$(b) = \psi'(\eta)$ where $\psi(\cdot)$ is the digamma function.

- For stable frailty $E(b) = -\left(\frac{1}{\alpha} - 1\right)\psi(1)$ and Var$(b) = \left(\frac{1}{\alpha^2} - 1\right)\frac{\pi^2}{6}$.

- For log-skew-t frailty: $E(b) = \left(\frac{\nu}{\pi}\right)^{1/2}\frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)}\delta$ and Var$(b) = (\rho^2 + \delta^2)\frac{\nu}{\nu-2} - \frac{\nu}{\pi}\left(\frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)}\right)^2\delta^2$.

- Equate the Var(b) under all three models and $E(b)$ under the last two models to obtain comparable densities.
- With this corr$(\log T_1, \log T_2)$ will be the same under all three models.

The expectation of b is not matched under the gamma and stable model since they are of opposite signs, always.

We thus obtain:

- For gamma frailty: $\eta = 1.14$.
- For stable $\alpha = 0.74$.
- For the log-skew-t model, we first choose, $\rho = 1$ and $\nu = 8$. Then $\delta = 0.23$.

Now corr$(\log T_1, \log T_2) = 0.45$ for all three models.

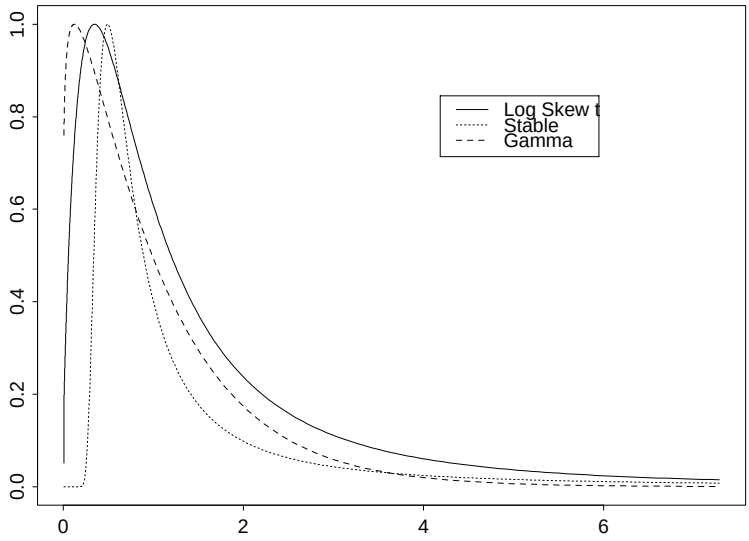


Figure 1: Rescaled densities of frailty distributions. The graphs are for (i) the log-skew-t distribution with $\rho = 1$, $\delta = 0.23$ and $\nu = 8$, (ii) positive stable with $\alpha = 0.74$, and (iii) the gamma distribution with $\eta = 1.14$.

- Three densities induce the same correlation on the log transformed survival times.
- The log-skew-t model has the heaviest tail.

Models for baseline hazard

Time is divided into g pre-specified intervals $I_k = (\tau_{k-1}, \tau_k]$ for $k = 1, 2, \ldots, g$ where

$$0 = \tau_0 < \tau_1 < \ldots < \tau_g < \infty,$$

τ_g being the last survival or censored time.

Assume that the baseline hazard is constant within each interval. That is,

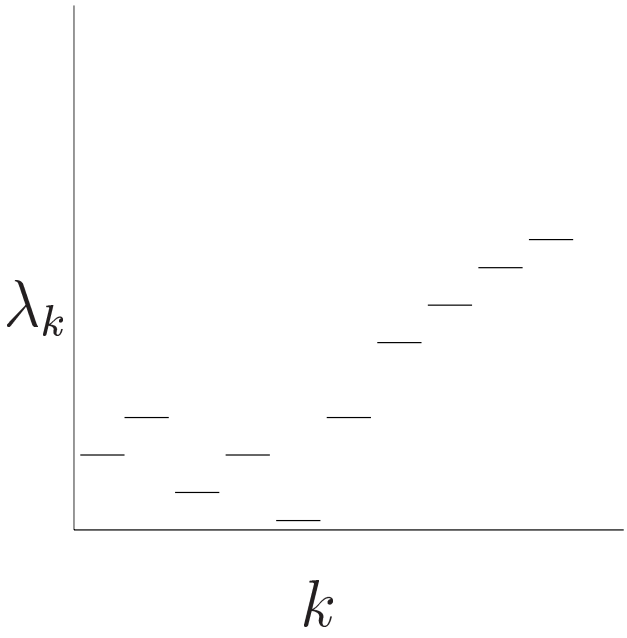
$$h_o(t_{ij}) = h_k, \text{ for } t_{ij} \in I_k.$$

Assume a martingale prior for $\lambda_k = \log(h_k)$, (see e.g. Sahu *et al.*, 1997),

$$\lambda_k | \lambda_1, \ldots, \lambda_{k-1} \sim N(\lambda_{k-1}, 1)$$

with $\lambda_0 = 0$.

A baseline hazard function



We further assume that the jump times $\tau_1, \tau_2, \ldots$ form a time-homogeneous Poisson process with rate a.

Advantage: The number and positions of the grid points need not be fixed in advance.

Given g, the martingale specification for λ_k implies that

$$\boldsymbol{\lambda} \sim N_g\left(\mathbf{0},\, C^{-1}\right)$$

where the $g \times g$ matrix C has all elements zero except for $c_{kk} = 2, k = 1, \ldots, g$ and $c_{kk+1} = -1, k = 1, \ldots, g-1$ and $c_{k-1k} = -1$ for $k = 2, \ldots, g$.

Posterior

Likelihood:

$$\prod_{i=1}^{n}\prod_{j=1}^{m}\left\{\prod_{k=1}^{g_{ij}}\exp(-h_k\Delta_k\theta_{ij}w_i)\right\}$$

$$\left(h_{g_{ij}+1}\theta_{ij}w_i\right)^{\delta_{ij}}\exp\left\{-h_{g_{ij}+1}(t_{ij}-\tau_{g_{ij}})\theta_{ij}w_i\right\},$$

where $t_{ij}\in(\tau_{g_{ij}},\tau_{g_{ij}+1}]=I_{g_{ij}+1}$ and
$\theta_{ij}=\exp(\boldsymbol{\beta}^T\mathbf{z}_{ij})$.

Prior:

$$\prod_{i=1}^{n}\left[\pi(w_i|\cdots)\right]\pi(\cdots)\,\pi(g)\pi(\boldsymbol{\lambda}|g)\,\pi(\boldsymbol{\beta}).$$

Prior Sensitivity

Aim is to keep the hyper-parameters as vague as possible.

- For η, and ρ^{-2} we use gamma(b, b) with $b = 0.001$
- For $\delta \sim U(0, 5)$: positively skewed.
- For ν we adopt exponential with mean 10.
- For g the number of jump times we assume Poisson with mean 10 truncated between 1 and 20.

Inference was largely insensitive to these choices.

Using Reversible jump

Three types of moves:

(a) updating all the parameters except for g,

(b) birth of a new jump time,

(c) death of an existing jump time.

The move type (a) does not change the dimension, details in Sahu *et al.* (1997).

How to jump from the current point x to a new point y of possibly different dimension?

Draw u and v of appropriate dimensions, such that

dimension of (x, u) = dimension of (y, v).

The point y is obtained using a one-to-one transformation $(y, v) = T(x, u)$ which is accepted with probability $\alpha\{(x, u), (y, v)\} =$

$$\min\left\{1, \frac{\pi(y)q_2(v)}{\pi(x)q_1(u)}\left|\frac{\partial T(x,u)}{\partial(x,u)}\right|\right\}$$

where $\pi(\cdot)$ is the posterior distribution and $q_1(u)$ and $q_2(v)$ are the densities of u and v, respectively.

Consider the birth move. We draw:

$$\tau^* \sim \text{Uniform}(\tau_0, \tau_{\max}).$$

Suppose that $\tau^* \in (\tau_{k-1}, \tau_k)$.

Need to generate λ'_k and λ'_{k+1} in the proposal when the current point is λ_k.

There is one degree of freedom of proposing the move. We simulate u uniformly in $(-\epsilon, \epsilon)$ for some $\epsilon > 0$.

Now λ'_k is taken to be a convex combination of λ_{k-1} and $\lambda_k + u$, and λ'_{k+1} is taken to be a convex combination of $\lambda_k - u$ and λ_{k+1}.

In particular we take

$$\begin{aligned}\lambda'_k &= \frac{\tau^* - \tau_{k-1}}{\tau_k - \tau_{k-1}}\lambda_{k-1} + \frac{\tau_k - \tau^*}{\tau_k - \tau_{k-1}}(\lambda_k + u),\\ \lambda'_{k+1} &= \frac{\tau^* - \tau_{k-1}}{\tau_k - \tau_{k-1}}(\lambda_k - u) + \frac{\tau_k - \tau^*}{\tau_k - \tau_{k-1}}\lambda_{k+1}.\end{aligned}$$

Now we set $\lambda'_i = \lambda_i$ for $1 \leq i \leq k-1$ and $\lambda'_{i+2} = \lambda_{i+1}$ for $k \leq i \leq g-1$. Further, we let $\tau'_k = \tau^*$ and set $\tau'_i = \tau_i$ for $0 \leq i \leq k-1$ and $\tau'_{i+1} = \tau_i$ for $k \leq i \leq g$.

Other proposals can be considered as well.

Observe that the Jacobian for this type of move is

$$\frac{2\,\Delta'_k\,\Delta'_{k+1}}{\Delta_k^2}.$$

where $\Delta_k = \tau_k - \tau_{k-1}$.

The density $q_1(u)$ in the acceptance probability is $\frac{1}{2\epsilon} I(-\epsilon < u < \epsilon)$ and generation of v is not required.

Model Choice and Validation

$$\pi(y_r|\mathbf{y}_{obs}) = \int \pi(y_r|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y}_{obs})d\boldsymbol{\theta}.$$

Suppose

$$\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(B)} \sim \pi(\boldsymbol{\theta}|\mathbf{y}).$$

- Draw $y_r^{(j)}$ from $\pi(y_r|\boldsymbol{\theta}^{(j)})$.
- It is a sample from the predictive density. Gelfand (1996).

Litters Example

Three rats from each of 50 litters were observed to develop tumor. A randomly selected rat from each litter was given a drug and the other two were selected as controls and were given a placebo.

β	φ	δ	ν
0.69 (0.31)	0.51 (0.45)	–	–
(0.09, 1.30)	(0.06, 1.69)	–	–
0.78 (0.32)	0.40 (0.004)	–	–
(0.13, 1.41)	(0.37, 0.42)	–	–
0.74 (0.31)	0.23 (0.23)	0.39 (0.27)	12.8 (9.2)
(0.13, 1.35)	(0.05, 0.89)	(0.02, 0.99)	(2.07, 37.1)

Parameter estimates from the gamma (first row), stable (second row) and log-skew-t (last row) model.

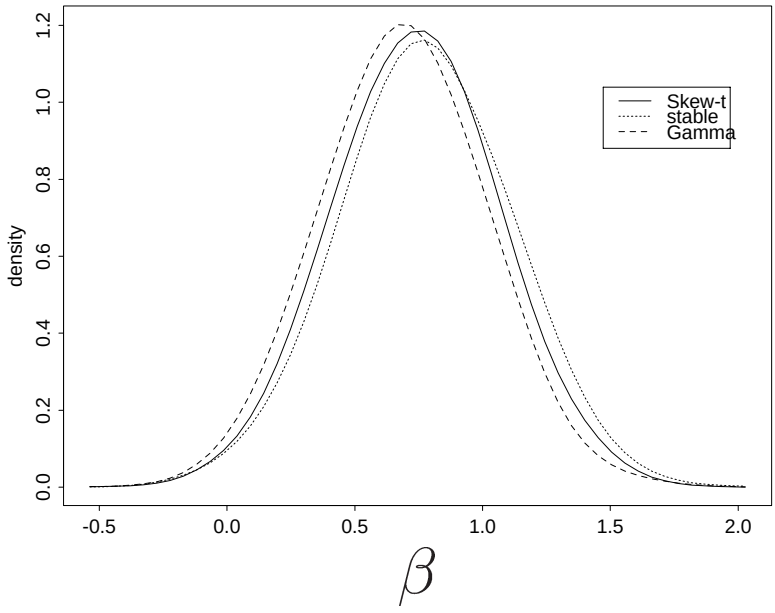


Figure 2: The kernel density estimates of β.

Diagnostic Checking

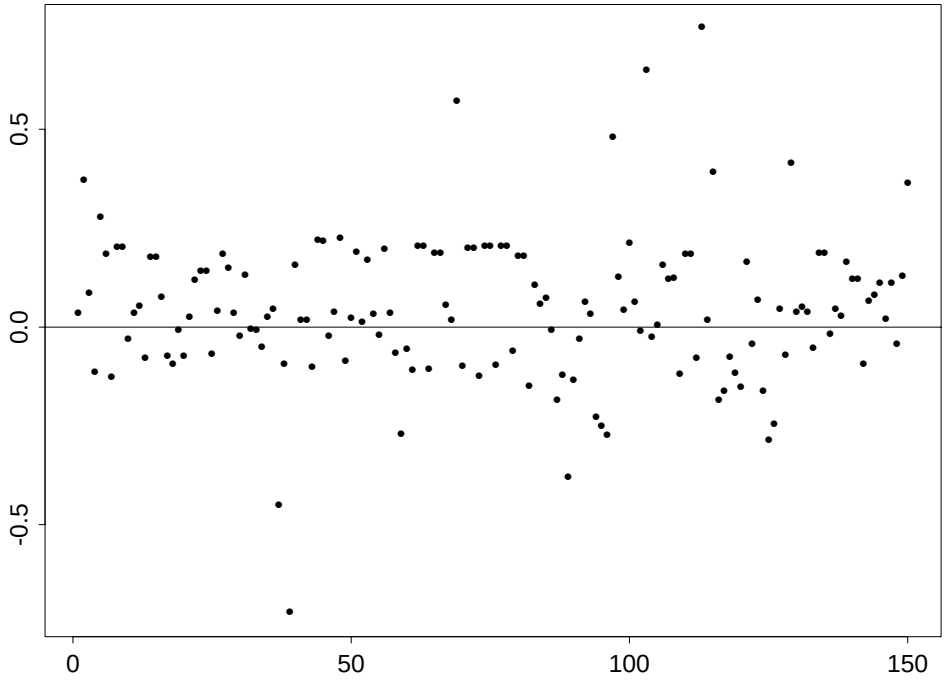


Figure 3: The CPO plot for comparing the log-skew-t model versus the stable model. 93 out of 150 observations support the skew model. The log pseudo-Bayes factor is 6.3.

Fitted Frailty Distributions

Recall the theoretical comparison made before.

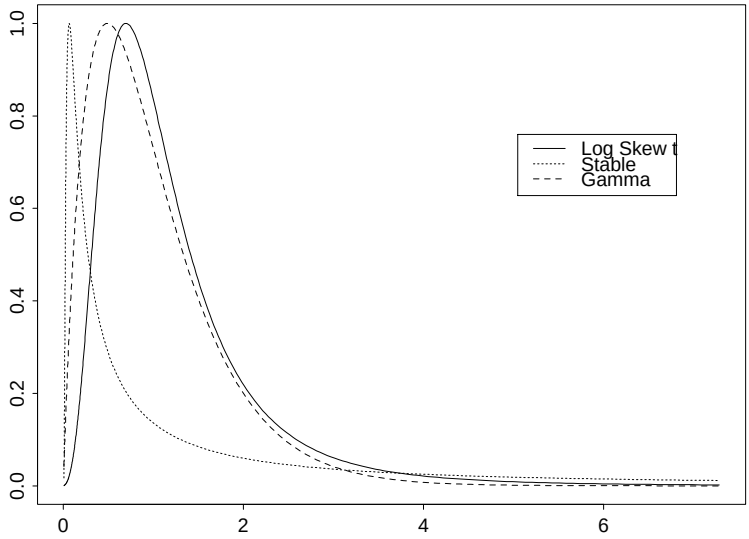


Figure 4: The fitted frailty distributions for the litters example.

- The tail of the log-skew-t is heavier than the gamma but lighter than the stable.

Kidney Infection Data

This is a well known data set from 38 kidney patients of which 28 are female. For each patient two recurrence times were recorded. Some were censored.

- Right censored data.
- Covariates: sex and age.
- Survival times: T_{ij}, jth recurrence time for the ith patient.

Sex has a 'significant' effect.

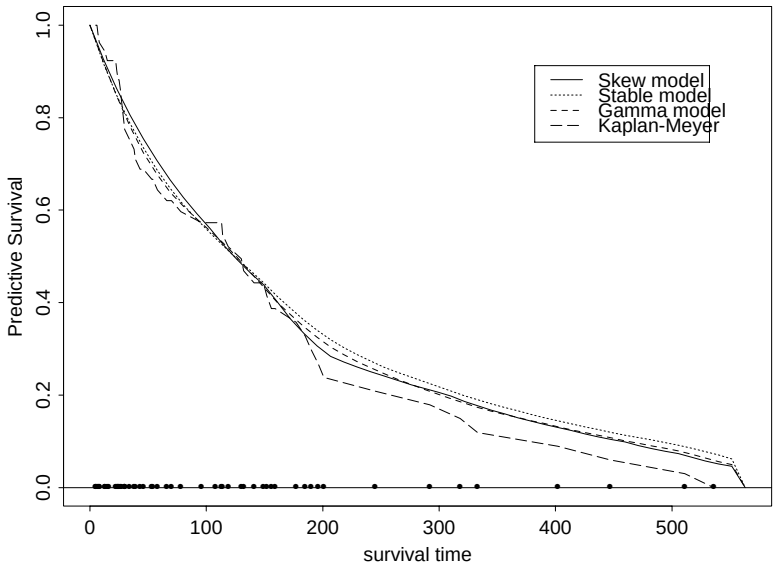


Figure 5: Predictive survival curves for a typical female patient for the kidney infection example.

- Most survival times are below 200.
- Around time=200 the Kaplan-Meyer estimate shows a sharp decrease in the survival function as the data suggests.
- The predictive survival function under the log-skew-t model adapts to this most rapidly. Other models follow suit, but at a slower pace.

Evaluate reversible jump

We get a smoother baseline hazard function.

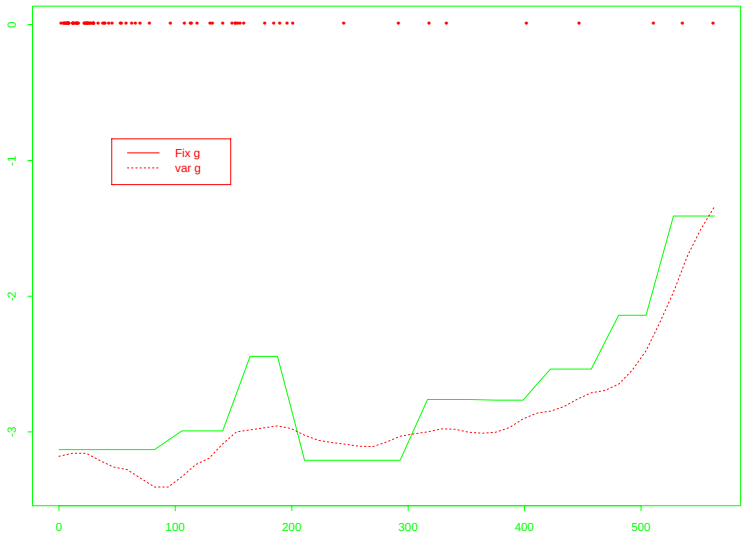


Figure 6: Posterior mean estimate of the log base line hazard function for fixed g and random g model for the kidney infection example.

The posterior distribution of g is not very sensitive under the three models or under different hyper parameter values.

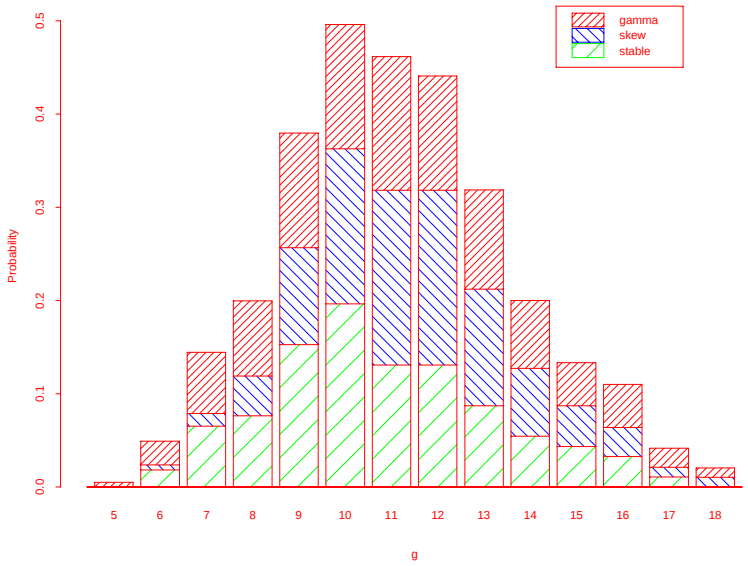


Figure 7: The posterior distribution of g under the three models for the litters example.

Discussion

Extended the multivariate survival models in two directions.

1. The log-skew-t frailty is more flexible than the gamma and stable frailty distributions.
   - It cannot take multi-modal shape.
2. The baseline hazard function is modeled using a flexible martingale process.
   - It imposes smoothing using its neighbors.
   - The jump times follow a time-homogeneous Poisson process.
   - This removes the ad-hoc assumptions on the number and position of the jump times.

Further Discussion

- We have developed the powerful reversible jump Markov chain Monte Carlo method for multivariate survival analysis.
- Our methods can be extended to many other scenarios including the models with time-dependent covariate effects.

The paper and the references are available from:

http://www.maths.soton.ac.uk/staff/Sahu/