Bayesian Methods for Survival and Longitudinal Data

By

Ming-Hui Chen

Department of Statistics, University of Connecticut

215 Glenbrook Road, U-4120, Storrs, CT 06269, USA

email: mhchen@stat.uconn.edu

Summary

Survival analysis arises in many fields of study such as medicine, biology, engineering, public health, epidemiology, and economics. This workshop aims to provide a comprehensive treatment of Semiparametric Bayesian methods for survival and longitudinal data. There will be four 2-hour sessions.

♠ Reference Book

J.G. Ibrahim, M.-H. Chen, and D. Sinha, *Bayesian Survival Analysis*, Springer-Velag, Second Printing, 2005.

Website: "`http://www.stat.uconn.edu/~mhchen/survbook`."

Outline

Session 1: Bayesian Semiparametric and Frailty Survival Models

Session 2: Bayesian Cure Rate Models

Session 3: Joint Models for Longitudinal and Survival Data

Session 4: Bayesian Model Assessment in Survival Analysis

Session 1: Bayesian Semiparametric and Frailty Survival Models

1.1 Introduction

1.2 Various Semiparametric Models

1.3 Frailty Models

Overview

In this session, we will start with a brief introduction and we will then discuss Bayesian models based on prior processes for the baseline hazard and cumulative hazard, construction of the likelihood function, and prior elicitation. Gamma processes, beta processes, correlated gamma processes, and frailty models will be introduced. Several examples and case studies will be presented to illustrate the various models.

Outline

- Bayesian models based on prior processes for the baseline hazard and cumulative hazard.
- Gamma processes
- Beta processes
- Correlated gamma processes
- Prior distributions
- Frailty models

1.1 Introduction

- These tutorial lectures focus on
  - Modelling
  - Priors
  - Computations
  - Applications
- Goals in these four sessions are to present
  - Modeling strategies and main ideas
  - Prior elicitation strategies
  - Advantages over other methods
  - MCMC implementation

1.2 Semiparametric Models

♠ Piecewise Constant Hazard Model

- Let $D = (n, \boldsymbol{y}, X, \nu)$ denote the observed data, where $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$, $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_n)'$ with $\nu_i = 1$ if the i^{th} subject failed and 0 otherwise, and X is the $n \times p$ matrix of covariates with i^{th} row $\boldsymbol{x}_i'$.
- One of the most convenient and popular models for semiparametric survival analysis is the piecewise constant hazard model.
- Discrete approximation to continuous time model.
- To construct this model, we first construct a finite partition of the time axis, $0 < s_1 < s_2 < \ldots < s_J$, with $s_J > y_i$ for all $i = 1, 2, \ldots, n$.
- Discrete vs. continuous time - Discrete time model converges to continuous time model as intervals become finer and finer.

- Thus, we have the J intervals $(0, s_1]$, $(s_1, s_2]$, ..., $(s_{J-1}, s_J]$. In the j^{th} interval, we assume a constant baseline hazard $h_0(y) = \lambda_j$ for $y \in I_j = (s_{j-1}, s_j]$.
- Letting $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_J)'$, we can write the likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ for the n subjects as

$$
\begin{aligned}
& L(\boldsymbol{\beta}, \boldsymbol{\lambda} | D) \\
= & \prod_{i=1}^{n} \prod_{j=1}^{J} \left(\lambda_j \exp(\boldsymbol{x}_i' \boldsymbol{\beta})\right)^{\delta_{ij}\nu_i} \exp\left\{ - \delta_{ij} \left[\lambda_j (y_i - s_{j-1}) \right.\right. \\
& \left.\left. + \sum_{g=1}^{j-1} \lambda_g (s_g - s_{g-1}) \right] \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \right\}, \qquad (1)
\end{aligned}
$$

where $\delta_{ij} = 1$ if the i^{th} subject failed or was censored in the j^{th} interval, and 0 otherwise, $\boldsymbol{x}_i' = (x_{i1}, x_{i2}, \ldots, x_{ip})$ denotes the $p \times 1$ vector of covariates for the i^{th} subject, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is the corresponding vector of regression coefficients.

- The indicator δ_{ij} is needed to properly define the likelihood over the J intervals. The semiparametric model in (1), sometimes referred to as a *piecewise exponential model*, is quite general and can accommodate various shapes of the baseline hazard over the intervals.
- If $J = 1$, the model reduces to a parametric exponential model with failure rate parameter $\lambda \equiv \lambda_1$.
- The piecewise exponential model is a useful and simple model for modeling survival data. It serves as the benchmark for comparisons with other semiparametric or fully parametric models for survival data.
- A common prior of the baseline hazard $\boldsymbol{\lambda}$ is the independent gamma prior $\lambda_j \sim \mathcal{G}(\alpha_{0j}, \lambda_{0j})$ for $j = 1, 2, \ldots, J$, where $\mathcal{G}(\alpha_{0j}, \lambda_{0j})$ denote the gamma distribution with parameters $(\alpha_{0j}, \lambda_{0j})$, with density given by

$$\pi(\lambda_j | \alpha_{0j}, \lambda_{0j}) \propto \lambda_j^{\alpha_{0j}-1} \exp(-\lambda_{0j}\lambda_j).$$

- Here α_{0j} and λ_{0j} are prior parameters which can be elicited through the prior mean and variance of λ_j.
- Another approach is to build a prior correlation among the λ_j's using a correlated prior $\boldsymbol{\psi} \sim N(\boldsymbol{\psi}_0, \Sigma_\psi)$, where $\psi_j = \log(\lambda_j)$ for $j = 1, 2, \ldots, J$.
- The likelihood in (1) is based on continuous survival data.
- The likelihood function based on grouped or discretized survival data is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}|D) \propto \prod_{j=1}^{J} G_j^*,$$

where

$$G_j^* = \exp\Big\{ -\lambda_j \Delta_j \sum_{k \in \mathcal{R}_j - \mathcal{D}_j} \exp(\boldsymbol{x}_k' \boldsymbol{\beta}) \Big\} \times \prod_{l \in \mathcal{D}_j} \left[1 - \exp\{-\lambda_j \Delta_j \exp(\boldsymbol{x}_l' \boldsymbol{\beta})\}\right], \quad (2)$$

$\Delta_j = s_j - s_{j-1}$, $\mathcal{R}_j$ is the set of patients at risk, and $\mathcal{D}_j$ is the set of

patients having failures in the j^{th} interval.

- When one considers the piecewise constant baseline hazard model with $h_j = \Delta_j \lambda_j$ and $\Delta_j = s_j - s_{j-1}$, there is a great similarity between the two likelihoods.

♠ Informative Prior Specifications: The Power Prior

- In many cancer and AIDS clinical trials, current studies often use treatments that are very similar or slight modifications of treatments used in previous studies. We refer to data arising from previous similar studies as *historical data*.
- In carcinogenicity studies, for example, large historical databases exist for the control animals from previous experiments.
- In all of these situations, it is natural to incorporate the historical data into the current study by quantifying it with a suitable prior distribution on the model parameters.

- From a Bayesian perspective, historical data from past similar studies can be very helpful in interpreting the results of the current study
- To fix ideas, suppose we have historical data from a similar previous study, denoted by $D_0 = (n_0, \boldsymbol{y}_0, X_0)$ where n_0 is the sample size of the historical data, $\boldsymbol{y}_0$ is the $n_0 \times 1$ response vector, and X_0 is the $n_0 \times p$ matrix of covariates based on the historical data.
- The power prior is defined to be the likelihood function based on the historical data D_0, raised to a power a_0.
- Here, $0 \leq a_0 \leq 1$ is a scalar parameter that controls the influence of the historical data on the current data.

- We consider the power prior for an arbitrary regression model (Ibrahim and Chen, 2000). Let the data from the *current* study be denoted by $D = (n, \boldsymbol{y}, X)$, where n denotes the sample size, $\boldsymbol{y}$ denotes the $n \times 1$ response vector, and X denotes the $n \times p$ matrix of covariates. Further, denote the likelihood for the current study by $L(\boldsymbol{\theta}|D)$, where $\boldsymbol{\theta}$ is a vector of indexing parameters.

- Now suppose we have historical data from a similar previous study, denoted by $D_0 = (n_0, \boldsymbol{y}_0, X_0)$. Further, let $\pi_0(\boldsymbol{\theta}|.)$ denote the prior distribution for $\boldsymbol{\theta}$ before the historical data D_0 is observed. We shall call $\pi_0(\boldsymbol{\theta}|.)$ the *initial prior* distribution for $\boldsymbol{\theta}$.

- Given a_0, we define the *power prior* distribution of $\boldsymbol{\theta}$ for the current study as

$$\pi(\boldsymbol{\theta}|D_0, a_0) \propto L(\boldsymbol{\theta}|D_0)^{a_0} \pi_0(\boldsymbol{\theta}|c_0), \tag{3}$$

  where c_0 is a specified hyperparameter for the initial prior, and a_0 is a scalar prior parameter that weights the historical data relative to the likelihood of the current study.

- The prior parameter c_0 controls the impact of $\pi_0(\boldsymbol{\theta}|c_0)$ on the entire prior, and the parameter a_0 controls the influence of the historical data on $\pi(\boldsymbol{\theta}|D_0, a_0)$.
- The parameter a_0 can be interpreted as a relative precision parameter for the historical data. It is reasonable to restrict the range of a_0 to be between 0 and 1, and thus we take $0 \leq a_0 \leq 1$.
- One of the main roles of a_0 is that it controls the heaviness of the tails of the prior for $\boldsymbol{\theta}$. As a_0 becomes smaller, the tails of (3) become heavier.
- Setting $a_0 = 1$, (3) corresponds to the update of $\pi_0(\boldsymbol{\theta}|c_0)$ using Bayes theorem. That is, with $a_0 = 1$, (3) corresponds to the posterior distribution of $\boldsymbol{\theta}$ from the previous study.
- When $a_0 = 0$, then the prior does not depend on the historical data, and in this case, $\pi(\boldsymbol{\theta}|D_0, a_0 = 0) \equiv \pi_0(\boldsymbol{\theta}|c_0)$. Thus, $a_0 = 0$ is equivalent to a prior specification with no incorporation of historical data. Therefore, (3) can be viewed as a generalization of the usual

Bayesian update of $\pi_0(\boldsymbol{\theta}|c_0)$.

- The parameter a_0 allows the investigator to control the influence of the historical data on the current study. Such control is important in cases where there is heterogeneity between the previous and current study, or when the sample sizes of the two studies are quite different.
- The hierarchical power prior specification is completed by specifying a (proper) prior distribution for a_0. This leads to a joint power prior distribution for $(\boldsymbol{\theta}, a_0)$ of the form

$$\pi(\boldsymbol{\theta}, a_0|D_0) \propto L(\boldsymbol{\theta}|D_0)^{a_0}\pi_0(\boldsymbol{\theta}|c_0)\pi(a_0|\boldsymbol{\gamma}_0), \tag{4}$$

where $\boldsymbol{\gamma}_0$ is a specified hyperparameter vector. A natural choice for $\pi(a_0|\boldsymbol{\gamma}_0)$ is a beta prior.

♠ **Example 1.1: Melanoma data.**

- Melanoma incidence is increasing at a rate that exceeds all solid tumors. Although education efforts have resulted in earlier detection of melanoma, patients who have deep primary melanoma ($> 4\ mm$) or melanoma metastatic to regional draining lymph nodes, classified as *high-risk melanoma* patients, continue to have high relapse and mortality rates of 60% to 75% (see KirkwoodKirkwood, J.M. et al., 2000).
- Recently, several post-operative (adjuvant) chemotherapies have been proposed for this class of melanoma patients, and the one which seems to provide the most significant impact on relapse-free survival and survival is interferon alpha-2b (IFN). This chemotherapy was used in two recent Eastern Cooperative Oncology Group (ECOG) phase III clinical trials, E1684 and E1690.

- The first trial, E1684, was a two-arm clinical trial comparing high-dose interferon (IFN) to observation (OBS). There were a total of $n_0 = 286$ patients enrolled in the study, accrued from 1984 to 1990, and the study was unblinded in 1993. The results of this study suggested that IFN has a significant impact on relapse-free survival (RFS) and survival (OS), which led to U.S. Food and Drug Administration (FDA) approval of this regimen as an adjuvant therapy for high-risk melanoma patients.
- The ECOG trial E1690 was a three-arm phase III clinical trial, and had treatment arms consisting of high-dose interferon, low-dose interferon, and observation. This study had $n = 427$ patients on the high-dose interferon arm and observation arm combined. E1690 was initiated right after the completion of E1684.

- We consider the E1690 melanoma data as the current data (D). We use E1684 as the historical data (D_0), and incorporate the E1684 data via the power prior in (3).
- The power prior for (β, λ) for model (1) is given by

$$\pi(\beta, \lambda, a_0|D_0) \propto L(\beta, \lambda|D_0)^{a_0} \pi_0(\beta, \lambda)\ a_0^{\kappa_0 - 1}(1 - a_0)^{\xi_0 - 1},$$

  where (κ_0, ξ_0) are specified hyperparameters for the prior distribution of a_0, $\pi_0(\beta, \lambda)$ is the initial prior distribution for (β, λ), and $L(\beta, \lambda|D_0)$ is the likelihood function in (1) with D_0 in place of D.

- For the initial prior $\pi_0(\beta, \lambda)$, we assume that β and λ are independent, where β has a uniform prior and λ has a Jeffreys's prior. This leads to the joint initial improper prior

$$\pi_0(\beta, \lambda) \propto \prod_{j=1}^{J} \lambda_j^{-1}. \tag{5}$$

- We consider the treatment covariate alone in the example here, and

thus β is one dimensional.

- Table 1.1 shows results based on several values of a_0 using the initial prior in (5). The value $a_0 = 0$ corresponds to a Bayesian analysis of E1690 using noninformative priors, that is, not using any historical data. A value of $a_0 = 1$ corresponds to giving the historical and current data equal weight. In Table 1.1, HR denotes the hazard ratio of OBS to IFN, SD denotes the posterior standard deviation, and 95% HPD denotes 95% Highest Posterior Density intervals.
- We see from Table 1.1 that as more weight is given to the historical data, the posterior hazard ratios increase and the HPD intervals become narrower and do not include 1.
- This is reasonable since the posterior hazard ratios based on the E1684 data alone were much larger than E1690 alone, and therefore as more weight is given to E1684, the greater the posterior hazard ratios and the narrower the HPD intervals.

- Thus, the incorporation of E1684 into the current analysis via the power prior sharpens the assessment between IFN and OBS and leads to more definitive conclusions about the effect of IFN.
- This example thus demonstrates the effect of incorporating historical data into an analysis.

TABLE 1.1. Posterior Estimates of Hazard Ratio for E1690 using E1684 as Historical Data

$E(a_0 \mid D, D_0)$	HR	SD	95% HPD
0	1.30	0.17	(0.99, 1.64)
0.05	1.30	0.16	(0.99, 1.63)
0.30	1.33	0.15	(1.03, 1.63)
1	1.36	0.13	(1.11, 1.62)

♠ Models Using a Gamma Process

- The gamma process is perhaps the most commonly used nonparametric prior process for the Cox model.
- The seminal paper by Kalbfleisch (1978, JRSSB) describes the gamma process prior for the baseline cumulative hazard function.
- The gamma process can be described as follows. Let $\mathcal{G}(\alpha, \lambda)$ denote the gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$.
- Let $\alpha(t), t \geq 0$, be an increasing left continuous function such that $\alpha(0) = 0$, and let $Z(t), t \geq 0$, be a stochastic process with the properties:

  (i) $Z(0) = 0$;

  (ii) $Z(t)$ has independent increments in disjoint intervals; and

  (iii) for $t > s$, $Z(t) - Z(s) \sim \mathcal{G}(c(\alpha(t) - \alpha(s)), c)$.

- Then the process $\{Z(t) : t \geq 0\}$ is called a gamma process and is denoted by $Z(t) \sim \mathcal{GP}(c\alpha(t), c)$. We note here that $\alpha(t)$ is the mean of the process and c is a weight or confidence parameter about the mean.
- The sample paths of the gamma process are almost surely increasing functions.

♠ Gamma Process on Cumulative Hazard

- Under the Cox model, the joint probability of survival of n subjects given the covariate matrix X is given by

$$P(\mathbf{Y} > \boldsymbol{y}|\boldsymbol{\beta}, X, H_0) = \exp\left\{-\sum_{j=1}^{n} \exp(\boldsymbol{x}_j'\boldsymbol{\beta}) H_0(y_j)\right\}.$$

- The gamma process is often used as a prior for the cumulative baseline hazard function $H_0(y)$.

- In this case, we take

$$H_0 \sim \mathcal{GP}(c_0 H^*, c_0), \tag{6}$$

  where $H^*(y)$ is an increasing function with $H^*(0) = 0$.

- H^* is often assumed to be a known parametric function with hyperparameter vector $\boldsymbol{\gamma}_0$. For example, if H^* corresponds to the exponential distribution, then $H^*(y) = \gamma_0 y$, where γ_0 is a specified hyperparameter. If $H^*(y)$ is taken as Weibull, then $H^*(y) = \eta_0 y^{\kappa_0}$, where $\boldsymbol{\gamma}_0 = (\eta_0, \kappa_0)'$ is a specified vector of hyperparameters.

- The marginal survival function is given by

$$P(\mathbf{Y} > \boldsymbol{y}|\boldsymbol{\beta}, X, \boldsymbol{\gamma}_0, c_0)$$
$$= \prod_{j=1}^{n} [\phi(iV_j)]^{c_0(H^*(y_{(j)}) - H^*(y_{(j-1)}))}, \tag{7}$$

  where ϕ is the characteristic function of an infinitely divisible distribution function with unit mean, $V_j = \sum_{l \in \mathcal{R}_j} \exp(x_l'\boldsymbol{\beta})$, $\mathcal{R}_j$ is the risk set at time $y_{(j)}$, and $y_{(1)} < y_{(2)} < \ldots, < y_{(n)}$ are distinct ordered times.

- For continuous data, when the ordered survival times are all distinct, the likelihood of $(\boldsymbol{\beta}, \gamma_0, c_0)$ can be obtained by differentiating (7). Note that this likelihood, used by Kalbfleisch (1978) and Clayton (1991), and among others, is defined only when the observed survival times are distinct.

- Sinha, Ibrahim, and Chen (2003, Biometrika) provide a Bayesian justification of Cox's partial likelihood via the gamma rocess prior.
- The extension to the cases when ties are present is also considered by Chen, Ibrahim, and Shao (2004). Chen, Ibrahim, and Shao (2004) further develop an efficient Markov chain Monte Carlo algorithm to sample from the resulting posterior distribution.

♠ Gamma Process with Grouped-Data Likelihood

- Again, we construct a finite partition of the time axis, $0 < s_1 < s_2 < \ldots < s_J$, with $s_J > y_i$ for all $i = 1, \ldots, n$. Thus, we have the J disjoint intervals $(0, s_1]$, $(s_1, s_2]$, ..., $(s_{J-1}, s_J]$, and let $I_j = (s_{j-1}, s_j]$.
- The observed data D is assumed to be available as grouped within these intervals, such that $D = (X, \mathcal{R}_j, \mathcal{D}_j : j = 1, 2, \ldots, J)$, where $\mathcal{R}_j$ is the risk set and $\mathcal{D}_j$ is the failure set of the j^{th} interval I_j.

- Let h_j denote the increment in the cumulative baseline hazard in the j^{th} interval, that is

$$h_j = H_0(s_j) - H_0(s_{j-1}), \quad j = 1, 2, \ldots, J.$$

- The gamma process prior in (6) implies that the h_j's are independent and

$$h_j \sim \mathcal{G}(\alpha_{0j} - \alpha_{0,j-1}, c_0), \tag{8}$$

  where $\alpha_{0j} = c_0 H^*(s_j)$, and H^* and c_0 are defined in the previous subsection.

- Thus, the hyperparameters (H^*, c_0) for h_j consist of a specified parametric cumulative hazard function $H^*(y)$ evaluated at the endpoints of the time intervals, and a positive scalar c_0 quantifying the degree of prior confidence in $H^*(y)$.

- Now writing $H_0 \sim \mathcal{GP}(c_0 H^*, c_0)$ implies that every disjoint increment in H_0 has the prior given by (8).

- Thus, the grouped data representation can be obtained as

$$P(y_i \in I_j|\boldsymbol{h}) = \exp\left\{-\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\sum_{k=1}^{j-1} h_k\right\}[1-\exp\{-h_j\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}],$$

  where $\boldsymbol{h} = (h_1, h_2, \ldots, h_J)'$.

- This leads to the grouped data likelihood function

$$L(\boldsymbol{\beta}, \boldsymbol{h}|D) \propto \prod_{j=1}^{J} G_j,$$

  where

$$G_j = \exp\Big\{-h_j \sum_{k\in\mathcal{R}_j-\mathcal{D}_j} \exp(\boldsymbol{x}_k'\boldsymbol{\beta})\Big\} \prod_{l\in\mathcal{D}_j} \Big[1-\exp\{-h_j\exp(\boldsymbol{x}_l'\boldsymbol{\beta})\}\Big].$$

- A typical prior for $\boldsymbol{\beta}$ is a $N_p(\boldsymbol{\mu}_0, \Sigma_0)$ distribution. Thus, the joint posterior of $(\boldsymbol{\beta}, \boldsymbol{h})$ can be written as

$$\pi(\boldsymbol{\beta}, \boldsymbol{h}|D) \propto \prod_{j=1}^{J} \left[G_j h_j^{(\alpha_{0j} - \alpha_{0,j-1})-1} \exp(-c_0 h_j) \right] \times \exp\left\{ -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)\Sigma_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0) \right\}.$$

- To sample from the joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{h})$, it can be shown that $[\boldsymbol{\beta}|\boldsymbol{h}, D]$ is log-concave in $\boldsymbol{\beta}$ and thus the adaptive rejection algorithm can be used efficiently to sample the components of $\boldsymbol{\beta}$.
- Moreover, $[\boldsymbol{h}|\boldsymbol{\beta}, D]$ is also log-concave in the components of $\boldsymbol{h}$. We can thus carry out the following Gibbs sampling scheme:

(i) Sample from

$$\pi(\beta_j|\boldsymbol{\beta}^{(-j)}, \boldsymbol{h}, D)$$
$$\propto \prod_{j=1}^{J} G_j \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)\Sigma_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)\right\},$$

using the adaptive rejection algorithm for $j = 1, 2, \ldots, p$.

(ii) Sample from

$$\pi(h_j|\boldsymbol{h}^{(-j)}, \boldsymbol{\beta}, D) \propto h_j^{(\alpha_{0j}-\alpha_{0,j-1})-1}$$
$$\times \exp\left\{-h_j\Big(\sum_{k\in\mathcal{R}_j-\mathcal{D}_j} \exp(\boldsymbol{x}_k'\boldsymbol{\beta}) + c_0\Big)\right\}, \qquad (9)$$

where $\boldsymbol{h}^{(-j)}$ denote the $\boldsymbol{h}$ vector without the j^{th} component. The full conditional distribution in (9) can be well approximated by a gamma distribution, and thus a more efficient Gibbs sampling scheme would be to replace (9) by

(ii*) Sample from $[\boldsymbol{h}|\boldsymbol{\beta}, D]$ using independent samples from a conditional posterior approximated by

$$h_j \sim \mathcal{G}\left(\alpha_{0j} - \alpha_{0,j-1} + d_j, c_0 + \sum_{k \in \mathcal{R}_j - \mathcal{D}_j} \exp(x_k'\boldsymbol{\beta})\right).$$

♠ Gamma Process on Baseline Hazard

- An alternative specification of the semiparametric Cox model is to specify a gamma process prior on the hazard rate itself.
- We construct the likelihood by using a piecewise constant baseline hazard model and use only information about which interval the failure times fall into. Let $0 = s_0 < s_1 < \ldots < s_J$ be a finite partition of the time axis and let

$$\delta_j = h_0(s_j) - h_0(s_{j-1})$$

denote the increment in the baseline hazard in the interval $(s_{j-1}, s_j]$, $j = 1, 2, \ldots, J$, and $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_J)'$.

- For an arbitrary individual in the population, the survival function for the Cox model at time y is given by

$$\begin{aligned} S(y|\boldsymbol{x}) &= \exp\left\{-\eta \int_0^y h_0(u)\, du\right\} \\ &\approx \exp\left\{-\eta \left(\sum_{i=1}^{J} \delta_i (y - s_{i-1})^+\right)\right\}, \end{aligned} \tag{10}$$

  where $h_0(0) = 0$, $(u)^+ = u$ if $u > 0$, 0 otherwise, and $\eta = \exp(\boldsymbol{x}'\boldsymbol{\beta})$.

- This first approximation arises since the specification of $\boldsymbol{\delta}$ does not specify the entire hazard rate, but only the δ_j.

- Let p_j denote the probability of a failure in the interval $(s_{j-1}, s_j]$, $j = 1, 2, \ldots, J$. Using (10), we have

$$\begin{aligned} p_j =& S(s_{j-1}) - S(s_j) \approx \exp\left\{-\eta \sum_{l=1}^{j-1} \delta_l (s_{j-1} - s_{l-1})\right\} \\ &\times \left[1 - \exp\left\{-\eta (s_j - s_{j-1}) \sum_{l=1}^{j} \delta_l\right\}\right]. \end{aligned}$$

- Thus, in the j^{th} interval $(s_{j-1}, s_j]$, the contribution to the likelihood function for a failure is p_j, and $S(s_j)$ for a right censored observation.
- For $j = 1, 2, \ldots, J$, let d_j be the number of failures, $\mathcal{D}_j$ be the set of subjects failing, c_j be the number of right censored observations and $\mathcal{C}_j$ is the set of subjects that are censored. Also, let $D = (n, \boldsymbol{y}, X, \boldsymbol{\nu})$ denote the data.
- The grouped data likelihood function is thus given by

$$L(\boldsymbol{\beta}, \boldsymbol{\delta}|D) = \prod_{j=1}^{J} \left\{ \exp\left\{ -\delta_j(a_j + b_j) \right\} \times \prod_{k \in \mathcal{D}_j} [1 - \exp\{-\eta_k T_j\}] \right\}, \tag{11}$$

where $\eta_k = \exp(\boldsymbol{x}_k'\boldsymbol{\beta})$,

$$a_j = \sum_{l=j+1}^{J} \sum_{k \in \mathcal{D}_l} \eta_k (s_{l-1} - s_{j-1}),$$

$$b_j = \sum_{l=j}^{J} \sum_{k \in \mathcal{C}_l} \eta_k (s_l - s_{j-1}),$$

and

$$T_j = (s_j - s_{j-1}) \sum_{l=1}^{j} \delta_l.$$

- We note that the grouped data likelihood involves the following approximation: Instead of conditioning on exact event times, we condition on the set of failures and set of right censored events in each interval, and thus we approximate continuous right censored data by grouped data.

♢ Prior Elicitation

- We demonstrate the use of the power prior for the gamma process model.
- Let $D_0 = (n_0, \boldsymbol{y}_0, X_0, \boldsymbol{\nu}_0)$ denote the data from the previous study (i.e., historical data), where n_0 denotes the sample size of the previous study, y_0 denotes a right censored vector of survival times with censoring indicators $\boldsymbol{\nu}_0$, and X_0 denotes the $n \times p$ matrix of covariates.
- Let $\pi_0(\boldsymbol{\beta}, \boldsymbol{\delta})$ denote the initial prior for $(\boldsymbol{\beta}, \boldsymbol{\delta})$.
- Let a_0 denote the weight parameter, $0 \leq a_0 \leq 1$.
- The power prior distribution for $(\boldsymbol{\beta}, \boldsymbol{\delta})$ is given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{\delta} | D_0, a_0) \ \propto \ \{L(\boldsymbol{\beta}, \boldsymbol{\delta} | D_0)\}^{a_0} \ \pi_0(\boldsymbol{\beta}, \boldsymbol{\delta}),$$

  where $L(\boldsymbol{\beta}, \boldsymbol{\delta} | D_0)$ is the likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\delta})$ based on the historical data D_0 and thus, $L(\boldsymbol{\beta}, \boldsymbol{\delta} | D_0)$ is (11) with D replaced by $D_0 = (n_0, \boldsymbol{y}_0, X_0, \boldsymbol{\nu}_0)$.

- To simplify the prior specification, we take

$$\pi_0(\boldsymbol{\beta}, \boldsymbol{\delta}) = \pi_0(\boldsymbol{\beta}|c_0)\pi_0(\boldsymbol{\delta}|\boldsymbol{\theta}_0),$$

  where c_0 and $\boldsymbol{\theta}_0$ are fixed hyperparameters.

- Specifically, we take $\pi_0(\boldsymbol{\beta}|c_0)$ to be a p dimensional multivariate normal density, $N_p(0, c_0 W_0)$, with mean 0 and covariance matrix $c_0 W_0$, where c_0 is a specified scalar and W_0 is a $p \times p$ diagonal matrix.

- We take $\pi_0(\boldsymbol{\delta}|\boldsymbol{\theta}_0)$ to be a product of J independent gamma densities, each with mean f_{0i}/g_{0i} and variance f_{0i}/g_{0i}^2, $i = 1, 2, \ldots, J$. So, we get

$$\pi_0(\boldsymbol{\delta}|\boldsymbol{\theta}_0) \propto \prod_{i=1}^{J} \delta_i^{f_{0i}-1} \exp\{-\delta_i g_{0i}\},$$

  where $\boldsymbol{\theta}_0 = (f_{01}, g_{01}, \ldots, f_{0J}, g_{0J})'$.

- The prior specification is completed by specifying a prior for a_0. We specify a beta prior for a_0 $(0 \leq a_0 \leq 1)$, so that

$$\pi(a_0|\alpha_0, \lambda_0) \propto a_0^{\alpha_0 - 1}(1 - a_0)^{\lambda_0 - 1},$$

thus obtaining the joint prior

$$\begin{aligned}\pi(\boldsymbol{\beta}, \boldsymbol{\delta}, a_0|D_0) \propto & L(\boldsymbol{\beta}, \boldsymbol{\delta}|D_0)^{a_0}\pi_0(\boldsymbol{\beta}|c_0) \\ & \times \pi_0(\boldsymbol{\delta}|\boldsymbol{\theta}_0)\pi(a_0|\alpha_0, \lambda_0).\end{aligned}$$

♠ Example 1.2: Myeloma Data

- Our main goal in this example is to illustrate the behavior of the power prior and the sensitivity of the posterior estimates to the choice of a_0 and c_0.
- The current dataset is E2479 (Study 2) and the historical dataset, which consists of a similar study with $n_0 = 65$ in multiple myeloma conducted several years earlier, is labeled Study 1. Two superimposed Kaplan-Meier plots for the two studies are displayed in Figure 1.1.

- A total of $n = 339$ observations were available from E2479, with 8 observations being right censored. Our analysis used $p = 8$ covariates. These are blood urea nitrogen (x_1), hemoglobin (x_2), platelet count (x_3) (1 if normal, 0 if abnormal), age (x_4), white blood cell count (x_5), bone fractures (x_6), percentage of the plasma cells in bone marrow (x_7), and serum calcium (x_8).
- To ease the computational burden, we standardized all of the variables. The standardization helped the numerical stability in the implementation of the adaptive rejection algorithm for sampling the regression coefficients from the posterior distribution. Study 1 consisted of $n_0 = 65$ observations of which 17 were right censored.

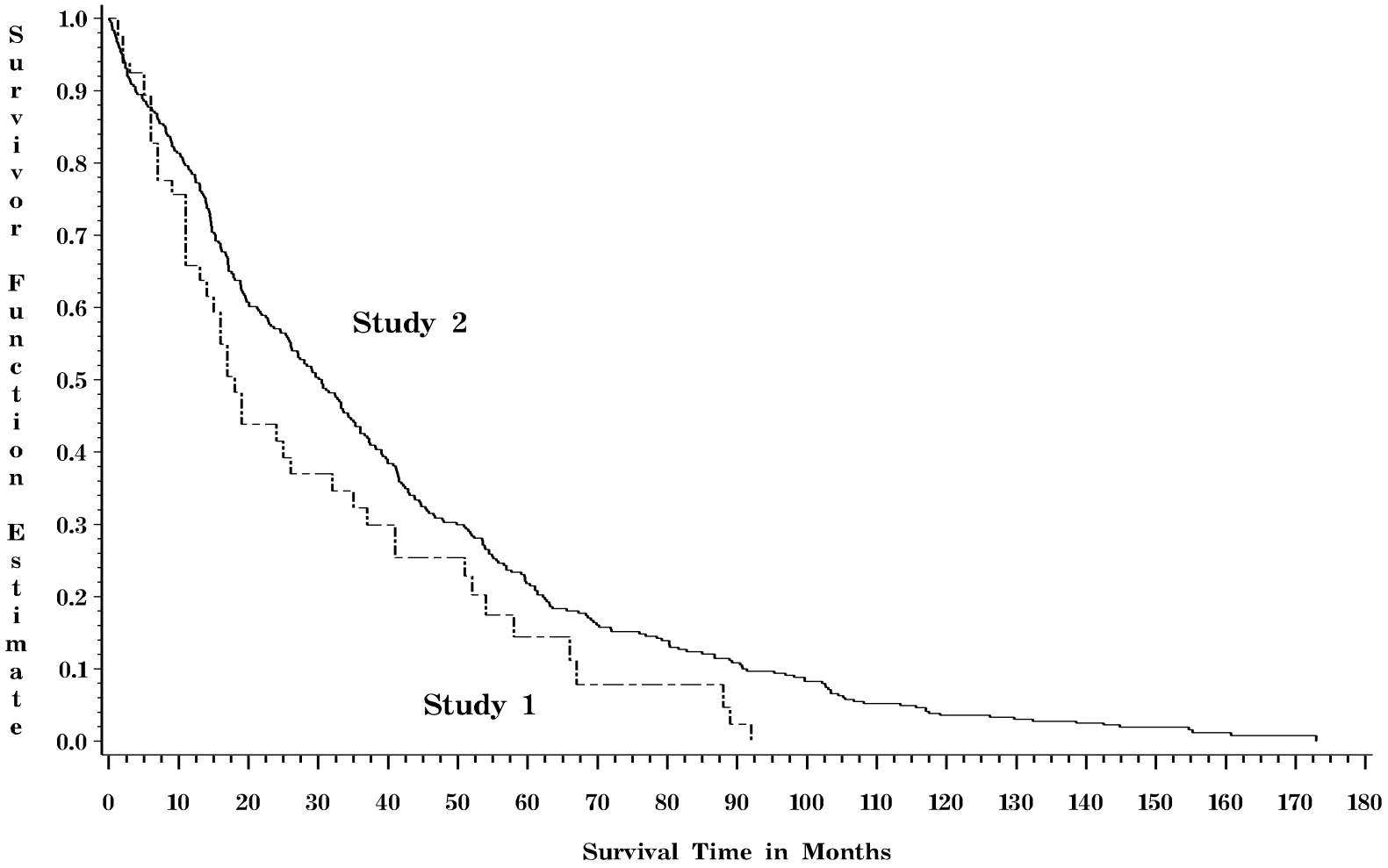

FIGURE 1.1. Survival curves for Study 1 and Study 2.

- We used $J = 28$ with the intervals chosen so that with the combined data sets from the historical and current data, at least one failure or censored observation falls in each interval. This technique for choosing J is reasonable and preserves the consistency in the interpretation of Δ for the two studies.
- In addition, we take $f_{0i} = s_i - s_{i-1}$ if $s_i - s_{i-1} \geq 1$ and $f_{0i} = 1.1$ if $s_i - s_{i-1} < 1$, and $g_{0i} = 0.001$.
- For the last interval, we take $g_{0i} = 10$ for $i = J$ since very little information in the data is available for this last interval.
- The above choices of f_{0i} and g_{0i} ensure the log-concavity of $\pi_0(\boldsymbol{\delta} \mid \theta_0)$, as this is required in sampling $\boldsymbol{\delta}$ from its conditional prior and posterior distributions.
- To obtain the posterior estimates of $\boldsymbol{\beta}$, 50,000 Gibbs iterations were used after convergence. Tables 1.2 and 1.3 show the posterior estimates of $\boldsymbol{\beta}$ under the prior parameters $(\mu_{a_0}, \sigma_{a_0}) = (0, 0)$, $(0.5, 0.06)$, $(1, 0)$, and $c_0 = 3$, 10. We note that $\sigma_{a_0} = 0$ implies

$a_0 = 0$ or $a_0 = 1$ with probability 1.

- From these two tables, it can be seen that the posterior means of $\boldsymbol{\beta}$ are very similar, and the posterior standard deviations are slightly smaller when μ_{a_0} is getting larger. Thus, the results are not too sensitive to these values of $(\mu_{a_0}, \sigma_{a_0})$ and c_0. This may be partially explained by the relatively small sample size of the historical data.
- From Table 1.2, we see that as a_0 increases, the posterior standard deviations of the regression coefficients decrease, and thus the precision in the estimates is improved. Thus, one of the advantages in incorporating historical data is that it increases the precision in estimation.
- Another feature in Tables 1.2 and 1.3 is that the estimates are remarkably robust with respect to the choices of $(\mu_{a_0}, \sigma_{a_0})$ and c_0, thus revealing a desirable property in the power prior.

TABLE 1.2. Posterior Estimates of β for Myeloma Data with $c_0 = 3$.

	Posterior Mean			Posterior Std. Error		
Variable	$a_0 = 0$	$\mu_{a_0} = 0.5$	$a_0 = 1$	$a_0 = 0$	$\mu_{a_0} = 0.5$	$a_0 = 1$
x_1	0.133	0.134	0.153	0.046	0.046	0.044
x_2	−0.178	−0.182	−0.186	0.046	0.045	0.043
x_3	−0.069	−0.076	−0.087	0.049	0.049	0.046
x_4	0.406	0.411	0.371	0.049	0.049	0.046
x_5	0.229	0.236	0.218	0.054	0.054	0.051
x_6	−0.004	−0.002	0.015	0.051	0.051	0.048
x_7	0.440	0.446	0.399	0.058	0.058	0.054
x_8	0.235	0.241	0.224	0.057	0.057	0.053

TABLE 1.3. Posterior Estimates of β for Myeloma Data with $c_0 = 10$.

	Posterior Mean			Posterior Std. Error		
Variable	$a_0 = 0$	$\mu_{a_0} = 0.5$	$a_0 = 1$	$a_0 = 0$	$\mu_{a_0} = 0.5$	$a_0 = 1$
x_1	0.131	0.134	0.153	0.047	0.046	0.045
x_2	−0.180	−0.182	−0.189	0.046	0.046	0.044
x_3	−0.072	−0.076	−0.090	0.049	0.049	0.046
x_4	0.413	0.411	0.377	0.050	0.050	0.047
x_5	0.226	0.236	0.224	0.055	0.054	0.052
x_6	−0.005	−0.002	0.015	0.052	0.051	0.048
x_7	0.449	0.446	0.406	0.058	0.058	0.055
x_8	0.241	0.241	0.228	0.057	0.057	0.054

♠ Beta Process Models

- We first discuss time-continuous right censored survival data without covariates.
- In this context, write the definition of the cumulative hazard $H(t)$ as

$$H(t) = -\log(S(t)), \tag{12}$$

  where $S(t)$ is the survival function.
- The gamma process can be defined on $H(t)$ when this definition of the cumulative hazard is appropriate.
- A more general way of defining the hazard function, which is valid even when the survival time distribution is not continuous, is to use the definition of Hjort (1990).

- General formulae for the cumulative hazard function $H(t)$ are

$$H(t) = \int_{[0,t]} \frac{dF(u)}{S(u)}, \tag{13}$$

where

$$F(t) = 1 - S(t) = 1 - \prod_{[0,t]} \{1 - dH(t)\}. \tag{14}$$

The cumulative hazard function $H(t)$ defined here is equal to (12) when the survival distribution is absolutely continuous.

- Hjort (1990) presents what he calls a beta process with independent increments as a prior for $H(.)$.

- A beta process generates a proper cdf $F(t)$, as defined in (13), and has independent increments of the form

$$dH(s) \sim \mathcal{B}(c(s)dH^*(s), c(s)(1 - dH^*(s))), \tag{15}$$

where $\mathcal{B}(a, b)$ denotes the beta distribution with parameters (a, b).

- Due to the complicated convolution property of independent beta distributions, the exact distribution of the increment $H(s)$ is only approximately beta over any finite interval, regardless of how small the length of the interval might be.
- It is possible to deal with the beta process for the baseline cumulative hazard appropriately defined under a Cox model with time continuous data, but survival data in practice is commonly grouped within some grid intervals, where the grid size is determined by the data and trial design.
- So for practical purposes, it is more convenient and often sufficient to use a discretized version of the beta process long with grouped survival data.

- Within the spirit of the definition of the cumulative hazard function $H(t)$ defined in (13), a discretized version of the Cox model can be defined as

$$S(s_j|\boldsymbol{x}) = P(T > s_j|\boldsymbol{x}) = \prod_{k=1}^{j}(1-h_k)^{\exp(\boldsymbol{x}'\boldsymbol{\beta})},$$

where h_k is the discretized baseline hazard rate in the interval $I_k = (s_{k-1}, s_k]$. The likelihood can thus be written as

$$L(\boldsymbol{\beta}, \boldsymbol{h}) = \prod_{j=1}^{J}\left((1-h_j)^{\sum_{i\in\mathcal{R}_j-\mathcal{D}_j}\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}\right) \times \prod_{l\in D_j}\left(1-(1-h_j)^{\exp(x_l'\boldsymbol{\beta})}\right),$$

where $\boldsymbol{h} = (h_1, h_2, \ldots, h_J)'$.

- To complete the discretized beta process model, we specify independent beta priors for the h_k's. Specifically, we take $h_k \sim \mathcal{B}(c_{0k}\alpha_{0k}, c_{0k}(1-\alpha_{0k}))$, and independent for $k = 1, 2, \ldots, J$.

- Though it is reasonable to assume that the h_k's are independent from each other a priori, the assumption of an exact beta distribution of the h_k's is only due to an approximation to the true time-continuous beta process.
- Thus, according to the time-continuous beta process, the distribution of the h_k's is not exactly beta, but it can be well approximated by a beta distribution only when the width of I_k is small.
- Under the discretized beta process defined here, the joint prior density of $\boldsymbol{h}$ is thus given by

$$\pi(\boldsymbol{h}) \propto \prod_{j=1}^{J} h_j^{c_{0j}\alpha_{0j}-1}(1-h_j)^{c_{0j}(1-\alpha_{0j})-1}.$$

- A typical prior for $\boldsymbol{\beta}$ is a $N_p(\boldsymbol{\mu}_0, \Sigma_0)$ prior, which is independent of $\boldsymbol{h}$.

- Assuming an arbitrary prior for $\boldsymbol{\beta}$, the joint posterior of $(\boldsymbol{\beta}, h)$ can be written as

$$\pi(\boldsymbol{\beta}, \boldsymbol{h}|D) \propto L(\boldsymbol{\beta}, \boldsymbol{h}|D)\pi(\boldsymbol{h})\pi(\boldsymbol{\beta}) = \prod_{j=1}^{J} \left((1-h_j)^{\sum_{i \in \mathcal{R}_j - \mathcal{D}_j} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})} \right)$$

$$\times \prod_{l \in \mathcal{D}_j} \left(1 - (1-h_j)^{\exp(\boldsymbol{x}_l'\boldsymbol{\beta})} \right) \prod_{j=1}^{J} h_j^{c_{0j}\alpha_{0j}-1} (1-h_j)^{c_{0j}(1-\alpha_{0j})-1} \pi(\boldsymbol{\beta}).$$

- We can use the Gibbs sampler to sample from the joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{h})$.

♠ Correlated Gamma Processes

- Again, consider the piecewise constant hazard model. We construct a finite partition of the time axis, $0 < s_1 < s_2 < \ldots < s_J$, with $s_J > y_i$ for all $i = 1, 2, \ldots, n$, and assume a constant baseline hazard $h_0(y) = \lambda_k$ for $y \in I_k = (s_{k-1}, s_k]$.
- The grouped data likelihood function of $(\boldsymbol{\lambda}, \boldsymbol{\beta})$ can now be written as

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}|D) = \prod_{k=1}^{J} \left[\lambda_k^{d_k} \left(\prod_{i \in \mathcal{D}_k} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \right) \right.$$
$$\left. \times \exp\left\{ -\lambda_k (s_k - s_{k-1}) \left(\sum_{i \in \mathcal{R}_k} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \right) \right\} \right],$$

where $\mathcal{R}_k$ is the set of patients at risk and $\mathcal{D}_k$ is the set of patients having failures in the k^{th} interval.

- The gamma process prior of Kalbfleisch (1978) discussed earlier assumes independent cumulative hazard increments. This is unrealistic in most applied settings, and does not allow for borrowing of strength between adjacent intervals.
- Modeling dependence between hazard increments has been discussed by Gamerman (1991), who proposes a Markov prior process for the $\{\log(\lambda_k)\}$, by modeling

$$\log(\lambda_k) = \log(\lambda_{k-1}) + \epsilon_k,$$

$$E(\epsilon_k) = 0, \quad \text{and} \quad \text{Var}(\epsilon_k) = \sigma_k^2.$$

- Arjas and Gasbarra (1994) introduced a first-order autoregressive structure on the increment of the hazards by taking

$$\lambda_k | \lambda_{k-1} \sim \mathcal{G}(\alpha_k, \alpha/\lambda_{k-1})$$

for $k > 1$.

- Nieto-Barajas and Walker (2000) propose dependent hazard rates with a Markovian relation, given by $\lambda_1 \sim \mathcal{G}(\alpha_1, \gamma_1)$,

$$u_k|\lambda_k, v_k \sim \mathcal{P}(v_k\lambda_k), \;\; v_k|\xi_k \sim \mathcal{E}\left(1/\xi_k\right), \tag{16}$$

$$\lambda_{k+1}|u_k, v_k \sim \mathcal{G}(\alpha_{k+1} + u_k, \gamma_{k+1} + v_k), \tag{17}$$

  and

$$\boldsymbol{\beta} \sim \pi(\boldsymbol{\beta}),$$

  for $k \geq 1$, where $\pi(\boldsymbol{\beta})$ denotes the prior for $\boldsymbol{\beta}$, which can be taken to be a normal distribution, for example.

- Following Nieto-Barajas and Walker (2000), the innovation in the process is the introduction of two latent processes $\{u_k\}$ and $\{v_k\}$, which are not observable and cannot be expressed in terms of observable quantities.

- The role of the processes $\{u_k\}$ and $\{v_k\}$ is fundamental, since they allow us to control the strength of the correlation between different hazard increments.

- These processes can be characterized by examining the following conditional expectations:

$$E(\lambda_{k+1}|u_k, v_k) = \frac{\alpha_{k+1} + u_k}{\gamma_{k+1} + v_k},$$

$$E(u_k|\lambda_k, v_k) = v_k\lambda_k, \;\; E(u_k|\lambda_k) = \lambda_k\xi_k,$$

and

$$E(\lambda_{k+1}|\lambda_k, v_k) = \frac{\alpha_{k+1} + v_k\lambda_k}{\gamma_{k+1} + v_k},$$

where α_k, γ_k, and ξ_k are specified hyperparameters.

- Note that we do not force the process to have the same first-order autoregressive structure, in that the dependence can vary along the time axis.

- We also note that the higher the value of v_k, the closer the process is to first-order autoregressive.

- When $v_k = 0$, we have the independence model of Kalbfleisch (1978).

- Allowing v_k to be a random process in (16) enables us to learn about the degree of correlation between increments in adjacent time intervals through the data.
- This alternative Markov prior process generalizes the independent gamma process, but keeps the convenient conjugacy property of a gamma prior with the Poisson distribution.
- When α_k, γ_k and v_k are constant over time, the hazard increments λ_k are marginally distributed as independent and identically distributed (i.i.d.) gamma variates, and the process is a stationary process.
- Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_J)'$, $\boldsymbol{u} = (u_0, u_1, \ldots, u_J)'$, $\boldsymbol{v} = (v_0, v_1, \ldots, v_J)'$, and define $u_0 = v_0 = 0$. With this process, the joint posterior is

given by

$$\begin{aligned}
&\pi(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}|D) \propto L(\boldsymbol{\lambda}, \boldsymbol{\beta}|D) \\
&\times \prod_{k=1}^{J} \{\pi(\lambda_k|u_{k-1}, v_k)\pi(u_k|\lambda_k, v_k)\pi(v_k|\xi_k)\}\, \pi(\boldsymbol{\beta}) \\
&\propto \prod_{k=1}^{J} \left[\lambda_k^{d_k} A_k \frac{\lambda_k^{\alpha_k+u_{k-1}-1} \exp\left(-\lambda_k(\gamma_k + v_{k-1})\right)}{\Gamma(\alpha_k + u_{k-1})}\right. \\
&\left. \times (\gamma_k + v_{k-1})^{\alpha_k+u_{k-1}} \frac{\exp(-v_k\lambda_k)(v_k\lambda_k)^{u_k}}{u_k!} \exp(-\frac{v_k}{\xi_k})\right] \\
&\times \exp\left(-\sum_{k=1}^{J} \lambda_k(s_k - s_{k-1})B_k\right) \pi(\boldsymbol{\beta}),
\end{aligned}$$

where d_k is the number of events in $I_k = (s_{k-1}, s_k]$,

$$A_k = \prod_{i \in \mathcal{D}_k} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}), \text{ and } B_k = \sum_{i \in \mathcal{R}_k} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}).$$

- After some algebra, we obtain

$$\begin{aligned}
&\pi(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{v}|D) \\
&\propto \exp\left\{-\sum_{k=1}^{J} \lambda_k \left((s_k - s_{k-1})B_k + \gamma_k + v_{k-1} + v_k\right)\right\} \\
&\quad \times \prod_{k=1}^{J} \frac{A_k(\gamma_k + v_{k-1})^{\alpha_k + u_{k-1}} v_k^{u_k} \exp(-\frac{v_k}{\xi_k})}{\Gamma(\alpha_k + u_{k-1})u_k!} \\
&\quad \times \prod_{k=1}^{J} \lambda_k^{d_k + \alpha_k + u_{k-1} + u_k - 1} \times \pi(\boldsymbol{\beta}).
\end{aligned}$$

1.3 Frailty Models

♠ Introduction

- In studies of survival, the hazard function for each individual may depend on a set of risk factors or explanatory variables but usually not all such variables are known or measurable.
- This unknown and unobservable risk factor of the hazard function is often termed as the individual's heterogeneity or *frailty*.
- Frailty models are becoming increasing popular in multivariate survival analysis since they allow us to model the association between the individual survival times within subgroups or *clusters* of subjects.

♠ Proportional Hazards Model with Frailty

- The most common type of frailty model is called the shared-frailty model, which is an extension of the Cox proportional hazards model.
- This model can be derived as follows. Let y_{ij} denote the survival time for the j^{th} subject in the i^{th} cluster, $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, m_i$. Thus, m_i represents the number of subjects in the i^{th} cluster, and therefore we have a total of $N = \sum_{i=1}^{n} m_i$ subjects.
- In the shared frailty model, we assume that the conditional hazard function of y_{ij} given the unobserved frailty random variable w_i for the i^{th} cluster and the fixed covariate vector x_{ij} is given by

$$h(y|w_i, x_{ij}) = h_0(y) w_i \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}), \tag{18}$$

$i = 1, 2, \ldots n$, $j = 1, 2, \ldots, m_i$, where $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown regression coefficients, $h_0(.)$ is an unknown baseline hazard function common to every subject, and $\boldsymbol{x}_{ij}$ is the $p \times 1$ covariate vector for the j^{th} subject in the i^{th} cluster, and may be time dependent.

- A common method is to use a parametric distribution for the frailty w_i. Finite mean frailty distributions such as the gamma and log-normal are very popular in the literature in spite of their theoretical limitations.
- Other alternatives include using infinite mean distributions such as the positive stable distribution.
- The gamma distribution is the most commonly used finite mean distribution to model the frailty term w_i.
- For finite mean frailty distributions, we need the mean of the frailty distribution to be unity in order for the parameters of the model to be identifiable.
- Thus, for the gamma frailty model, the w_i's are assumed to be i.i.d. with

$$w_i \sim \mathcal{G}(\kappa^{-1}, \kappa^{-1}), \tag{19}$$

  where κ is the (unknown) variance of the w_i's.
- Thus, larger values of κ imply greater heterogeneity among clusters.

- Letting $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)'$, we have

$$\pi(\boldsymbol{w}|\kappa) \propto \prod_{i=1}^{n} w_i^{\kappa^{-1}-1} \exp(-\kappa^{-1} w_i). \tag{20}$$

- As mentioned in Anderson et al. (1993), in spite of promising results by several authors, formal and completely satisfactory justifications of these likelihood-based methods await more results on their asymptotic and convergence properties.
- Moreover, none of these likelihood-based methods directly maximize the full likelihood given the data, and the small sample properties of these estimators have yet to be studied.
- Thus, Bayesian methods are attractive for these models since they easily allow an analysis using the full likelihood and inference does not rely on asymptotics.

♠ Weibull Model with Gamma Frailties

- Let ν_{ij} denote the censoring indicator variable, taking value 1 if the j^{th} subject ($j = 1, 2, \ldots, m_i$) of the i^{th} cluster ($i = 1, 2, \ldots, n$) fails and 0 otherwise. Hence, y_{ij} is a failure time if $\nu_{ij} = 1$ and a censoring time otherwise.
- Further, let $\boldsymbol{\nu} = (\nu_{11}, \nu_{12}, \ldots, \nu_{nm_n})'$, $\boldsymbol{y} = (y_{11}, y_{12}, \ldots, y_{nm_n})'$, and $X = (X_1, X_2, \ldots, X_n)$, where X_i is the $m \times p$ matrix of covariates for the i^{th} cluster.
- Let $D = (X, \boldsymbol{\nu}, \boldsymbol{y}, \boldsymbol{w})$ denote the complete data and let $D_{obs} = (X, \boldsymbol{\nu}, \boldsymbol{y})$ denote the observed data.
- Here, we only allow right censored survival data and assume that the censoring is noninformative.
- Let the Weibull baseline hazard function be given by

$$h_0(y_{ij}) = \gamma\alpha y_{ij}^{\alpha-1},$$

where (γ, α) are the parameters of the Weibull distribution.

- The hazard function is given by

$$h(y_{ij}|\boldsymbol{x}_{ij}, w_i) = \gamma\alpha w_i y_{ij}^{\alpha-1}\theta_{ij}, \tag{21}$$

where $\theta_{ij} = \exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta})$, and the complete data likelihood is given by

$$L(\boldsymbol{\beta}, \gamma, \alpha|D) = \prod_{i=1}^{n}\prod_{j=1}^{m_i}\left(\gamma\alpha y_{ij}^{\alpha-1} w_i\theta_{ij}\right)^{\nu_{ij}} \exp\left\{-\gamma y_{ij}^{\alpha}\theta_{ij}w_i\right\}. \tag{22}$$

- The likelihood function of $(\boldsymbol{\beta}, \gamma, \alpha)$ based on the observed data D_{obs} can be obtained by integrating out the w_i's from (22) with respect to the density $\pi(\boldsymbol{w}|\kappa)$ given in (20).
- The observed data likelihood is far too complicated to work with, and thus it is difficult to evaluate the joint posterior distribution of $(\boldsymbol{\beta}, \gamma, \alpha)$ analytically.
- To circumvent this problem, we use the Gibbs sampler to generate samples from the joint posterior distribution.

- Let $\pi(\cdot)$ denote the prior density of its argument and let

$$S = \sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij}^{\alpha} \theta_{ij} w_i. \tag{23}$$

The full conditional distribution of each w_i is a gamma distribution, i.e.,

$$(w_i | \boldsymbol{\beta}, \alpha, \gamma, D_{obs}) \sim \mathcal{G}\left\{ \kappa^{-1} + \sum_{j=1}^{m_i} \nu_{ij}, \kappa^{-1} + \gamma \sum_{j=1}^{m_i} y_{ij}^{\alpha} \theta_{ij} \right\}, \tag{24}$$

for $i = 1, 2, \ldots, n$.

- Letting $\eta = \kappa^{-1}$, the full conditional distribution of η is given by

$$\pi(\eta | \boldsymbol{\beta}, \boldsymbol{w}, D_{obs}) \propto \prod_{i=1}^{n} w_i^{\eta - 1} \eta^{-n\eta} \frac{\exp\left\{-\eta \sum_{i=1}^{n} w_i\right\}}{[\Gamma(\eta)]^n} \pi(\eta). \tag{25}$$

- The full conditional of $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta}|\eta, \boldsymbol{w}, D_{obs}) \propto \exp\left\{\boldsymbol{\beta}' \sum_{i=1}^{n}\sum_{j=1}^{m_i} \nu_{ij}\boldsymbol{x_{ij}} - \gamma S\right\} \pi(\boldsymbol{\beta}). \tag{26}$$

- If a priori $\gamma \sim \mathcal{G}(\rho_1, \rho_2)$, then the full conditional of γ is a gamma distribution given by

$$(\gamma|\boldsymbol{\beta}, \alpha, \boldsymbol{w}, D_{obs}) \sim \mathcal{G}\left\{\rho_1 + \sum_{i=1}^{n}\sum_{j=1}^{m_i} \nu_{ij},\ \rho_2 + S\right\}. \tag{27}$$

- Finally, the full conditional of α is given by

$$\pi(\alpha|\boldsymbol{\beta}, \gamma, \boldsymbol{w}, D_{obs}) \propto \left(\prod_{i=1}^{n}\prod_{j=1}^{m_i} y_{ij}^{\nu_{ij}}\right)^{\alpha-1} \alpha^{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\nu_{ij}} \exp\{-\gamma S\}\, \pi(\alpha). \tag{28}$$

- A priori, it is common to take $\alpha \sim \mathcal{G}(a_1, a_2)$, so that $\pi(\alpha) \propto \alpha^{a_1-1}\exp(-\alpha a_2)$.

- With these choices of priors, it can be shown that each of the above full conditional densities in (24)–(28) is log-concave.
- The frailty model in (21) is a multiplicative frailty model. The modeling strategy used in the `BUGS` software manual is based on an additive frailty model,

$$h(y_{ij}|\boldsymbol{x}_{ij}, b_i) = \xi_{ij}\alpha y_{ij}^{\alpha-1}, \tag{29}$$

where

$$\log(\xi_{ij}) = \zeta + \boldsymbol{x}_{ij}'\boldsymbol{\beta} + b_i, \tag{30}$$

the b_i's are assumed i.i.d. $N\left(0, \kappa^{-1}\right)$, and κ is given a $\mathcal{G}(\phi_1, \phi_2)$ prior. The prior for α is $\mathcal{G}(a_1, a_2)$, and ζ is given a normal prior.

- It is expected that both the multiplicative and additive hazard Weibull model formulations will yield similar inferences since this additive frailty model is actually a multiplicative frailty model with a log-normal frailty distribution.

♠ Gamma Process Prior for $H_0(t)$

- Clayton (1991) derives the Bayesian frailty model assuming a gamma process for the cumulative baseline hazard $H_0(t)$.
- When the gamma process prior on $H_0(t)$ as developed earlier is used, the complete (time-continuous) data likelihood of $(\boldsymbol{\beta}, H_0)$ is given by

$$L(\boldsymbol{\beta}, H_0) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} \left(w_i \exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta}) dH_0(y_{ij})\right)^{\nu_{ij}} \exp\left\{-\exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta}) w_i H_0(y_{ij})\right\}$$

$$\propto \left(\prod_{i=1}^{n} w_i^{d_i}\right) \exp\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i} \nu_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\beta}\right) \left(\prod_{i=1}^{n}\prod_{j=1}^{m_i} (dH_0(y_{ij}))^{\nu_{ij}}\right)$$

$$\times \exp\left(-\sum_{i=1}^{n}\sum_{j=1}^{m_i} \exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta}) w_i H_0(y_{ij})\right). \qquad (31)$$

- As discussed earlier, a grouped data version of (31) is given by

$$L(\boldsymbol{\beta}, \boldsymbol{h}|w, D) = \prod_{j=1}^{J} G_j h_j^{c_0 \alpha_{0j} - 1} \exp(-c_0 h_j), \tag{32}$$

where

$$G_j = \exp(-h_j \sum_{k \in \mathcal{R}_j} \exp(\boldsymbol{x}_k' \boldsymbol{\beta} + \log(w_k)))$$
$$\times \prod_{l \in \mathcal{D}_j} (1 - \exp(-\exp(h_j \boldsymbol{x}_l' \boldsymbol{\beta} + \log(w_l)))),$$

$\boldsymbol{h} = (h_1, h_2, \ldots, h_J)'$, the h_j's are independent, $h_j \sim \mathcal{G}(\alpha_{0j} - \alpha_{0,j-1}, c_0)$ as defined in (8), $\mathcal{R}_j$ is the set of patients at risk, and $\mathcal{D}_j$ is the set of patients having failures in the j^{th} interval as defined earlier.

- The joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{w}|D)$ is given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{w}|D_{obs}) \propto L(\boldsymbol{\beta}, \boldsymbol{h}|D)\pi(\boldsymbol{w})\pi(\boldsymbol{\beta}), \tag{33}$$

  where $\pi(\boldsymbol{\beta})$ is the prior distribution for $\boldsymbol{\beta}$, and $\pi(\boldsymbol{w})$ is given by (19).

- Clayton (1991) assumes a normal prior for $\boldsymbol{\beta}$, although other priors can certainly be used.

- To sample from (33), we need to sample from the following full conditional distributions:

$$\begin{aligned} (i) \quad & [\boldsymbol{\beta}|\boldsymbol{w}, \boldsymbol{h}, D_{obs}]; \\ (ii) \quad & [\boldsymbol{w}|\boldsymbol{\beta}, \boldsymbol{h}, D_{obs}]; \text{ and} \\ (iii) \quad & [\boldsymbol{h}|\boldsymbol{\beta}, \boldsymbol{w}, D_{obs}]. \end{aligned}$$

♠ Piecewise Exponential Model for $h_0(t)$

- Piecewise exponential models provide a very flexible framework for modeling univariate survival data.
- Although in a strict sense it is a parametric model, a piecewise exponential hazard can approximate any shape of a nonparametric baseline hazard.
- Therefore, in practice, defining a prior for a piecewise exponential hazard is the same as defining a prior process for the nonparametric hazard function.
- We construct the piecewise exponential model in the same way discussed earlier.
- We first divide the time axis into J prespecified intervals $I_k = (s_{k-1}, s_k]$ for $k = 1, 2, \ldots, J$, where $0 = s_0 < s_1 < \ldots < s_J < \infty$, s_J being the last survival or censored time and assume the baseline hazard to be constant within intervals.

That is,

$$h_0(y) = \lambda_k, \quad \text{for} \quad y \in I_k. \tag{34}$$

- Although modeling the baseline hazard through an independent increment prior process such as the gamma or beta process is attractive and common, in many applications, it often turns out that prior information is available on the smoothness of the hazard rather than the actual baseline hazard itself.
- For the frailty model, the marginal hazard is always a nonlinear function of the baseline hazard involving covariates and frailty parameters as well. However, the ratio of marginal hazards at the nearby time-points given the same covariates is approximately equal to the ratio of baseline hazards at these points.
- Thus, in these cases, it is more attractive to consider some type of correlated prior process for the baseline hazard.

- To correlate the λ_k's in adjacent intervals, a discrete-time martingale process is used, similar to that of Arjas and Gasbarra (1994) for the univariate survival model.
- Given $(\lambda_1, \lambda_2, \ldots, \lambda_{k-1})$ we specify that

$$\lambda_k | \lambda_1, \lambda_2, \ldots, \lambda_{k-1} \sim \mathcal{G}\left(\alpha_k, \frac{\alpha_k}{\lambda_{k-1}}\right) \tag{35}$$

  for $k = 1, 2 \ldots, J$, where $\lambda_0 = 1$, and $E(\lambda_k | \lambda_1, \lambda_2, \ldots, \lambda_{k-1}) = \lambda_{k-1}$.
- The parameter α_k in (35) controls the amount of smoothness available and a small α_k indicates less information on the smoothing of the λ_k's. If $\alpha_k = 0$, then λ_k and λ_{k-1} are independent.
- When $\alpha_k \to \infty$, then the baseline hazard is the same in the intervals I_k and I_{k-1}: i.e., $\lambda_k = \lambda_{k-1}$. Thus, in the limiting case, we get a constant baseline hazard.

- A version of the above process which can also be used, where $\log(\lambda_k) = \xi_k$ and took

$$\xi_k|\xi_{k-1} \sim N(\xi_{k-1}, \tau^2), \quad k = 1, 2, \ldots, J \tag{36}$$

with $\xi_0 = 0$. Taking this further, we can assume a second difference prior process for ξ_k, i.e.,
$\xi_k|\xi_{k-2}, \xi_{k-1} \sim N(2\xi_{k-1} - \xi_{k-2}, \tau^2), k = 3, 4, \ldots, J$, and so forth.

- A few remarks are in order on the choice of the number of grid intervals J. It is clear that a very large J will make the model nonparametric.
- However, too large a J will produce unstable estimators of the λ_k's and too small a choice may lead to poor model fitting. Hence, a robust choice of J should be considered here.
- Note that the maximum likelihood estimate of λ_k depends on the number of failures, d_k, in the k^{th} interval I_k and is 0 if d_k is zero.
- One advantage of the Bayesian approach with the correlated prior process described here is to smooth out such jumps to zero.

- A random choice of J will make the posterior distribution have a varying dimension and sampling techniques other than the Gibbs sampler, e.g., reversible jump MCMC can be used to compute the posterior distribution.
- The above models can also be easily altered to accommodate monotone baseline hazard functions.
- Suppose that one intends to model the λ_k's with the constraint $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_J$. Then we can assume that

$$\lambda_k - \lambda_{k-1} \sim \mathcal{G}(\alpha_k, \alpha_k),\ k = 1, 2, \ldots, J$$

  instead of (35) or (36).
- The likelihood can now be derived as follows. The j^{th} subject of the i^{th} cluster has a constant hazard of $h_{ij} = \lambda_k \theta_{ij} w_i$ in the k^{th} interval $(k = 1, 2, \ldots, J)$ given the unobserved frailty w_i. If the subject has survived beyond the k^{th} interval, i.e., $y_{ij} > s_k$, the likelihood contribution is $\exp\{-\lambda_k \Delta_k \theta_{ij} w_i\}$, where $\Delta_k = s_k - s_{k-1}$.

- If the subject has failed or is censored in the k^{th} interval, i.e., $s_{k-1} < y_{ij} \leq s_k$, then the likelihood contribution is $(\lambda_k \theta_{ij} w_i)^{\delta_{ij}} \exp\{-\lambda_k (y_{ij} - s_{k-1}) \theta_{ij} w_i\}$. Hence, the complete data likelihood is given by

$$
\begin{aligned}
& L(\boldsymbol{\beta}, \boldsymbol{\lambda} | D) \\
\propto & \prod_{i=1}^{n} \prod_{j=1}^{m_i} \left[\left\{ \prod_{k=1}^{g_{ij}} \exp(-\lambda_k \Delta_k \theta_{ij} w_i) \right\} \left(\lambda_{g_{ij}+1} \theta_{ij} w_i \right)^{\delta_{ij}} \right. \\
& \left. \times \exp\left\{ -\lambda_{g_{ij}+1} (y_{ij} - s_{g_{ij}}) \theta_{ij} w_i \right\} \right],
\end{aligned}
$$

where g_{ij} is such that $y_{ij} \in (s_{g_{ij}}, s_{g_{ij}+1}] = I_{g_{ij}+1}$.

♠ Example 1.3: Kidney Infection Data

- We consider the following four models to fit the kidney infection data:

  Model I: Piecewise exponential model with gamma priors for the λ_k's as in (35).

  Model II: Weibull baseline hazard with multiplicative gamma frailties.

  Model III: Piecewise exponential baseline hazard with normal priors for the $\log(\lambda_k)$'s as in (36).

  Model IV: Weibull baseline hazard with additive frailties.

- The proportional hazard's component of each of the above models is

$$\theta_{ij} = \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}) = \exp(\beta_{sex} sex_i + \beta_{age} age_{ij}),$$

  where $sex_i = 1$ if the i^{th} patient is a female and 0 otherwise, age_{ij} is the age at the j^{th} infection of the i^{th} patient.

- The `BUGS` codes and datasets for these models are given at the website "`http://www.stat.uconn.edu/~mhchen/survbook`."
- The following values of hyperparameters were used in this example. For the prior on κ^{-1}, we take $\phi_1 = \phi_2 = 0.001$. Each component of $\boldsymbol{\beta}$ was assumed a priori normal with 0 mean and variance 10^3. The same prior was assumed for ζ in Model IV. For Model I all the α_k's were assumed to be 0.01.
- For Model III, τ^2 was taken as 10^4 to make it comparable with the corresponding prior precision for the λ_k's in Model I. For Models III and IV we took $\rho_1 = \rho_2 = 0.001$, and $a_1 = a_2 = 0.001$.
- We first investigated different choices of the grid size J for Models I and III. We experimented with three choices of $J = 5$, 10, and 20.
- The $J = 5$ case seemed to give worse model fitting than the $J = 10$ case and the last choice of J did not provide substantially better results than the $J = 10$ case. Hence we decided to use $J = 10$ throughout.

- Model fitting and/or model choice were not very sensitive to small variations on the values of the other hyperparameters as given above.
- Widely different α_k's in Model I did change the estimates a little bit. However, that did not alter the model choice ordering as reported below.
- Table 1.4 shows the posterior mean, standard deviation, and 95% credible intervals for β_{sex}, β_{age}, κ. We show the estimates of α and γ (for Model IV $\mu = \exp\{\zeta\}$) in Table 1.5.
- In both Tables 1.4 and 1.5, posterior means are followed by standard deviations in the first row, and 95% credible intervals are shown in the second row.
- The estimates of β_{sex} show that the female patients have a slightly lower risk for infection.

- The estimates of κ from different models show that there is strong posterior evidence of a high degree of heterogeneity in the population of patients. Some patients are expected to be very prone to infection compared to others with the same covariate value.
- This is not very surprising, as in the dataset there is a male patient with infection times 8 and 16, and there is also another male patient with infection times 152 and 562.
- The high posterior means of κ also provide evidence of a strong positive association between two infection times for the same patient.
- The above analysis suggests that Models I and III are very close to each other while Models II and IV are also somewhat similar.

TABLE 1.4. Parameter Estimates from Different Models for Kidney Infection Data

	β_{sex}	β_{age}	κ
Model I	-1.493 (0.468)	0.006 (0.013)	0.499 (0.283)
	$(-2.430, -0.600)$	$(-0.018, 0.032)$	(0.061, 1.160)
Model II	-1.888 (0.564)	0.007 (0.013)	0.585 (0.307)
	$(-3.034, -0.846)$	$(-0.018, 0.032)$	(0.115, 1.317)
Model III	-1.500 (0.480)	0.007 (0.013)	0.523 (0.285)
	$(-2.467, -0.624)$	$(-0.018, 0.036)$	(0.089, 1.195)
Model IV	-1.69 (0.529)	0.006 (0.014)	0.816 (0.507)
	$(-2.780, -0.699)$	$(-0.019, 0.036)$	(0.079, 2.050)

TABLE 1.5. Parameter Estimates of α and μ for Weibull Models II and IV for Kidney Infection Data

	Model II	Model IV
α	1.278 (0.190)	1.22 (0.160)
	(0.937, 1.692)	(0.916, 1.540)
μ	0.016 (0.015)	0.013 (0.014)
	(0.001, 0.058)	(0.001, 0.053)

♠ Positive Stable Frailties

- The choice of the gamma distribution for frailties arises partly for mathematical convenience, since this distribution is conjugate to the likelihood for w_i and partly because, in the case of bivariate survival data without covariates, integration over the unknown frailty yields a class of bivariate survival time distributions with appealing properties.
- This choice also has some less desirable consequences, however. The marginal relationship between the hazard and covariates no longer follows the proportional hazards model, since the marginal hazard function $h(y|\boldsymbol{x})$ is given by

$$h(y|\boldsymbol{x}) = \frac{h_0(y)\exp(\boldsymbol{x}'\boldsymbol{\beta})}{\kappa H_0(y)\exp(\boldsymbol{x}'\boldsymbol{\beta}) + 1}. \tag{37}$$

- Instead, we see in (37) that there is a convergence of hazards, at a rate determined by κ.

- In the multivariate case with covariates, this property has the consequence that information for estimation of κ comes partly from the coincidence of failure within clusters, and partly from the marginal convergence of hazards in relation to the covariates.
- Several authors have pointed out that this is not a desirable property for the multivariate model since it renders interpretation of κ difficult.
- This problem persists with any other finite mean frailty distribution, such as the log-normal.
- The assumption of a positive stable distribution of the w_i's avoids this problem, since the proportional hazards assumption for covariates then remains true marginally.

- Also, in the finite mean frailty model, the unconditional effect of a covariate, which is measured by the hazard ratio between unrelated subjects (i.e., with different frailties) is always less than its conditional effect, measured by the hazard ratio among subjects with the same frailty.
- In particular, suppose we consider two subjects from different clusters with respective covariates $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Let $S_1(y)$ and $S_2(y)$ denote the corresponding unconditional survivor functions derived under a frailty specification.
- The covariate effects, as measured by the hazard ratio, are always attenuated and further, $S_1(y)$ and $S_2(y)$ are usually not related via a proportional hazards model.
- The degree of attenuation of the core effect is not easy to quantify unless both conditional and marginal specifications have a proportional hazards structure.

- If the frailty distribution is an infinite variance positive stable distribution with Laplace transform

$$E(\exp(-sw)) = \exp(-s^{\alpha})$$

for $0 < \alpha < 1$, then $S_1(y)$ and $S_2(y)$ will have proportional hazards since

$$S(y|\boldsymbol{x}) = \exp\left\{-(\theta H_0(y))^{\alpha}\right\} = \exp\left\{-\exp(\alpha \boldsymbol{x}'\boldsymbol{\beta}) H_0^{\alpha}(y)\right\}, \qquad (38)$$

where $H_0(y)$ denotes the cumulative baseline hazard function.

- From (38), it is easy to quantify the attenuation of the covariate effect, since the parameter α is clearly the attenuation of $\boldsymbol{\beta}$ in the marginal hazard.

- Thus, we no longer need to choose between conditional and unconditional Cox model specifications, since a single specification can be interpreted either way.

- Although the positive stable frailty model is conceptually simple, estimation of the resulting model parameters is complicated due to the lack of a closed form expression for the density function of a stable random variable.
- The Bayesian framework for this model using MCMC methods offers an attractive alternative to frequentist methods. Specifically, it greatly reduces the difficulty in estimating α, the parameter that characterizes the tail behavior in stable distributions.

♠ Bivariate Measures of Dependence

- Oakes (1989) proposes a local measure of dependence, given by

$$\theta^*(y_1, y_2) = \frac{S(y_1, y_2)\Delta_1\Delta_2 S(y_1, y_2)}{(\Delta_1 S(y_1, y_2))(\Delta_2 S(y_1, y_2))}, \tag{39}$$

  where Δ_j denotes the operator $-\frac{\partial}{\partial y_j}$, $j = 1, 2$.
- This function, which was introduced by Clayton (1978), may be interpreted as the ratio of the conditional distribution of Y_1 given $Y_2 = y_2$, to that of Y_1 given $Y_2 > y_2$.

- This measure has proved to be a useful quantity for assessing the degree of association between survival times for a given frailty distribution.
- For a model with gamma frailties, that is, $w \sim \mathcal{G}(\kappa^{-1}, \kappa^{-1})$, it can be shown that the joint survival function of (Y_1, Y_2) is given by

$$S(y_1, y_2) = \frac{1}{\left[\frac{1}{\kappa} - \log(S_1(y_1)) - \log(S_2(y_2))\right]^{1/\kappa}},$$

where $S_1(y_1)$ and $S_2(y_2)$ are the baseline survival functions. For this model, it can be shown that $\theta^*(y_1, y_2) = 1 + \kappa$, so that $\theta^*(y_1, y_2)$ is free of (y_1, y_2) for this model. Also, $\theta^*(y_1, y_2) = 1 + \kappa$ implies that the larger the κ, the larger the correlation between Y_1 and Y_2, and thus this measure of dependence sheds light on the role of κ when using gamma frailties. In particular, when $\kappa \to 0$, $\theta^*(y_1, y_2) \to 1$, indicating independence between Y_1 and Y_2.

- For a model with positive stable frailties, the Laplace transform of w is $E(\exp(-sw)) = \exp(-s^{\alpha})$, and the bivariate survival function is

$$S(y_1, y_2) = \exp\left\{-\left[-\log(S_1(y_1)) - \log(S_2(y_2))\right]^{\alpha}\right\}. \qquad (40)$$

The parameter α $(0 < \alpha < 1)$ is a scalar parameter that is a measure of association between (Y_1, Y_2). Small values of α indicate high association between (Y_1, Y_2). As $\alpha \to 1$, this implies less association between (Y_1, Y_2) which can be seen from (40). For (40),

$$\theta^*(y_1, y_2) = \alpha^{-1}(1-\alpha)\left[-\log(S_1(y_1)) - \log(S_2(y_2))\right]^{-\alpha} + 1. \qquad (41)$$

Thus (41) decreases in (y_1, y_2). Therefore, the association between (y_1, y_2) is greater when (y_1, y_2) are small and the association decreases over time. As $\alpha \to 0$, Y_1 and Y_2 achieve maximal dependence. When $\alpha \to 1$, $\theta^*(y_1, y_2) \to 1$, indicating independence between Y_1 and Y_2.