

HIERARCHICAL MODELLING OF MULTIVARIATE SURVIVAL DATA

By

Paul Antony Gustafson

B. Sc. (Mathematics) University of British Columbia

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
DEPARTMENT OF STATISTICS

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

1991

© Paul Antony Gustafson, 1991

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of Statistics

The University of British Columbia
Vancouver, Canada

Date Dec. 9, 1991.

Abstract

Hierarchical models based on conditional independence are investigated as a means of modelling multivariate survival times. The model structure follows Clayton (1978), Hougaard (1986b), and Oakes (1986,1989). Both approximate Bayesian and maximum likelihood estimation in these models is investigated via simulation. Predicting a component of a response vector on the basis of other components of the response and other vectors is also studied, using Bayes and empirical Bayes methods. Application to real data is detailed, using the techniques of estimation and prediction discussed.

Table of Contents

Abstract	ii
List of Tables	vi
List of Figures	vii
Acknowledgement	viii
1 Introduction	1
2 Models and Techniques	3
2.1 Conditionally Independent Hierarchical Models	3
2.2 Hazard Rate and Frailty	4
2.3 Multivariate Frailty Model as a CIHM.	6
2.4 The Weibull Model	9
2.5 Properties as a One-Stage Model	10
2.6 Bayesian Inference	11
2.7 Using Laplace's Method to Approximate Posterior Quantities	12
3 Estimation and Simulation	14
3.1 Estimating the Common Parameter	14
3.2 Prior Information	15
3.3 Reparameterizations	15
3.4 Weibull/Gamma Simulation Study	17

3.5	Weibull/Positive Stable Simulation Study	20
4	Prediction	24
4.1	Parametric Empirical Bayes Approach to Prediction	25
4.2	Full Bayesian Approach to Prediction	26
4.3	The PEB Approach in the Weibull Model	27
4.4	The Full Bayesian Approach in the Weibull Model	28
4.5	Full versus PEB Prediction	29
4.5.1	Exponential Prediction Example	30
4.6	Prediction in the Weibull/Gamma Model	32
4.7	Prediction in the Weibull/Positive Stable Model	33
4.8	Effect of Frailty Distribution on Prediction	36
5	Application	39
5.1	Description of Rat-Litter Data	39
5.2	Single Risk versus Competing Risks	39
5.3	Model Adequacy Checking	40
5.4	Single Risk : Tumour Development	42
5.5	Single Risk : Death due to Other Causes	44
5.6	Competing Risks : Single Frailty	44
5.7	Competing Risks : Two Frailties	45
5.8	Including Covariates	47
5.9	Assessing Model Choice	48
6	Discussion	53
6.1	Approaches to Frailty Models	53
6.2	Specification of Distributions	54

6.3	Techniques	56
6.4	Assessing Models	57
Appendices		60
A	Quadrant Dependence of Frailty Models	60
B	Female Rat Litter Data	61
C	Male Rat Litter Data	62
D	Performance of Weibull/Gamma Model on Diagnostic Plot	63
E	Derivatives of Survivor Function in Two-Frailty Model	64
Bibliography		66

List of Tables

3.1	Comparison of estimators for 99 simulated samples of 50 observations in the Weibull/gamma model	18
3.2	Comparison of estimators for 100 simulated samples of 200 observations in the Weibull/gamma model	19
3.3	Comparison of estimators for 100 simulated samples of 50 observations in the Weibull/positive stable model	21
3.4	Comparison of estimators for 100 simulated samples of 200 observations in the Weibull/positive stable model	22
4.5	Probabilities of intervals for $\log L$ in exponential example	32
5.6	Estimation with rat-litter data : single risk (tumour)	42
5.7	Estimation with rat-litter data : single risk (other causes)	44
5.8	Estimation with rat-litter data : competing risks, single frailty	46
5.9	Estimation with rat-litter data : competing risks, two frailties	48
5.10	Estimation with rat-litter data : single risk (other cause), proportional hazards model	49
5.11	Predictive scores	51

List of Figures

4.1	Predictive density ratio as a function of Y (exponential example)	31
4.2	Approximate full and empirical predictive densities (gamma frailty) . . .	34
4.3	Histograms of observed predictive density ratios: approximate full versus empirical Bayes	35
4.4	Approximate full and empirical predictive densities (stable frailty)	37
5.5	Diagnostic plots based on Kaplan-Meier estimates of the marginals . . .	43
5.6	Histograms of predictive density ratios : positive stable frailty versus gamma frailty	50

Acknowledgement

I would like to thank Mohan Delampady for his guidance and supervision throughout the development of this thesis. Additionally, I am grateful to Harry Joe and James Zidek for their careful reading of the manuscript, and their subsequent suggestions. Finally, I would like to thank my wife, Réka Vásárhelyi, for her continued patience and support.

Chapter 1

Introduction

Often in experimental sciences and engineering, the time taken for objects or entities to fail or die is studied. The statistical methods and models used to interpret the resultant data are commonly referred to under the broad heading of survival analysis. Kalbfleish & Prentice (1980), Lawless (1982), and Cox & Oakes (1984) cover the main ideas of survival analysis and act as guides to the large body of relevant literature.

A small fraction of this literature deals with survival analysis within the Bayesian paradigm. Some such works are Cornfield & Detre (1977), Martz & Waller (1982), Kalbfleish (1985), and Gammerman & West (1987). Another small fraction is concerned with multivariate survival analysis. In this context, by multivariate we mean situations where several survival times somehow belong to a single cohesive unit or family. Holt & Prentice (1974) and Wild (1983) are two works in this category. Also related are works on the reliability of systems of several components, such as Lindley & Singpurwalla (1986).

The literature is much sparser in the area of intersection of these two categories (which we can call Bayesian multivariate survival analysis). In fact, there is apparently no literature dealing with multivariate survival from an explicitly Bayesian perspective. However, the type of multivariate survival models developed and used by Clayton (1978), Hougaard (1986b), and Oakes(1986,1989), though developed from a frequentist perspective, have an implicit hierarchical Bayes structure. We aim to bring this Bayesianity to the fore in our investigation of these models. For completeness, it should be noted that there are hierarchical structures for multivariate survival other than the one we will investigate.

For an example, see Lindley & Singpurwalla (1986).

To meet our aim, in Chapter 2 we introduce the model from the point of view of a hierarchical model as well as from a frailty model perspective. We also introduce relevant aspects of Bayesian inference. Chapter 3 is concerned with inference, using simulation to examine techniques. Then, Chapter 4 focusses on the problem of predicting a survival time given the survival times of other components of the same unit. The ideas of Chapters 3 and 4 are put to use in the analysis of a real data set, contained in Chapter 5. The data is lifetimes of rats—for each of a number of litters, a multivariate response (lifetimes of litter members) is observed. The hierarchical frailty model is employed as an attempt to account for factors common to a litter. The situation is complicated by three possible mechanisms for terminating individual observations (tumour appearance, sacrifice, natural death), as well as a treatment suspected to promote tumour appearance. Finally, Chapter 6 attempts to address some of the issues raised in the first five chapters.

Chapter 2

Models and Techniques

2.1 Conditionally Independent Hierarchical Models

A Bayesian model is hierarchical if there are several stages of information. A two-stage model would involve a response Y coming from a density $f(y | \theta)$. In turn θ would come from a density $g(\theta | \lambda)$. In general, θ would be referred to as the parameter, while λ is the hyperparameter. The distribution g on θ could represent prior information or structural information. Additional prior information could be specified in the form of a completely specified distribution on λ . Alternately, λ could be estimated from the data. Usually, the investigator would wish to make inferences about the parameter θ , although there is nothing to prevent making inferences about the hyperparameter λ .

Kass & Steffey (1989) formalize the class of conditionally independent hierarchical models (CIHM's) as follows. Say we have observation vectors Y_i of length n_i , $i = 1, \dots, k$, parameterized by $\theta_i \in \Theta$, $i = 1, \dots, k$ and by $\lambda \in \Lambda$, where Θ and Λ have dimension q and r respectively. Then we have a CIHM if:

- Stage 1. conditionally on $(\theta_1, \dots, \theta_k)$ and λ , the Y_i are independent with densities $f(y_i | \theta_i, \lambda)$ ($i = 1, \dots, k$) belonging to a family $\{f(y | \theta, \lambda) : \theta \in \Theta, \lambda \in \Lambda\}$;
- Stage 2. conditionally on λ , $\theta_1, \dots, \theta_k$ are independent and identically distributed with density from $\{g(\theta | \lambda) : \lambda \in \Lambda\}$.

We can think of observing k units or families, with the i^{th} unit consisting of n_i components or individuals. Then the θ_i 's are unit-specific parameters, while λ is a parameter

common to all units. A fully Bayesian approach to this model would incorporate a prior distribution on λ . Then given data, one could in principle determine the joint posterior distribution of any subset of $(\theta, \lambda) = (\theta_1, \dots, \theta_k, \lambda_1, \dots, \lambda_r)$.

Of course such a model can be collapsed to a one-stage model parameterized by λ . Then $Y = (Y_1, \dots, Y_k)$ has density given by:

$$p(y \mid \lambda) = \prod_{i=1}^k E_{\lambda}^{\theta_i} f(y_i \mid \theta_i, \lambda) \quad (2.1)$$

where the expectation is with respect to the distribution on θ_i , and thus depends on λ . If we were only interested in the θ_i 's, we could take a parametric empirical Bayes (PEB) approach. This would involve finding a maximum likelihood estimate $\hat{\lambda}$ by maximizing (2.1), and then proceeding as if $\hat{\lambda}$ were the true value of λ . This approach is discussed in Kass & Steffey (1989). Of course if the n_i 's are small, we cannot hope to get useful information about individual θ_i 's. In this case we can only hope for information about λ (providing k is suitably large). So in some cases we can only expect to get information about the distribution of the unit-specific parameters, rather than information about the parameters themselves.

Before we employ CIHM's, we wish to introduce the notion of frailty. Hopefully, this will lend a physical interpretation to the model structure.

2.2 Hazard Rate and Frailty

Often when working with distributions on the positive reals, and especially when studying survival times, distributions are characterized by their hazard rate or hazard function. If a continuous non-negative random variable has distribution function F , then its hazard rate h is given by

$$h(t) = -\frac{d}{dt} \log[1 - F(t)]$$

A common method of including covariates in a hazard rate model was introduced by Cox (1972). If an individual has observed covariate vector z then his hazard rate is assumed to be

$$h(t \mid z) = g(z^t \beta) h_0(t)$$

where β is an unknown parameter vector, g is a fixed function (usually the exponential function), and h_0 is an unknown function, the baseline hazard.

Modelling concomitant variables, as above, is a way to account for heterogeneity in a population. However, in many situations it is reasonable to expect that a population has sources of heterogeneity not accounted for by covariates. For this reason, Vaupel, Manton & Stallard (1979) introduced an unobservable positive scalar quantity, frailty, in a hazard rate model. If an individual has frailty θ , then his hazard is modelled as

$$h(t \mid \theta) = \theta h_0(t)$$

where again h_0 is a baseline hazard common to a population. As in the Cox model, the effect on the hazard is multiplicative. So one can think of frailty as an unobserved covariate. (To be consistent with the Cox model, frailty would be a function of an unobserved covariate vector.) A model with covariates and frailty is also reasonable, both would have a multiplicative effect on the baseline hazard.

If one makes no assumptions about the frailty of individuals in a population, it is unreasonable to expect to make inferences about the population. Hougaard (1984, 1986a) suggests assuming that within a population, frailty is distributed as some parametric family. A gamma distribution is mathematically convenient, and has been used in several studies (e.g. Clayton, 1978). Hougaard presents arguments in favour of a stable distribution of frailties. Other distributions could be used; however, as we shall see, only distributions with closed form Laplace and inverse Laplace transforms (such as the gamma and positive stable distributions) will be computationally feasible.

It is also clear that if the frailties in a population are all independent realizations from a distribution then unless very strong assumptions are made, it will be difficult to identify the frailty distribution in the presence of the baseline hazard, and vice-versa. There are, however, situations where one expects that certain individuals will share a common frailty. For instance it would be reasonable to assume that members of a family have the same frailty, and that conditionally on this frailty the family member lifetimes are independent. In such a context the frailty might be a function of genetic and environmental factors (Hougaard, 1986a). With two or more individuals per frailty value, it becomes reasonable to estimate both the frailty distribution, and the (parameterized) baseline hazard.

2.3 Multivariate Frailty Model as a CIHM.

A model with observations on families, as mentioned above, fits into the framework of a CIHM. If $Y_i = (Y_{i1}, \dots, Y_{in_i})$ is the vector of lifetimes from the i^{th} family, then we take θ_i to be the family frailty. We partition $\lambda = (\lambda^{(a)}, \lambda^{(b)})$ so that $\lambda^{(a)}$ parameterizes the frailty distribution, while $\lambda^{(b)}$ parameterizes the baseline hazard.

In general we might expect to have several baseline hazards; that is, we might expect our observations to be classified as coming from various groups. For instance in a clinical trial we might have a treatment group and a control group, or with data on fathers and sons we would treat fathers as one group and sons as another group. If we have p groups, then we can specify the first stage of the model by specifying functions h_1, \dots, h_p such that $Y_{ij} \mid \theta_i$ has hazard

$$\theta_i h_k(y_{ij} \mid \lambda^{(b)}) \tag{2.2}$$

when Y_{ij} belongs to the k^{th} group. Note that since θ_i is positive, (2.2) will be a valid hazard function precisely when h_k is a valid hazard function.

Because we will be dealing with survival times and failure times, we wish to allow for right-censored observations. In the usual way, we introduce an indicator d_{ij} taking value 0 if Y_{ij} is censored at y_{ij} , and 1 if $Y_{ij} = y_{ij}$ is observed. We assume that given a family's frailty, censoring arises in a way that makes the likelihood of independent observations a product of densities and survivor functions. This is true under fairly weak conditions—for a discussion, see Lawless (1982, pp. 31-44). Then, suppressing the dependence of the h_k 's on $\lambda^{(b)}$, the first-stage likelihood is

$$L(\theta, \lambda \mid y, d) = \prod_{i=1}^k L(\theta_i, \lambda \mid y_i, d_i) \quad (2.3)$$

where

$$L(\theta_i, \lambda \mid y_i, d_i) = \prod_{j=1}^{n_i} \left\{ \left[\theta_i h_{k_{ij}}(y_{ij}) \exp(-\theta_i H_{k_{ij}}(y_{ij})) \right]^{d_{ij}} \left[\exp(-\theta_i H_{k_{ij}}(y_{ij})) \right]^{1-d_{ij}} \right\} \quad (2.4)$$

in which k_{ij} denotes the group in which Y_{ij} belongs, while $H_k(y) = \int_0^y h_k(t) dt$. Simplifying and integrating with respect to the θ_i 's gives the unconditional (collapsed) likelihood

$$L(\lambda \mid y, d) = \prod_{i=1}^k \left\{ \left[\prod_{\{j: d_{ij}=1\}} h_{k_{ij}}(y_{ij}) \right] E_{\lambda^{(a)}}^{\theta_i} \left[\theta_i^{D_i} \exp(-\theta_i H_i(y_i)) \right] \right\} \quad (2.5)$$

where $D_i = \sum_{j=1}^{n_i} d_{ij}$, and $H_i(y_i) = \sum_{j=1}^{n_i} H_{k_{ij}}(y_{ij})$.

One possibility for the second-stage is that the θ_i 's come from a gamma distribution. Note that in this case, (2.5) is relatively easy to evaluate. However, if the gamma distribution has unconstrained shape and scale parameters, one might expect an identifiability problem with the likelihood (2.5) (this is certainly the case with the Weibull model of the next section). A natural restriction is to fix the mean frailty at 1. This is the approach taken by Clayton (1978). Hence an individual with average frailty has baseline hazard, and scale is accounted for in the first stage of the model. We choose to parameterize the gamma frailty distribution so that $\lambda^{(a)} = \lambda_0$, and the frailty density is

$$g(\theta) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\alpha\theta) \quad (2.6)$$

where $\alpha = -(\log \lambda_0)^{-1}$, and $\lambda_0 \in (0, 1]$. This parameterization has the advantage that λ_0 lies in a finite interval, with the important special case of unconditional independence within families occurring at an endpoint ($\lambda_0 = 1$). It can also be shown that the limiting case as λ_0 decreases to 0 is the Fréchet upper bound (perfect positive dependence) (Oakes, 1989). Assuming (2.6), (2.5) can be evaluated using

$$E_{\lambda_0}^{\theta} [\theta^D \exp(-H\theta)] = \frac{\Gamma(D + \alpha)}{\Gamma(\alpha)} \frac{\alpha^{\alpha}}{(H + \alpha)^{(D+\alpha)}} \quad (2.7)$$

where again $\alpha = -(\log \lambda_0)^{-1}$.

Hougaard's suggestion is a positive stable distribution of frailties. These distributions are characterized by their Laplace transforms (moment generating functions). Specifically,

$$E_{\lambda_0}^{\theta} [\exp(-s\theta)] = \exp(-s^{\lambda_0}), \quad \lambda_0 \in (0, 1] \quad (2.8)$$

The stable distributions have long right-tails. In fact, when $\lambda_0 < 1$, the k^{th} moment of the variate defined in (2.8) exists if and only if $k < \lambda_0$, and so in particular, the mean is infinite. The density is complicated, given by

$$g(\theta \mid \lambda_0) = \frac{1}{\pi\theta} \sum_{k=1}^{\infty} \frac{\Gamma(k\lambda_0 + 1)}{k!} (-\theta^{-\lambda_0})^k \sin(\lambda_0 k\pi)$$

A good discussion of stable variates is found in Feller (1971, pp. 169-76).

As in the gamma case, $\lambda_0 = 1$ means the distribution is degenerate at $\theta = 1$, and we have independence within families, while as λ_0 decreases to 0, we obtain the Fréchet upper bound (Oakes, 1989). We can evaluate (2.5) by differentiating (2.8) D times with respect to s . The differentiation gives a recursive form for (2.5) (Hougaard, 1986b). In particular,

$$E_{\lambda_0}^{\theta} [\theta^D \exp(-H\theta)] = (\lambda_0 H)^D \exp(-H^{\lambda_0}) R(\lambda_0, D, H) \quad (2.9)$$

where

$$R = \begin{cases} 1 & \text{if } D = 0 \\ \sum_{m=0}^{D-1} C_{D,m} H^{-m\lambda_0} & \text{otherwise} \end{cases}$$

and

$$C_{k,m} = \begin{cases} 1 & m = 0; \\ C_{k-1,m} + C_{k-1,m-1}[(k-1)\lambda_0^{-1} - (k-m)] & m = 1, \dots, k-2; \\ \lambda_0^{1-k} \frac{\Gamma(k-\lambda_0)}{\Gamma(1-\lambda_0)} & m = k-1. \end{cases}$$

A potentially important difference between gamma frailty and positive stable frailty has been noted by Hougaard (1984,1986a). If a population has frailty distribution given by density $g(\theta)$, then one can consider the distribution of frailties amongst those individuals living at time $t > 0$. We denote this distribution $g_t(\theta)$. For any second-stage distribution, with gamma frailty the coefficient of variation of g_t is constant in t , while with positive stable frailty, the coefficient of variation is decreasing in t . (Hougaard (1984) suggests that a normalized measure of variability, like the coefficient of variation of the frailty distribution, is an appropriate measure of heterogeneity.) So, in this sense, the population becomes more homogeneous with time under stable frailty, but not under gamma frailty. Since it seems sensible that the sub-population of long-lived individuals be more homogeneous than the whole population, this supports the use of stable frailty rather than gamma frailty.

So, gamma and stable distributions are candidates for the second stage of the model. A possible first stage is suggested in the next section.

2.4 The Weibull Model

A useful parametric family for survival times is the Weibull distributions. Following Hougaard (1986b), we introduce a two-group model, with groups 1 and 2 being treatment and control groups respectively. We take $\lambda_{(b)} = (\lambda_1, \lambda_2, \lambda_3)$, and specify the conditional hazards by

$$h_1(y \mid \lambda_1, \lambda_3) = \lambda_1(\lambda_3 y^{\lambda_3-1})$$

$$h_2(y \mid \lambda_2, \lambda_3) = \lambda_2(\lambda_3 y^{\lambda_3 - 1}) \quad (2.10)$$

So the groups have Weibull conditional hazards, with common shape parameter, but different scale parameters. Under both gamma and stable frailty, this leads to an identifiable unconditional likelihood.

2.5 Properties as a One-Stage Model

The properties of a frailty model treated as a multivariate distribution have been studied by several authors. Often they arise in the context of bivariate distributions with given marginals (these can be classified by bivariate distributions with uniform marginals, often referred to as copulas—see Genest and MacKay, 1986a,b).

If the marginal survivor functions of T_1 and T_2 are specified as S_1 and S_2 respectively, then, as noted in Oakes (1989), the joint survivor function $S(t) = P(T_1 > t_1, T_2 > t_2)$ is given by

$$S(t) = p[q(S_1(t_1)) + q(S_2(t_2))] \quad (2.11)$$

where p is the Laplace transform of the frailty density, and q is the inverse function of p . Oakes (1989) points out that this is a subclass of the Archimedian bivariate distributions considered by Genest & MacKay (1986a,b). For another generalization, see Marshall & Olkin (1988).

There are of course a number of possible measures of the association between T_1 and T_2 . These measures tend to classify the association as positive, in the sense that T_1 and T_2 have a tendency to be simultaneously large or simultaneously small. A useful characterization is the notion of quadrant dependence given in Lehmann (1966). (T_1, T_2) are positively quadrant dependent if

$$S(t) \geq S_1(t_1)S_2(t_2) \quad \forall t_1, t_2 \quad (2.12)$$

It is possible to show that if the log of the Laplace transform of the frailty density is convex then the bivariate frailty model is positively quadrant dependent for any choice of second-stage distributions (Appendix A). From (2.7) and (2.8) we see this convexity condition holds for both gamma and positive stable frailty. By results in Lehmann (1966), this implies that the following measures of dependence/association are non-negative ¹ : $Cov(T_1, T_2)$ (provided it exists), Kendall's τ (Kendall, 1938), Spearman's ρ_s (Spearman, 1904), and Blomqvist's quadrant measure (Blomqvist, 1950).

Oakes (1989) suggests using Kendall's τ in frailty models as it depends only on the second-stage (frailty) distribution. The definition is

$$\tau = E \left\{ \text{sign} \left[(T_1^{(1)} - T_1^{(2)})(T_2^{(1)} - T_2^{(2)}) \right] \right\}$$

where $T^{(i)} (i = 1, 2)$ are independent copies of T . Genest & MacKay (1986b) show that a model of the form (2.11) satisfies

$$\tau = 4 \int_0^\infty sp(s)p''(s)ds - 1 \quad (2.13)$$

With our parameterizations, under gamma frailty $\tau = (2\alpha + 1)^{-1}$, where $\alpha = -(\log \lambda_0)^{-1}$, while under positive stable frailty, $\tau = 1 - \lambda_0$ (Oakes, 1989). Estimating τ from data and subsequently estimating the frailty distribution parameter is discussed in Oakes (1986). The same author suggests a method of assessing the frailty distribution choice without specifying conditional hazards (Oakes, 1989).

2.6 Bayesian Inference

Of course from a one-stage model perspective, we can only make inferences about λ . Any Bayesian inference about λ will involve finding the posterior distribution of λ . If the prior

¹This is actually a special case of a result in Genest & MacKay (1986a) concerning a stochastic ordering (concordance ordering) on copulas. Our case is obtained by comparing our copula to the independence copula. The condition on the log of the Laplace transform can be weakened to superadditivity.

information about λ is given by distribution $\Pi(\lambda)$, then the posterior is given by

$$\Pi(\lambda \mid y) = \frac{\left[\prod_{i=1}^k L(\lambda \mid y_i) \right] \Pi(\lambda)}{\int \left[\prod_{i=1}^k L(\lambda \mid y_i) \right] \Pi(\lambda) d\lambda}$$

The mean and variance of a component λ_i with respect to the posterior are known as the posterior mean and posterior variance of λ_i , and are good summaries of the investigator's post-experimental knowledge about λ_i . Alternatively, the posterior distribution of λ_i can be computed according to

$$\Pi(\lambda_i \mid y) = \int \int \dots \int \Pi(\lambda \mid y) d\lambda_1 \dots d\lambda_{i-1} d\lambda_{i+1} \dots d\lambda_r$$

This distribution can be plotted for a visual display of the knowledge of λ_i . Of course if λ were 1 or 2-dimensional, then the complete posterior could be plotted. Unfortunately, this is seldom the case. Bayesian analysis commonly requires the use of approximation techniques.

2.7 Using Laplace's Method to Approximate Posterior Quantities

Tierney & Kadane (1986) discuss the use of Laplace's method to approximate posterior quantities. For example if a model with k observations is parameterized by vector λ , with log-likelihood $\mathcal{L}(\lambda)$ and prior $\pi(\lambda)$, then the posterior mean of a function $g(\lambda)$ can be written

$$E(g(\lambda) \mid \text{data}) = \frac{\int \exp(kL^*(\lambda)) d\lambda}{\int \exp(kL(\lambda)) d\lambda}$$

where $L(\lambda) = k^{-1}(\log \pi(\lambda) + \mathcal{L}(\lambda))$ and $L^*(\lambda) = k^{-1}(\log g(\lambda) + \log \pi(\lambda) + \mathcal{L}(\lambda))$. The Laplace approximation amounts to replacing both numerator and denominator integrands with second-order Taylor approximations about their respective global maximas. This

results in integrals of unnormalized multivariate normal density functions. The approximation becomes

$$\hat{E}(g(\lambda) \mid \text{data}) = \left(\frac{\det \Sigma^*}{\det \Sigma} \right)^{1/2} \exp\{k(L^*(\hat{\lambda}^*) - L(\hat{\lambda}))\}$$

where $\hat{\lambda}$ and $\hat{\lambda}^*$ maximize L and L^* respectively, and Σ and Σ^* are the respective negative inverse Hessians evaluated at these maxima. The thrust of Tierney and Kadane's argument is that when g is a positive function, the numerator and denominator integrands are of similar shape. Thus the two Laplace approximations make similar errors, and so in taking the ratio the error is reduced. In fact while a single integral approximation has $O(k^{-1})$ error, the ratio approximation has $O(k^{-2})$ error.

This method could be used to estimate the posterior means and variances of components of λ in our model, using the likelihood in (2.5). Note that this method requires only maximization of slightly modified likelihood functions, and so should be feasible whenever maximum likelihood estimation is feasible. In general, the method can be adapted to estimate marginal posterior densities.

Laplace's method will be used to approximate posterior means in the simulated data sets of the next chapter.

Chapter 3

Estimation and Simulation

3.1 Estimating the Common Parameter

As mentioned previously, when family size is small, as in a bivariate model, we seek information about the common parameter λ . Using the unconditional likelihood (2.5), there are several possible methods of point estimation of λ . The most common would be maximum likelihood estimation; however, with a prior distribution on λ , Bayesian estimation is more general and appropriate. Neither method will produce closed-form estimators, given the form of the likelihood (the univariate Weibull model with censoring has no closed-form MLE). However, using an iterative procedure, the MLE should not be hard to obtain. Possibly, computation of posterior means would be more difficult, but approximating posterior means using Laplace's method should not be much harder than finding the MLE, given the similarity in method.

We propose to simulate data from the Weibull model of section 2.4, with one member of each group per family. Hougaard (1986b) found maximum likelihood estimates in the positive stable frailty case, using a heavily censored trivariate sample with 50 observations (the sample is lifetimes of female rat litters, with one treated sibling per litter, found in Mantel, Bohidar & Ciminera, 1977). Using a quasi-Newton minimization routine, we were able to reproduce Hougaard's results. The routine was also successfully applied to the same data through the gamma frailty model. These estimates were used to choose hopefully realistic parameter values for simulation. Analytic derivatives of the likelihood

are complicated in both models; for simplicity, we use the derivative-free routine in our simulations. We found in general that we could maximize the likelihood (and the slightly modified likelihoods needed for Laplace's method) fairly routinely. Of course, to do Bayesian estimation, we must also specify prior distributions.

3.2 Prior Information

Recognizing our relative paucity of prior information, we use vague priors on λ . One such prior is

$$\Pi_1(\lambda) = \begin{cases} 1 & \lambda_0 \in (0, 1], \lambda_j > 0 \ (j = 1, 2, 3) \\ 0 & \text{otherwise} \end{cases}$$

We note from (2.10) that λ_1^{-1/λ_3} is a scale parameter for Y_{i1} and similarly λ_2^{-1/λ_3} is a scale parameter for Y_{i2} (a conditional scale factor is also an unconditional scale factor).

Computing the necessary Jacobian, we see that the scale invariant prior will be

$$\Pi_2(\lambda) = \begin{cases} \lambda_1^{-1} \lambda_2^{-1} \lambda_3^{-2} & \lambda_0 \in (0, 1], \lambda_j > 0 \ (j = 1, 2, 3) \\ 0 & \text{otherwise} \end{cases}$$

We try both priors in the simulation study.

3.3 Reparameterizations

There are of course an infinite number of possible reparameterizations of our model. The point maximizing the likelihood is invariant under transformation of the parameter space. The posterior mean of a parameter in a given parameterization can be computed using an arbitrary parameterization for computation. For instance suppose we are interested in the posterior mean of λ_0 , but we parameterize using $\kappa = g(\lambda)$. The likelihood becomes $L'(\kappa) = L(g^{-1}(\kappa))$. Our prior becomes $\Pi'(\kappa) = \Pi(g^{-1}(\kappa)) \left| \frac{d}{d\kappa} g^{-1}(\kappa) \right|$ with respect to which we want to find the posterior mean of $(g^{-1}(\kappa))_0$. So reparameterization just induces

a change of variables in integration; the true posterior mean is unaffected. However, when using Laplace's method, the Taylor approximations to the integrands could be better with one parameterization rather than another. So, a prudent choice of transformation g could potentially improve the estimation, when using approximate posterior means.

In the Weibull model, the parameter space of the parameterization of interest, is truncated. ($\lambda_0 \in (0, 1]$, $\lambda_j > 0, j = 1, 2, 3$). However, the approximating integrals in Laplace's method are over all of \mathcal{R}^4 . Perhaps choosing a parameterization that similarly has support on \mathcal{R}^4 might improve the approximation. To this end we introduce reparameterization 1:

$$\begin{aligned}\kappa_0 &= -\log(-\log \lambda_0) \\ \kappa_1 &= \log \lambda_1 \\ \kappa_2 &= \log \lambda_2 \\ \kappa_3 &= \log \lambda_3\end{aligned}\tag{3.14}$$

which has Jacobian

$$\left| \frac{d}{d\kappa} g^{-1}(\kappa) \right| = \exp(\kappa_1 + \kappa_2 + \kappa_3 - \kappa_0 - e^{-\kappa_0})$$

A second reparameterization was used by Hougaard (1986b) in performing maximum likelihood estimation in the positive stable model:

$$\begin{aligned}\kappa_0 &= \lambda_0 \\ \kappa_1 &= \lambda_1^{\lambda_0} \\ \kappa_2 &= \lambda_2^{\lambda_0} \\ \kappa_3 &= \lambda_0 \lambda_3\end{aligned}\tag{3.15}$$

for which the Jacobian is

$$\left| \frac{d}{d\kappa} g^{-1}(\kappa) \right| = \left(\frac{1}{\kappa_0} \right)^3 (\kappa_1 \kappa_2)^{1/\kappa_0 - 1}$$

We try using the original parameterization and the two reparameterizations for computation in the simulation study.

3.4 Weibull/Gamma Simulation Study

The obvious way to simulate data from the Weibull Model is to generate each family's frailty independently from the frailty distribution, and then generate the lifetimes within a family as independent Weibull variates with parameters depending on the frailty. So in the gamma frailty case, it is a straightforward exercise in simulating gamma and Weibull variates. We fix the parameter value at $\lambda = (.6, 1.0, 1.0, 3.0)$, and generate 100 samples of size 50 each ¹ and 100 samples of size 200 each (the samples of size 200 are extensions of the samples of size 50). The parameter value was chosen on the basis of being close to preliminary parameter estimates for the data analyzed in Chapter 5.

We treat the original parameterization as the parameterization of interest. Using each parameterization, we compute the posterior mode of λ , and the approximate posterior means of the components of λ . Note that the MLE will be posterior mode of λ using the original parameterization and prior Π_1 . For each of the six estimators (and four components of λ), we report the mean of the 100 estimates. Also, as a realized loss in estimation, we report the sums of squares of the 100 deviations between parameter and estimate. The results are displayed in Tables 3.1 and 3.2.

The simulations with sample size 50 indicate that the approximate posterior means with respect to Π_2 outperform those computed with respect to Π_1 , in terms of realized loss. The only exceptional cases occur with component λ_0 , using the original parameterization and reparameterization II. The difference in loss is most noticeable in components λ_1 and λ_2 using the original parameterization and reparameterization I. Using reparameterization II seems to slightly improve the performance of the approximate posterior means; however, the estimates do not seem to be particularly sensitive to the selection between the original parameterization and reparameterization I. The performance of the

¹One sample of size 50 produced a likelihood which our routine could not maximize; therefore, results are provided based on the other 99 samples.

Table 3.1: Comparison of estimators for 99 simulated samples of 50 observations in the Weibull/gamma bivariate model with parameter vector $\lambda = (0.6, 1.0, 1.0, 3.0)$

Estimator	λ_0 (0.6)		λ_1 (1.0)		λ_2 (1.0)		λ_3 (3.0)	
	Mean	Loss	Mean	Loss	Mean	Loss	Mean	Loss
ORIG. PARAM.								
Prior I								
Approx. Post. Mean	.564	1.56	1.14	10.84	1.14	7.67	3.16	12.78
Posterior Mode (MLE)	.623	1.74	1.01	5.67	1.01	3.60	3.05	9.65
Prior II								
Approx. Post. Mean	.608	1.61	1.04	6.10	1.04	3.97	3.05	9.30
Posterior Mode	.663	2.08	.932	4.65	.934	3.10	2.95	8.72
REPARAM. I								
Prior I								
Approx. Post. Mean	.561	1.49	1.16	11.55	1.16	8.28	3.17	12.92
Posterior Mode	.560	1.42	1.12	9.28	1.12	6.37	3.16	12.31
Prior II								
Approx. Post. Mean	.599	1.26	1.05	6.32	1.05	4.12	3.06	9.25
Posterior Mode	.599	1.20	1.02	5.58	1.03	3.54	3.06	8.92
REPARAM. II								
Prior I								
Approx. Post. Mean	.592	.809	1.09	6.92	1.09	4.65	3.11	9.41
Posterior Mode	.495	3.74	1.18	17.84	1.18	12.71	3.26	22.07
Prior II								
Approx. Post. Mean	.617	1.02	1.02	5.04	1.02	3.26	3.02	7.76
Posterior Mode	.556	2.65	1.04	7.85	1.04	5.03	3.12	13.57

Table 3.2: Comparison of estimators for 100 simulated samples of 200 observations in the Weibull/gamma bivariate model with parameter vector $\lambda = (0.6, 1.0, 1.0, 3.0)$

Estimator	$\lambda_0 (0.6)$		$\lambda_1 (1.0)$		$\lambda_2 (1.0)$		$\lambda_3 (3.0)$	
	Mean	Loss	Mean	Loss	Mean	Loss	Mean	Loss
ORIG. PARAM.								
Prior I								
Approx. Post. Mean	.590	.318	1.04	1.56	1.04	1.39	3.05	2.37
Posterior Mode (MLE)	.606	.328	1.01	1.29	1.01	1.14	3.02	2.16
Prior II								
Approx. Post. Mean	.600	.308	1.02	1.34	1.02	1.18	3.02	2.15
Posterior Mode	.616	.350	.990	1.20	.989	1.06	2.99	2.09
REPARAM. I								
Prior I								
Approx. Post. Mean	.590	.317	1.04	1.57	1.04	1.40	3.05	2.37
Posterior Mode	.591	.309	1.03	1.48	1.03	1.32	3.04	2.34
Prior II								
Approx. Post. Mean	.600	.307	1.02	1.34	1.02	1.18	3.02	2.15
Posterior Mode	.601	.300	1.01	1.29	1.01	1.14	3.02	2.13
REPARAM. II								
Prior I								
Approx. Post. Mean	.591	.314	1.04	1.56	1.04	1.39	3.04	2.35
Posterior Mode	.581	.387	1.03	1.55	1.03	1.38	3.06	2.57
Prior II								
Approx. Post. Mean	.601	.305	1.02	1.34	1.02	1.18	3.02	2.14
Posterior Mode	.592	.357	1.01	1.34	1.01	1.18	3.03	2.29

MLE is relatively good, except in component λ_0 . Empirically then, one might prefer to use approximate posterior means when the frailty distribution parameter is of primary interest. Not surprisingly, with sample size 200, there is little difference in the estimators.

3.5 Weibull/Positive Stable Simulation Study

The positive stable frailty simulations were carried out in exactly the same fashion as in the gamma case. The true parameter value was fixed at $\lambda = (.9, 1.0, 1.0, 3.0)$ (again, this is based on estimated values from the data in Chapter 5). To simulate positive stable random variates, a formula due to Kanter (1975) was used. The stable variate θ is realized according to

$$\theta = \left[\frac{a(X_1)}{X_2} \right]^{\lambda_0^{-1}-1}$$

where X_1 and X_2 are independent; X_1 is uniform on $[0, \Pi]$, while X_2 has a standard exponential distribution. The function a is given by

$$a(x) = \frac{\sin[(1 - \lambda_0)x] \{\sin(\lambda_0 x)\}^{\frac{\lambda_0}{1-\lambda_0}}}{\{\sin x\}^{\frac{1}{1-\lambda_0}}}$$

This is a special case of the method for general stable variates presented in Chambers, Mallows & Stuck (1976). The simulation results are presented in Tables 3.3 and 3.4.

As previously, the simulations with sample size 50 indicate that using Π_2 rather than Π_1 improves the performance of the approximate posterior means, in terms of realized loss. However, it seems there is less sensitivity to reparameterization than in the gamma frailty case. The MLE again performs well in components λ_1 , λ_2 , λ_3 , but has higher realized loss than approximate posterior means in component λ_0 . The results with sample size 200 indicate very little difference between the various estimators.

The simulated samples and the resultant estimates are put to use in the study of a prediction problem in the next chapter. Also, both maximum likelihood and approximate

Table 3.3: Comparison of estimators for 100 simulated samples of 50 observations in the Weibull/positive stable bivariate model with parameter vector $\lambda = (0.9, 1.0, 1.0, 3.0)$

Estimator	$\lambda_0 (0.9)$		$\lambda_1 (1.0)$		$\lambda_2 (1.0)$		$\lambda_3 (3.0)$	
	Mean	Loss	Mean	Loss	Mean	Loss	Mean	Loss
ORIG. PARAM.								
Prior I								
Approx. Post. Mean	.865	.617	1.04	3.32	1.06	4.24	3.17	16.32
Posterior Mode (MLE)	.896	.561	1.00	2.70	1.02	3.52	3.10	13.92
Prior II								
Approx. Post. Mean	.879	.500	1.01	2.76	1.01	3.60	3.11	14.06
Posterior Mode	.909	.519	.973	2.48	.991	3.27	3.05	12.47
REPARAM. I								
Prior I								
Approx. Post. Mean	.850	.621	1.05	3.50	1.07	4.55	3.22	17.16
Posterior Mode	.839	.737	1.05	3.51	1.07	4.60	3.25	18.47
Prior II								
Approx. Post. Mean	.862	.469	1.01	2.88	1.03	3.81	3.16	14.54
Posterior Mode	.851	.562	1.01	2.93	1.03	3.88	3.19	15.61
REPARAM. II								
Prior I								
Approx. Post. Mean	.864	.755	1.04	3.39	1.06	4.36	3.17	16.44
Posterior Mode	.871	.735	1.01	3.00	1.03	3.84	3.16	17.34
Prior II								
Approx. Post. Mean	.881	.516	1.01	2.77	1.03	3.70	3.13	17.44
Posterior Mode	.887	.618	.979	2.67	.997	4.34	3.10	15.14

Table 3.4: Comparison of estimators for 100 simulated samples of 200 observations in the Weibull/positive stable bivariate model with parameter vector $\lambda = (0.9, 1.0, 1.0, 3.0)$

Estimator	$\lambda_0 (0.9)$		$\lambda_1 (1.0)$		$\lambda_2 (1.0)$		$\lambda_3 (3.0)$	
	Mean	Loss	Mean	Loss	Mean	Loss	Mean	Loss
ORIG. PARAM.								
Prior I								
Approx. Post. Mean	.892	.179	1.02	.606	1.01	.523	3.05	3.70
Posterior Mode (MLE)	.900	.186	1.01	.568	1.00	.489	3.03	3.57
Prior II								
Approx. Post. Mean	.894	.167	1.01	.599	1.00	.516	3.04	3.55
Posterior Mode	.903	.178	.997	.560	.995	.498	3.02	3.47
REPARAM. I								
Prior I								
Approx. Post. Mean	.887	.165	1.01	.610	1.01	.544	3.07	3.69
Posterior Mode	.880	.177	1.01	.618	1.01	.551	3.09	3.911
Prior II								
Approx. Post. Mean	.891	.153	1.01	.582	1.00	.519	3.05	3.46
Posterior Mode	.884	.161	1.01	.590	1.00	.526	3.07	3.64
REPARAM. II								
Prior I								
Approx. Post. Mean	.890	.176	1.01	.607	1.01	.540	3.06	3.75
Posterior Mode	.892	.191	1.01	.583	1.00	.517	3.05	3.85
Prior II								
Approx. Post. Mean	.894	.168	1.01	.579	1.00	.515	3.04	3.57
Posterior Mode	.896	.184	.999	.567	.977	.504	3.04	3.68

Bayesian estimation are utilized in the data analysis of Chapter 5.

Chapter 4

Prediction

In many instances it may be desirable to use information about a family's frailty in order to predict individual lifetimes. Of course, to do so would also require information about the individual's baseline hazard. For simplicity, we will assume that if we wish to predict potentially observable survival time $Y_{i,j}$, then we have observed all the other components of the i^{th} family (unobserved family members need not be included in the model). We denote these components by $Y_{i,(j)} = (Y_{i,1}, \dots, Y_{i,j-1}, Y_{i,j+1}, \dots, Y_{i,n_i})$. So, if we have observed m families, and the individual of interest belongs to the $(m+1)^{st}$ family, then we wish to study $Y_{m+1,j} \mid Y_1, \dots, Y_m, Y_{m+1,(j)}$. In such a situation, observing $Y_1, \dots, Y_m, Y_{m+1,(j)}$ gives information about the baseline effects and the frailty distribution; however, only $Y_{m+1,(j)}$ can give direct information about θ_{m+1} , the frailty of $Y_{m+1,j}$. The question of how much improvement in our knowledge about θ_{m+1} is realized by observing $Y_{m+1,(j)}$ is crucial to the prediction. Before delving into detail, we look at why this scenario might arise naturally.

Often we might have a situation where within a family, survival times of one group are generally measured before those of another group. In such a setting it may be of interest to predict the second group times given the first group times. For example, having observed the lifetimes of a number of father-son pairs, one might want to predict a son's lifetime given his father's lifetime. Similarly, if an individual has two responses, with one often larger than the other, such prediction might be relevant. For instance, Holt & Prentice (1974) report survival times of several skin-grafts per patient, with a mixture

of poor-match and good-match grafts. In a clinical situation, if a patient has grafts of both types, it might be desirable to predict the failure times of the good grafts, having observed the poor-match failure times. Additionally, in a clinical counselling situation, the onset times of a disease in family members could be used to predict the onset time in an unaffected individual.

Given a data set, one approach to the prediction problem is to assume the value of the underlying parameter λ is known and given by an estimator. This is known as an empirical Bayes procedure. Alternatively, one can incorporate the uncertainty about the value of λ by considering the posterior distribution of λ . Here, we consider both approaches.

4.1 Parametric Empirical Bayes Approach to Prediction

To take a Parametric Empirical Bayes (PEB) approach to prediction, we assume that based on the first m families and the partial observation of the $(m+1)^{st}$ family, we have an estimator $\hat{\lambda}$ of λ . This could be the maximum likelihood estimator, a vector of approximate posterior means, or another estimator. The PEB approximation amounts to assuming $\lambda = \hat{\lambda}$. So given $Y_1, \dots, Y_m, Y_{m+1,(j)}$, we assume θ_{m+1} is distributed with density $g(\theta_{m+1} | \hat{\lambda})$. If we treat this as a prior distribution for θ_{m+1} , and approximate the density of $Y_{m+1,(j)}$ by $f(y_{m+1,(j)} | \theta_{m+1}, \hat{\lambda})$, an application of Bayes theorem gives the density of $\theta_{m+1} | Y_1, \dots, Y_m, Y_{m+1,(j)}$ as proportional to $f(y_{m+1,(j)} | \theta_{m+1}, \hat{\lambda})g(\theta_{m+1} | \hat{\lambda})$. It follows that the predictive density of $Y_{m+1,j}$ is

$$\frac{\int f(y_{m+1,j} | \theta_{m+1}, \hat{\lambda}) f(y_{m+1,(j)} | \theta_{m+1}, \hat{\lambda}) g(\theta_{m+1} | \hat{\lambda}) d\theta_{m+1}}{\int f(y_{m+1,(j)} | \theta_{m+1}, \hat{\lambda}) g(\theta_{m+1} | \hat{\lambda}) d\theta_{m+1}} \quad (4.16)$$

We note that the denominator of (4.16) is simply the marginal unconditional density of $Y_{m+1,(j)}$ given $\lambda = \hat{\lambda}$. Similarly, given the conditional independence structure, the numerator of (4.16) is the unconditional joint density of $Y_{m+1} = (Y_{m+1,(j)}, Y_{m+1,j})$. So the PEB

predictive density is simply the conditional density of $Y_{m+1,j} \mid Y_{m+1,(j)}$ with parameter λ replaced by its estimator (the estimator contains the information from Y_1, \dots, Y_m). We denote this density by $p_1(y_{m+1,j} \mid y_{m+1,(j)}, \hat{\lambda})$.

4.2 Full Bayesian Approach to Prediction

To take a full Bayesian approach to prediction, we need the posterior distribution of $\theta_{m+1}, \lambda \mid Y_1, \dots, Y_m, Y_{m+1,(j)}$. By the definition of conditional probability, this will be the product of the densities of $\theta_{m+1} \mid \lambda, Y_1, \dots, Y_m, Y_{m+1,(j)}$ and $\lambda \mid Y_1, \dots, Y_m, Y_{m+1,(j)}$. But $\theta_{m+1} \mid \lambda, Y_1, \dots, Y_m, Y_{m+1,(j)} \equiv \theta_{m+1} \mid \lambda, Y_{m+1,(j)}$, which has posterior density

$$\frac{f(y_{m+1,(j)} \mid \theta_{m+1}, \lambda)g(\theta_{m+1} \mid \lambda)}{\int f(y_{m+1,(j)} \mid \theta_{m+1}, \lambda)g(\theta_{m+1} \mid \lambda)d\theta_{m+1}}$$

while the distribution of $\lambda \mid Y_1, \dots, Y_m, Y_{m+1,(j)}$ is just the posterior distribution of λ given all the observed data. We denote the corresponding density by $\Pi^*(\lambda)$. Hence the predictive density will be

$$\int \int f(y_{m+1,j} \mid \theta_{m+1}, \lambda) \left[\frac{f(y_{m+1,(j)} \mid \theta_{m+1}, \lambda)g(\theta_{m+1} \mid \lambda)}{\int f(y_{m+1,(j)} \mid \theta'_{m+1}, \lambda)g(\theta'_{m+1} \mid \lambda)d\theta'_{m+1}} \right] \Pi^*(\lambda)d\theta_{m+1}d\lambda$$

which simplifies to

$$\int \left[\frac{\int f(y_{m+1,j} \mid \theta_{m+1}, \lambda)f(y_{m+1,(j)} \mid \theta_{m+1}, \lambda)g(\theta_{m+1} \mid \lambda)d\theta_{m+1}}{\int f(y_{m+1,(j)} \mid \theta_{m+1}, \lambda)g(\theta_{m+1} \mid \lambda)d\theta_{m+1}} \right] \Pi^*(\lambda)d\lambda$$

or equivalently

$$E^{\Pi^*} \left\{ p_1(y_{m+1,j} \mid y_{m+1,(j)}, \lambda) \right\} \quad (4.17)$$

We denote this density by $p_2(y_{m+1,j} \mid \text{data})$. Note that the PEB predictive density can be obtained from (4.17) by assuming a posterior distribution on λ which is degenerate at $\hat{\lambda}$. As with other posterior quantities, we anticipate that it will not be possible to evaluate (4.17) exactly.

4.3 The PEB Approach in the Weibull Model

To evaluate PEB predictive densities, we can easily obtain conditional densities as ratios of appropriate joint densities. For example, in a bivariate model, suppose we want the conditional density of $Y_{i,2} \mid Y_{i,1}$. Suppressing the family index, we know an uncensored family has density

$$f(y_1, y_2 \mid \lambda) = h_1(y_1)h_2(y_2)E_\lambda^\theta[\theta^2 \exp(-\theta H(y))]$$

Similarly, by considering a family of size 1, we know

$$f(y_1 \mid \lambda) = h_1(y_1)E_\lambda^\theta[\theta \exp(-\theta H_1(y_1))]$$

Taking the ratio yields

$$p_1(y_2 \mid y_1, \lambda) = \frac{h_2(y_2)E_\lambda^\theta[\theta^2 \exp(-\theta H(y))]}{E_\lambda^\theta[\theta \exp(-\theta H_1(y_1))]} \quad (4.18)$$

Similar arguments give conditionals for censored cases:

$$\begin{aligned} f(y_2 \mid Y_1 > y_1, \lambda) &= \frac{h_2(y_2)E_\lambda^\theta[\theta \exp(-\theta H(y))]}{E_\lambda^\theta[\exp(-\theta H_1(y_1))]} \\ P(Y_2 > y_2 \mid Y_1 = y_1, \lambda) &= \frac{E_\lambda^\theta[\theta \exp(-\theta H(y))]}{E_\lambda^\theta[\theta \exp(-\theta H_1(y_1))]} \\ P(Y_2 > y_2 \mid Y_1 > y_1, \lambda) &= \frac{E_\lambda^\theta[\exp(-\theta H(y))]}{E_\lambda^\theta[\exp(-\theta H_1(y_1))]} \end{aligned}$$

Note that all these expressions can be easily evaluated in either the Weibull/gamma model or the Weibull/positive stable model. Assuming Y_1 is in group 1 and Y_2 is in group 2, (4.18) becomes

$$p_1(y_2 \mid y_1, \lambda) = \frac{(\alpha + 1)\lambda_2\lambda_3(\alpha + \lambda_1 y_1^{\lambda_3})^{(\alpha+1)} y_2^{\lambda_3-1}}{(\alpha + \lambda_1 y_1^{\lambda_3} + \lambda_2 y_2^{\lambda_3})^{(\alpha+2)}} \quad (4.19)$$

where $\alpha = -(\log \lambda_0)^{-1}$. When $\lambda_3 > 1$, this is a unimodal density; otherwise, it is strictly decreasing.

In the positive stable case, as mentioned in Hougaard(1986b), (4.18) is

$$p_1(y_2 | y_1, \lambda) = (\lambda_2 \lambda_3 y_2^{\lambda_3 - 1}) \lambda_0 \left[\frac{(\lambda_1 y_1^{\lambda_3} + \lambda_2 y_2^{\lambda_3})^2}{\lambda_1 y_1^{\lambda_3}} \right] \times \exp[(\lambda_2 y_2^{\lambda_3})^{\lambda_0} - (\lambda_1 y_1^{\lambda_3} + \lambda_2 y_2^{\lambda_3})^{\lambda_0}] \left[1 + (\lambda_0^{-1} - 1)(\lambda_1 y_1^{\lambda_3} + \lambda_2 y_2^{\lambda_3})^{-\lambda_0} \right] \quad (4.20)$$

Note that in both cases, $\lambda_2^{(\lambda_3^{-1})}$ is a scale parameter.

4.4 The Full Bayesian Approach in the Weibull Model

We note that with the complicated functional forms for the predictive densities (4.19) and (4.20), analytic evaluation of (4.17) would not be possible even if we had a nice form for the posterior density. We need to resort to some kind of approximation. Since a predictive density is necessarily a positive function, we could use the Laplace approximation method. However, this would have to be applied pointwise: a numerical maximization would be required for each point at which the predictive density is approximated. Furthermore, the approximate density may need to be normalized, requiring numerical integration.

A second possibility would be to replace the posterior distribution with a multivariate normal distribution. That is, if $\log(\text{likelihood} \times \text{prior})$ has mode μ and Hessian $-\Sigma$ at the mode, then replace Π^* with a $N(\mu, \Sigma^{-1})$ distribution. This is correct asymptotically (Berger, 1985, p. 224). However, due to the complicated integrand, Monte Carlo methods would still be required. For little additional cost (and potentially considerable gain) one could use this Normal distribution as an importance sampling distribution.

So, using the usual importance sampling scheme, if Π^0 is the $N(\mu, \Sigma^{-1})$ density function, and if $\lambda^{(1)}, \dots, \lambda^{(n)}$ are simulated vectors from this distribution, then our approximation is

$$p_2(z | \text{data}) \approx \sum_{i=1}^n c_i p_1(z | y_{m+1,2}, \lambda^{(i)}) \quad (4.21)$$

where $c_i \propto \frac{\Pi^*(\lambda^{(i)})}{\Pi^0(\lambda^{(i)})}$ and $\sum_{i=1}^n c_i = 1$

Note that we need only evaluate the posterior Π^* up to a constant, so it is sufficient just to evaluate the likelihood and prior. Hence, our estimate of p_2 is easily normalized.

It is of course of interest, within the context of the Weibull model as well as more generally, to discuss the relative merits of the PEB and full Bayesian approaches to prediction.

4.5 Full versus PEB Prediction

Generally, the difference between any parametric empirical Bayes approach and a full hierarchical approach is that the PEB approach neglects the experimenter's uncertainty about the value of λ . For this reason, we expect the PEB predictive density to be overconfident: to be more peaked and have thinner tails than the data can justify. However, the PEB approach does yield the correct functional form of the true conditional density, and if our estimate $\hat{\lambda}$ of λ is particularly good, we might expect the PEB predictive density to outperform the full predictive density.

Since we are comparing two specified probability density functions intended to describe observable $Y_{m+1,j} \mid Y_1, \dots, Y_m, y_{m+1,(j)}$, we choose to compare their performance by the density ratio

$$L(Y_1, \dots, Y_{m+1}) = \frac{p_2(Y_{m+1,j} \mid Y_1, \dots, Y_m, Y_{m+1,(j)})}{p_1(Y_{m+1,j} \mid Y_{m+1,(j)}, \hat{\lambda}(Y_1, \dots, Y_m, Y_{m+1,(j)}))} \quad (4.22)$$

Often we prefer to work with the log-ratio due to its symmetry. (A positive value corresponds to p_2 outperforming p_1 , while the corresponding negative value indicates p_1 outperforming p_2 by the same amount.) Of course in our model, p_1 and p_2 are complicated, so it is unlikely we can make any analytic inferences about the distribution of L . For this reason we mention a simple, though somewhat unrelated example of prediction. We speculate there is similar behaviour of likelihood ratios in the example and in our model.

4.5.1 Exponential Prediction Example

Let X_1, \dots, X_{n+1} be iid exponential with scale $1/\lambda$. Take a non-informative prior $\Pi(\lambda) = 1/\lambda$. It is straightforward to see that $\lambda \mid X_1, \dots, X_n$ has a gamma distribution with shape n , and inverse scale $\sum_{i=1}^n X_i$. So based on the first n observations, λ has posterior mean $n/\sum_{i=1}^n X_i$, which is also the MLE. A “plug-in” method neglecting uncertainty about λ would predict X_{n+1} as exponential with scale $\sum_{i=1}^n X_i/n$. The full predictive density would be the expectation of the density of X_{n+1} with respect to the gamma posterior distribution on λ . The ratio of the second density to the first is

$$L(X_1, \dots, X_{n+1}) = \frac{\frac{(\sum_{i=1}^n X_i)^n}{\Gamma(n)} \int [\theta \exp(-\theta X_{n+1})] \theta^{n-1} \exp(-(\sum_{i=1}^n X_i)\theta) d\theta}{\sum_{i=1}^n X_i \exp(-X_{n+1}(n/\sum_{i=1}^n X_i))}$$

This simplifies to

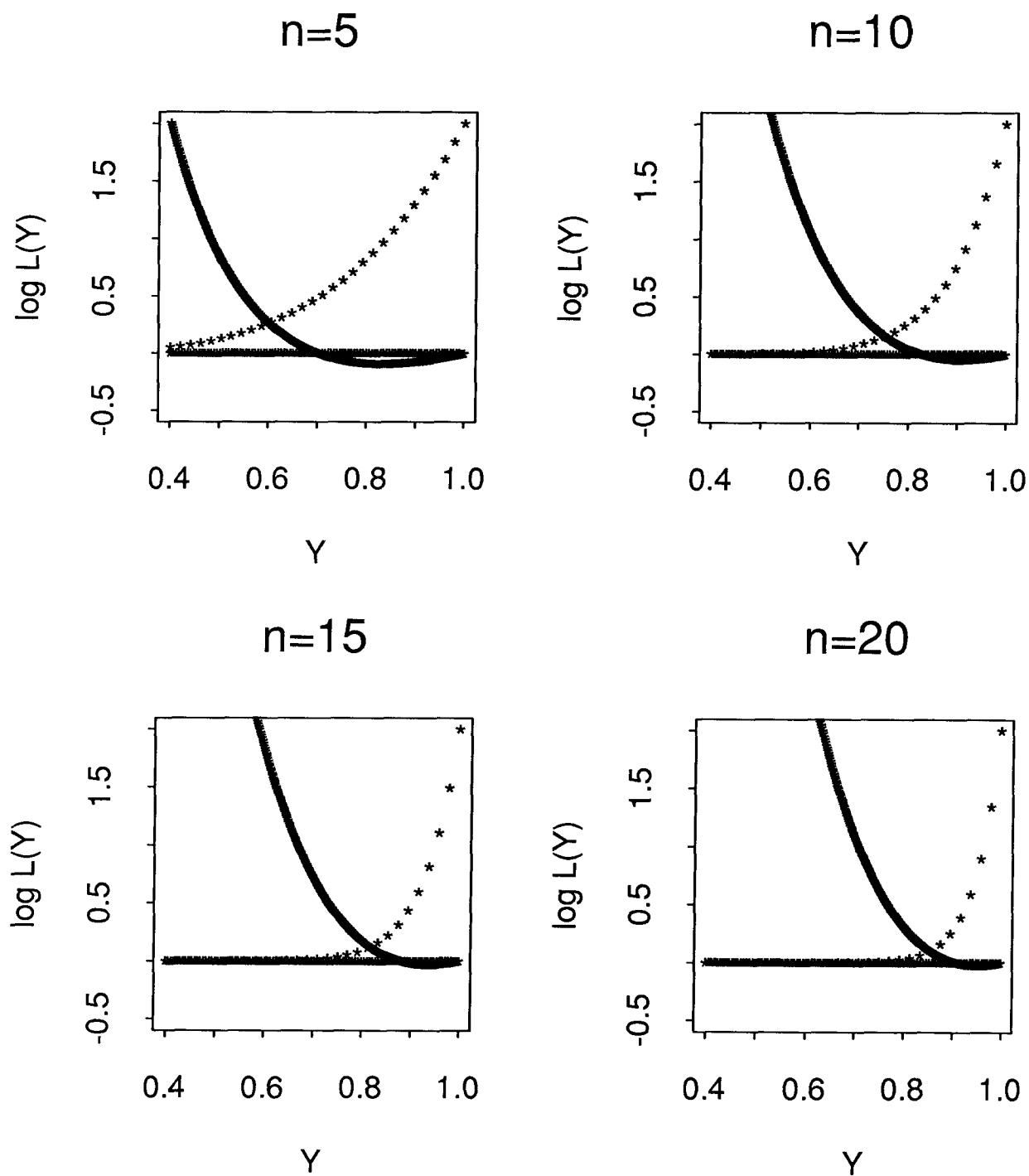
$$L(X_1, \dots, X_{n+1}) = \left[\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i + X_{n+1}} \right]^{n+1} \exp\left(\frac{nX_{n+1}}{\sum_{i=1}^n X_i}\right)$$

which can be written as

$$L(X_1, \dots, X_{n+1}) = Y^{n+1} \exp[n(Y^{-1} - 1)]$$

$$\text{where } Y = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^{n+1} X_i}$$

But by a well known property of independent gamma distributions, Y has a beta($n, 1$) distribution. Note that as a function of Y , L attains a minimum of $e(\frac{n}{n+1})^{n+1}$ at $Y = \frac{n}{n+1}$, but that $L(Y)$ increases without bound as Y decreases to the left of the minima. Furthermore $L(0) = 1$. However, the minimum occurs at the mean of Y 's distribution, whereas Y has little mass where L is large. To clarify, we superimpose $\log L(Y)$ and the density of Y for various n in Figure 4.1. So the region where $L < 1$ has high probability; however, L is almost surely larger than $k_n = e(\frac{n}{n+1})^{n+1}$. Conversely, large values of L are possible, but are unlikely. Based on the beta distribution function, we tabulate

Figure 4.1: Predictive density ratio as a function of Y (exponential example)Dotted Line proportional to density of Y

Horizontal Line equals zero

Table 4.5: Probabilities of intervals for $\log L$ in exponential example

n	k_n	$(\log k_n, 0)$	$(0, -\log k_n)$	$(-\log k_n, \infty)$
5	.91	.83	.05	.12
10	.95	.85	.04	.11
15	.97	.85	.05	.10
20	.98	.86	.04	.10

probabilities of $\log L$ lying in the intervals $(\log k_n, 0)$, $(0, -\log k_n)$, and $(-\log k_n, \infty)$ for various n in Table (4.5).

So, the “plug-in” prediction has a high probability of doing slightly better than the proper predictive density, but it cannot do a lot better. With small probability, it will do much worse. Thus the full predictive density is more robust, at the cost of usually being slightly outperformed. This is consistent with the conjectured sharper peaks and thinner tails of general PEB predictive densities mentioned earlier.

4.6 Prediction in the Weibull/Gamma Model

To try out the prediction in our model, we take the 99 samples of size 50 simulated from the Weibull/gamma model, and extend them with a 51st pair of observations. For each sample, we compute a PEB predictive density and an approximate full predictive density for $Y_{51,2} | Y_{51,1}, Y_1, \dots, Y_{50}$. For the PEB density we estimate λ by the approximate posterior means with respect to the flat prior Π_1 , computed using reparameterization II. (The simulations indicated approximate posterior means give better estimates of λ_0 , the parameter governing the frailty distribution, and hence the association within families. Since this association will be critical to our prediction, we use these estimators, even though they are slightly outperformed by the MLE in terms of the baseline parameters.)

For the approximate full predictive density, we use importance sampling, generating

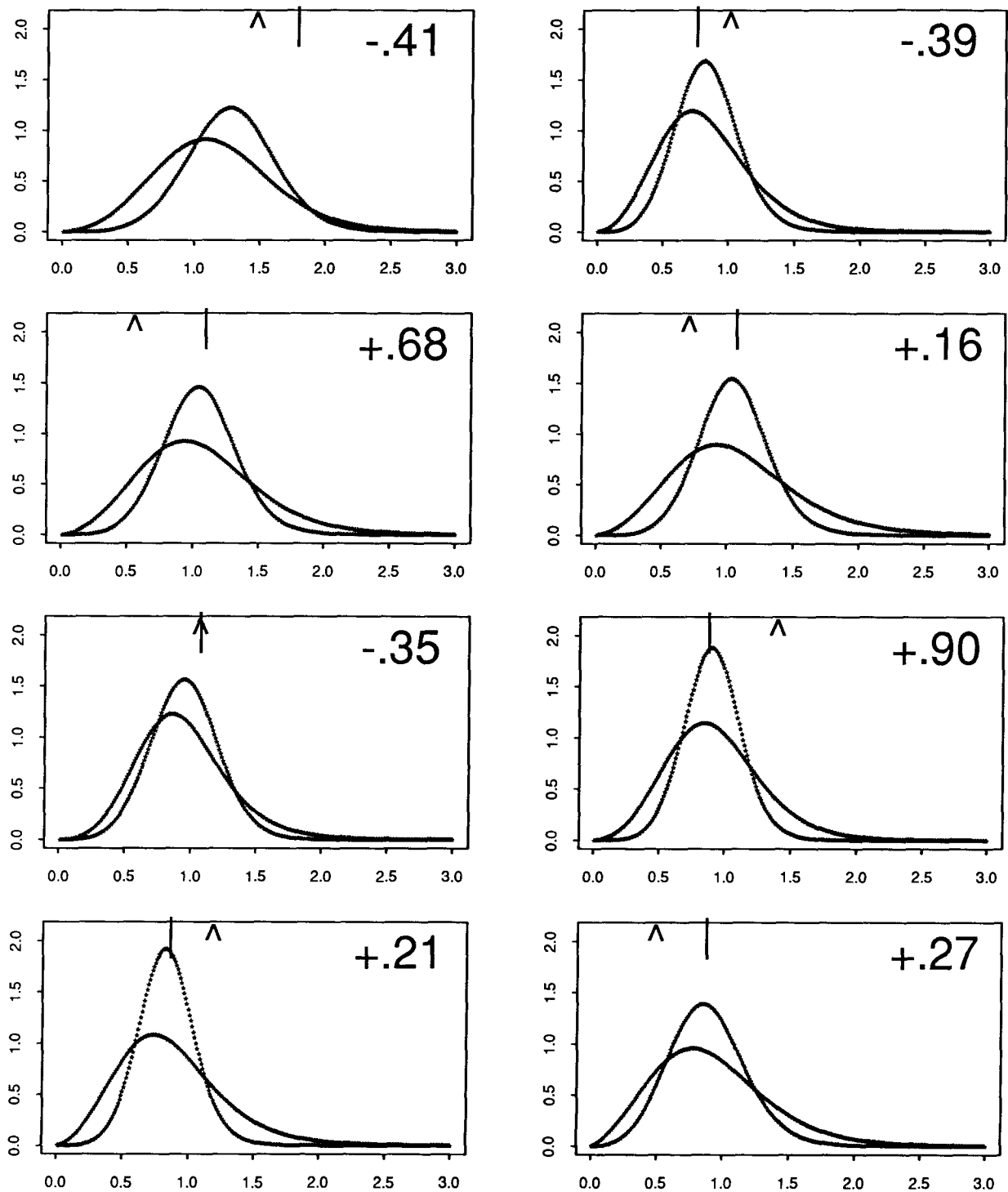
1000 multinormal vectors to approximate the integral. By estimating the error in the importance sampling, and by visually inspecting the effect of the importance sample size, it seems that this sample size is sufficient to control stochasticity introduced by the importance sample. We also use reparameterization II for this because the previous simulation results suggest that this parameterization leads to a ‘more normal’ posterior distribution than the others. If this is true, the importance sampling should be most efficient with this parameterization, given the choice of a multinormal importance density. Of course this makes our predictive density a sum of 1000 terms, which makes it somewhat hard to handle. However, we cannot really expect to do much better. Since the integral being approximated is 4-dimensional, using this sample size is analogous to approximating a 1-dimensional integral using $\sqrt[4]{1000} \approx 5.6$ simulated variates.

The two predictive densities are plotted for the first 8 samples in Figure 4.2. In each case the PEB predictive density is the more peaked and thinner-tailed of the pair. The potential overconfidence of the PEB densities is clear. It becomes a reality in light of Figure 4.3, a histogram of the log-predictive density ratios. While half the samples produce negative values (PEB beats full density) there are no large negative values, but there is a thick tail in the large positive values. The relative robustness of the full predictive density observed in the exponential example is empirically obvious in our model.

4.7 Prediction in the Weibull/Positive Stable Model

The same procedures mentioned above were carried out for the 100 samples of size 50 simulated from the Weibull/positive stable model. In this case, the maximum likelihood estimates of λ were used for the PEB prediction, given their performance in the simulation. Because of the proximity of the true value of λ_0 (0.9) to the parameter space

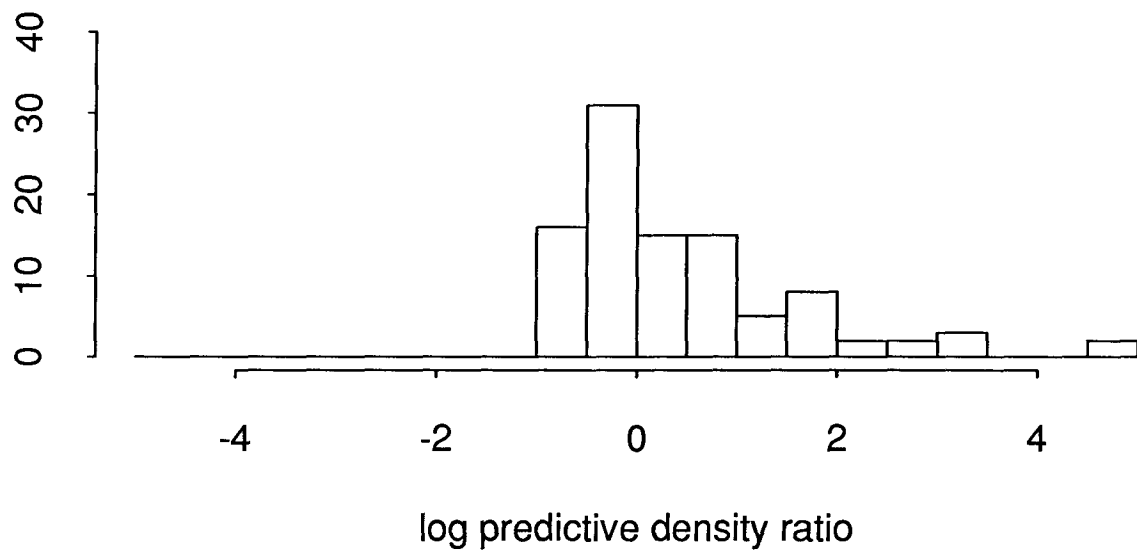
Figure 4.2: Approximate full and empirical predictive densities (gamma frailty)



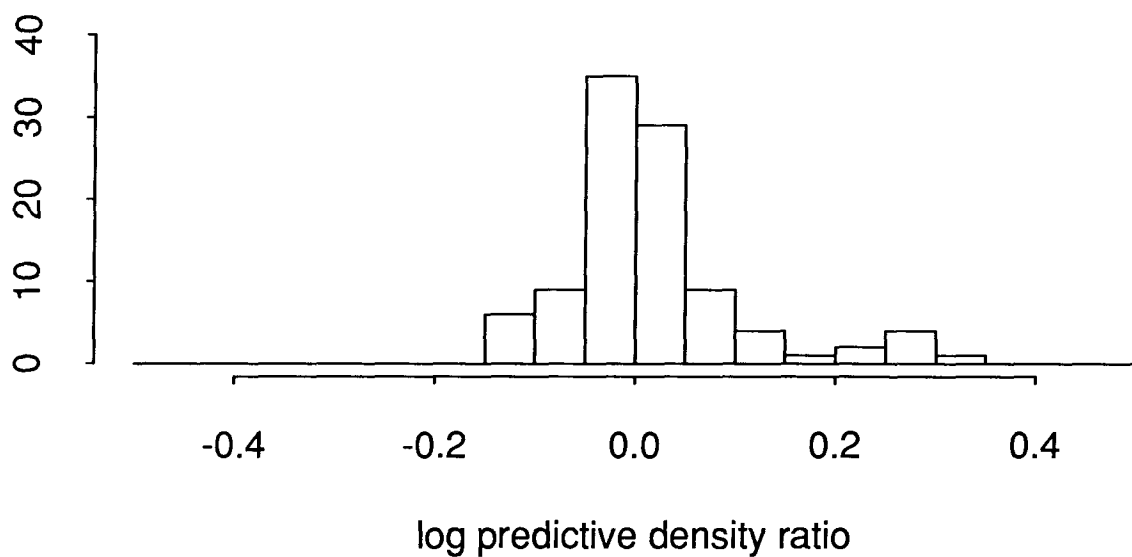
| marks predicting value $Y(51,1)$ ^ marks observed predicted value $Y(51,2)$
 number is log-likelihood ratio evaluated at observed $Y(51,2)$

Figure 4.3: Histograms of observed predictive density ratios: approximate full versus empirical Bayes

99 samples of size 50+1 with gamma frailty



100 samples of size 50+1 with stable frailty



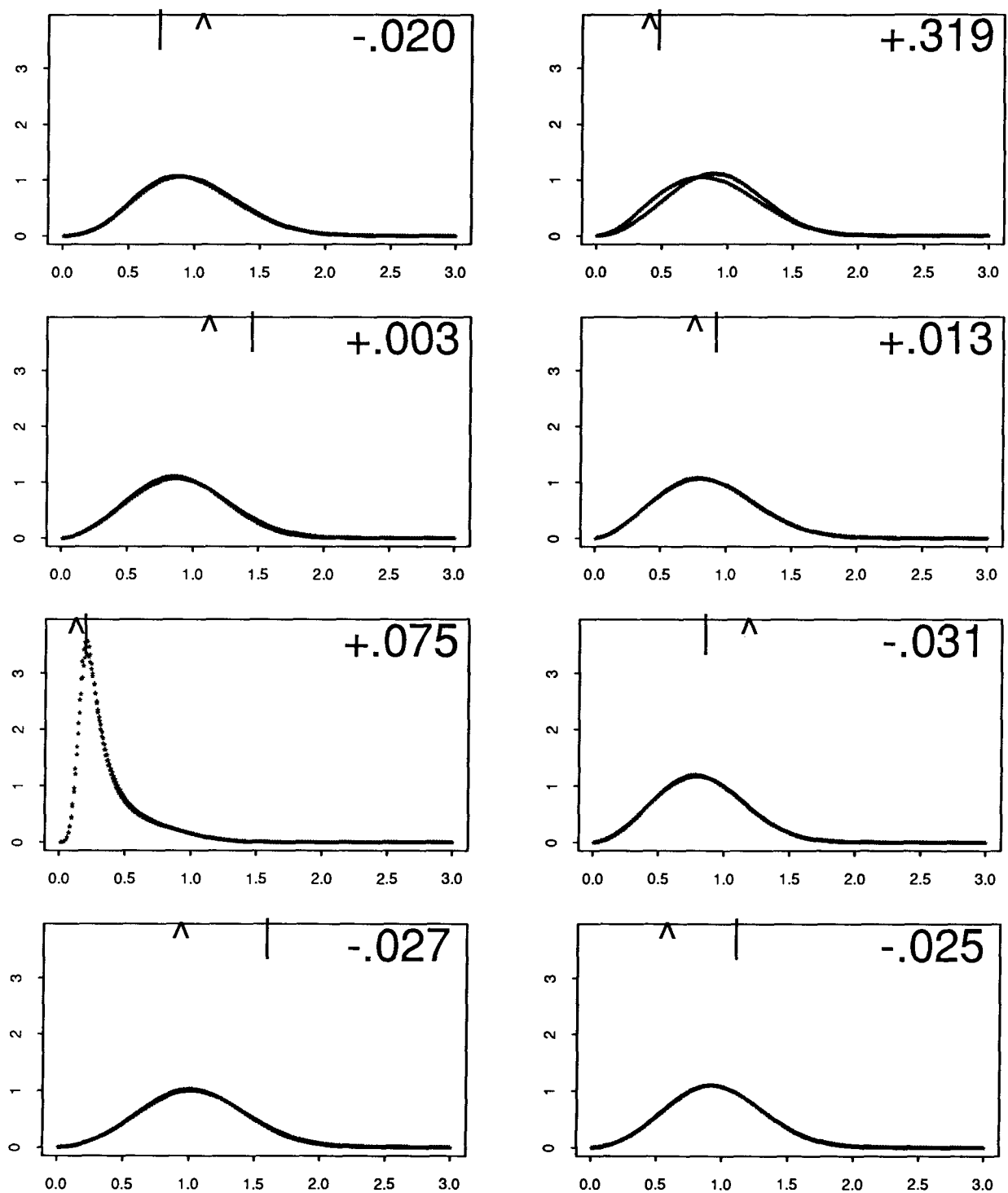
boundary ($\lambda_0 = 1$), many samples produced a substantial fraction of importance sample points outside the parameter space. However, this did not seem to hamper the accuracy of the approximation. The histogram of observed log-likelihood ratios is displayed in Figure 4.3. The right tail is again slightly heavier, but this trend is less pronounced than in the gamma frailty case. The important difference however is that the variability is roughly an order of magnitude smaller. Observing the plots of the two predictive densities in Figure 4.4 shows that in all but one of the first eight samples, the two densities are visually indistinguishable. It seems that with positive stable frailty and a sample of size 50, PEB prediction is virtually identical to the full Bayesian approach to prediction.

4.8 Effect of Frailty Distribution on Prediction

We speculate that the full Bayesian and Empirical Bayes predictive densities are so close in the stable frailty case for the following reason. Implicit in the prediction is an updating of knowledge regarding θ_{m+1} . Prior to observing $Y_{m+1,(j)}$, we know θ_{m+1} comes from a positive stable distribution, and we have information about the parameter, λ_0 . We use $Y_{m+1,(j)}$ to update (via Bayes theorem) the distribution on θ_{m+1} . Because the stable distribution has such a thick right-tail, we expect the information from $Y_{m+1,(j)}$ dominates the prior information. In other words, our knowledge about θ_{m+1} is not sensitive to our knowledge of λ_0 .

This would explain why replacing a distribution on λ_0 with a point mass might not greatly affect the prediction. Further, because the information on the remaining components of λ is so good compared to the knowledge about θ_{m+1} , interchanging degenerate and non-degenerate distributions on these components will not have a large effect. Conversely, with gamma frailty, the prior knowledge about λ_0 would be important in the updating process; consequently, there would be a noticeable difference between the full

Figure 4.4: Approximate full and empirical predictive densities (positive stable frailty)



| marks predicting value $Y(51,1)$ ^ marks observed predicted value $Y(51,2)$
 number is log-likelihood ratio evaluated at observed $Y(51,2)$

and PEB predictors under gamma frailty.

As we will see in the next chapter, predictive densities, while useful in their own right, can also be used as a tool in model selection criteria.

Chapter 5

Application

5.1 Description of Rat-Litter Data

We consider the data from a litter-matched experiment on rats, given in Mantel, Bohidar & Ciminera (1977). This data was analyzed in Hougaard (1986b), using likelihood methods. One-hundred litters of 3 animals each were observed. Fifty litters contained all female rats, the other 50 were exclusively male. One rat per litter was exposed to a possibly carcinogenic treatment. For each rat, a time (in weeks) and an indicator (T or D) was recorded. The data in its original form is given in Appendices B and C.

The times recorded are times from birth to a response. The response is appearance of a tumour (T), or death (D). The data description indicates that all animals were sacrificed after 104 weeks; however, it is clear that some animals (the surviving male rats listed as third in their litter) were sacrificed after 102 weeks. We will refer to deaths not due to sacrifice as deaths due to other causes. For ease of analysis, all times were divided by 100. Initially, we concentrate only on the female litters.

5.2 Single Risk versus Competing Risks

If the phenomenon of interest is tumour appearance, then we can treat those animals observed to die as censored observations. This was the approach taken in Hougaard (1986b). Note that this implies a mixture of censoring mechanisms. We treat each animal as having a time to tumour appearance T , and a censoring time L . Then either

we observe a value of T smaller than L (uncensored observation) or we know only that T exceeds L (censored observation). Following Lawless (1982, pp. 31-34), the sacrificed animals correspond to type I censoring (L is fixed before the experiment begins), while the animals observed to die correspond to random censoring (L is a random variable). For the conditional likelihood (2.3) to remain valid under random censoring, it is necessary to assume T and L are independently distributed given the frailty.

Of course it is entirely possible that death due to other causes (not tumours) is the response of interest, and we could treat tumour appearances as random censoring. The sacrificed animals would still correspond to type I censoring. As before the validity of (2.3) would require conditional independence of the tumour appearance time and the death due to other cause time.

It may be the case that both responses (tumour and death due to other causes) are of interest. Or, with data on families, it may be advantageous to model both responses, even if only one is of interest. We then require a competing risks model. Here, associated with each animal is a pair of random variables, T_1 (time to tumour appearance) and T_2 (time to death due to other causes). We observe either $(T_1 = t, T_2 > t)$ or $(T_1 > t, T_2 = t)$ in addition to type I censored observations $(T_1 > t, T_2 > t)$. Note that assuming independence of (T_1, T_2) given an animal's frailty is equivalent to the assumption required to justify the conditional likelihood in the single risk models. We explore several competing risk models based on this assumption.

5.3 Model Adequacy Checking

We would like to do some simple preliminary model checking with the data. One means by which to do this is examining marginal distributions. First, given a conditional Weibull distribution, we examine the unconditional marginal distribution induced by the frailty

distribution. Specifically, if the conditional hazard is $\theta\lambda_1(\lambda_3 t^{\lambda_3-1})$ it is easy to show that a gamma frailty distribution yields marginal hazard

$$\frac{\alpha\lambda_3 t^{\lambda_3-1}}{\frac{\alpha}{\lambda_1} + t^{\lambda_3}} \quad (5.23)$$

while the marginal hazard under positive stable frailty is

$$\lambda_1^{\lambda_0} (\lambda_0 \lambda_3 t^{\lambda_0 \lambda_3 - 1}) \quad (5.24)$$

So the positive stable frailty model has Weibull marginals as well as conditionals, but the shape parameters differ. Hence the marginal hazard is monotone (increasing, constant, or decreasing). Conversely with gamma frailty, the marginal hazard either increases to a maxima then decreases, or is monotone decreasing.

To do diagnostic marginal checks, we treat the female treatment rats as one sample, the first-listed female control rats as a second sample, and the second-listed as a third sample. Treating a single response as the response of interest, we proceed to check the adequacy of a Weibull distribution for each sample. The random component of the censoring means we don't know the ranks of even the uncensored observations, making a Weibull probability plot impossible. So, as suggested in Lawless (1982, p. 82) for each sample we plot $\log(-\log(\hat{S}(t)))$ as a function of $\log t$, where \hat{S} is the non-parametric Kaplan-Meier estimate of the survivor function. A true underlying Weibull marginal will manifest itself as a roughly linear plot with slope depending solely on the shape parameter. So, obvious non-linearity would contradict the Weibull/positive stable model, while linearity, but different slopes in treatment and control plots would contradict the assumption of a common shape parameter (conditional common shape λ_3 implies marginal common shape $\lambda_0 \lambda_3$ under stable frailty).

We would like to know how the Weibull/gamma model manifests itself on the above plot. It is possible to show that with this model, $\log(-\log S(t))$ will be increasing and convex for large $\log t$ (Appendix D).

Table 5.6: Estimation with rat–litter data : single risk (tumour)

gamma frailty			posterior mode (MLE)	approx. posterior mean
frailty parameter		λ_0	.61 (.29)	.58 (.25)
baseline	treatment	λ_1	.64 (.17)	.72 (.20)
hazard	control	λ_2	.260 (.068)	.288 (.080)
parameters	common shape	λ_3	3.93 (.57)	4.05 (.57)
positive stable frailty				
frailty parameter		λ_0	.906 (.095)	.871 (.083)
baseline	treatment	λ_1	.55 (.14)	.57 (.14)
hazard	control	λ_2	.214 (.058)	.223 (.059)
parameters	common shape	λ_3	4.10 (.63)	4.23 (.64)

The above-mentioned plots for both responses are given in Figure 5.5. The tumour plots appear to be roughly linear with common slope, supporting our Weibull/positive stable model. The other cause plots seem to show a slight concavity for large $\log t$, indicating perhaps the suitability of gamma frailty rather than stable frailty for this response. Of course this is appropriate only if the assumption of Weibull conditional hazards is appropriate.

5.4 Single Risk : Tumour Development

Using the Weibull model with conditional hazards $\theta\lambda_1(\lambda_3 t^{\lambda_3-1})$ for treatment animals and $\theta\lambda_2(\lambda_3 t^{\lambda_3-1})$ for controls, we compute the MLE and approximate standard errors in the usual fashion. We also compute approximate posterior means, and the square roots of approximate posterior variances with respect to Π_1 , the flat prior. Results for gamma frailty and for positive stable frailty are displayed in Table 5.6. The approximate posterior quantities were computed using the original parameterization. Given the success of using reparameterization II in the simulation study, we wished to try that technique here too.

Figure 5.5: Diagnostic plots based on Kaplan-Meier estimates of the marginals

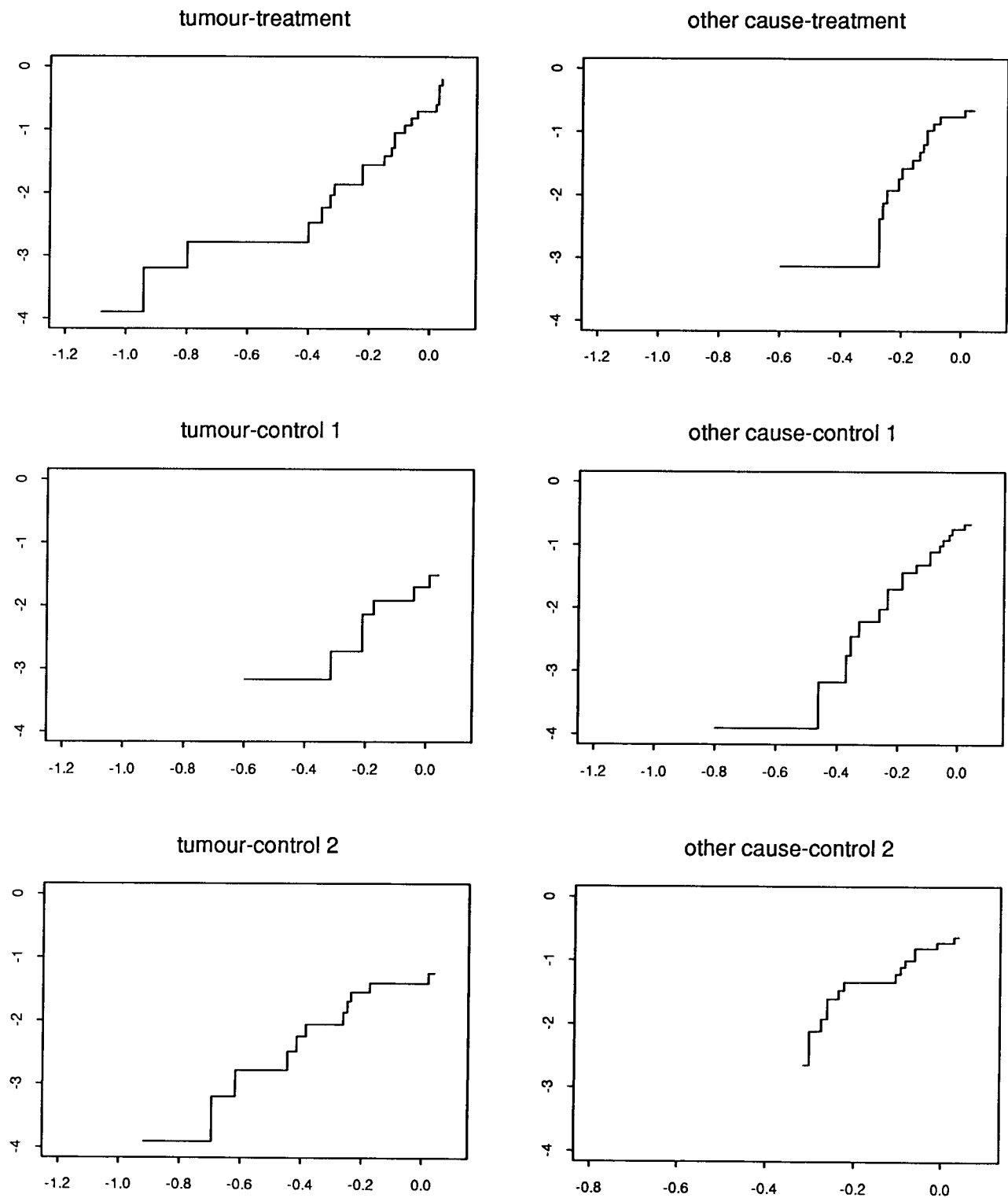


Table 5.7: Estimation with rat–litter data : single risk (other causes)

gamma frailty			posterior mode (MLE)	approx. posterior mean
frailty parameter		λ_0	.60 (.21)	.56 (.19)
baseline hazard parameters	treatment	λ_1	.54 (.16)	.62 (.19)
	control	λ_2	.51 (.11)	.55 (.13)
	common shape	λ_3	5.39 (.69)	5.55 (.70)
positive stable frailty				
frailty parameter		λ_0	.876 (.085)	.854 (.081)
baseline hazard parameters	treatment	λ_1	.44 (.12)	.47 (.13)
	control	λ_2	.392 (.084)	.408 (.086)
	common shape	λ_3	5.57 (.75)	5.70 (.75)

Unfortunately, we couldn't maximize the necessary modified likelihoods. In particular, the induced prior $\Pi'(\kappa)$ increases without bound as κ_0 decreases to 0. It would seem that the likelihood does not decrease commensurably quickly.

5.5 Single Risk : Death due to Other Causes

The same analysis applied to the tumour appearance times above was applied to the other cause times. The results are displayed in Table 5.7. Note that unlike the tumour risk, there is no discernible difference between treatment and control groups with other cause risk.

5.6 Competing Risks : Single Frailty

A simple approach to the competing risks problem is to assume that frailty is not risk-specific. Consequently, an animal's frailty acts multiplicatively on both its tumour hazard and its other cause hazard. Since treating the risks separately has produced similar

frailty parameter estimates for both risks, we have no immediate grounds to reject this assumption. The model structure remains the same; however, we now effectively have families of six members each (two lifetimes per animal), and three groups. The first group consists of tumour appearance times for treatment animals, the second group is tumour times for control animals, while the third group is other cause times. The conditional hazards are $\theta h(t)$, where h is group-dependent, according to:

$$h(t) = \begin{cases} \lambda_1(\lambda_3 t^{\lambda_3-1}) & \text{for group 1;} \\ \lambda_2(\lambda_3 t^{\lambda_3-1}) & \text{for group 2;} \\ \lambda_4(\lambda_5 t^{\lambda_5-1}) & \text{for group 3.} \end{cases} \quad (5.25)$$

So each family has one observation on group 1, two observations on group 2, and three observations on group 3. However, at most three observations from one family can be uncensored, since at most one lifetime per animal can be uncensored. Note that we assume the treatment has no effect on the other cause hazard (consistent with the estimation results from the last section). On the other hand, we do allow for different shape parameters for the different risks. The results of estimation in this model are given in Table 5.8.

5.7 Competing Risks : Two Frailties

There is no reason in general to expect that non-risk-specific frailty is realistic. On the other hand, assuming an individual has independent risk-specific frailties does not allow for an overall general health effect. A middle ground is suggested by Hougaard (1986b), in the context of distinguishing between sibling and twin relationships. Following his idea, we associate with an individual a general frailty θ , and risk-specific frailties ϕ_1 and ϕ_2 , corresponding to tumour appearance and other cause risk respectively. Generalizing

Table 5.8: Estimation with rat-litter data : competing risks, single frailty

gamma frailty			posterior mode (MLE)	approx. posterior mean
frailty parameter		λ_0	.78 (.15)	.72 (.14)
tumour appearance baseline	treatment	λ_1	.65 (.16)	.72 (.18)
	control	λ_2	.262 (.065)	.290 (.074)
	common shape	λ_3	3.98 (.57)	4.13 (.59)
other cause baseline	shape	λ_4	.508 (.088)	.55 (.10)
		λ_5	5.30 (.67)	5.45 (.67)
positive stable frailty				
frailty parameter		λ_0	.911 (.063)	.880 (.060)
tumour appearance baseline	treatment	λ_1	.59 (.14)	.63 (.15)
	control	λ_2	.228 (.055)	.241 (.058)
	common shape	λ_3	4.29 (.61)	4.43 (.61)
other cause baseline	shape	λ_4	.449 (.069)	.464 (.074)
		λ_5	5.52 (.68)	5.67 (.68)

from the previous section, the conditional hazards become

$$h(t \mid \theta, \phi_1, \phi_2) = \begin{cases} \theta \phi_1 \lambda_1 (\lambda_3 t^{\lambda_3 - 1}) & \text{for group 1;} \\ \theta \phi_1 \lambda_2 (\lambda_3 t^{\lambda_3 - 1}) & \text{for group 2;} \\ \theta \phi_2 \lambda_4 (\lambda_5 t^{\lambda_5 - 1}) & \text{for group 3.} \end{cases} \quad (5.26)$$

It is assumed that (ϕ_1, ϕ_2) are independently distributed according to a positive stable distribution parameterized by λ_6 , while θ^{λ_6} independently follows a positive stable distribution parameterized by λ_0 . Unfortunately, it appears this structure is chosen to make integration to the unconditional likelihood tractable, rather than for physical reasons. It is easy to show that (suppressing the family index), the unconditional survivor function is

$$P(Y_1 > y_1, \dots, Y_6 > y_6) = \exp \left(- \left[H_a^{\lambda_6} + H_b^{\lambda_6} \right]^{\lambda_0} \right) \quad (5.27)$$

where $H_a + H_b = H$, such that H_a contains the terms corresponding to tumour times, while H_b 's terms correspond to other cause times.

For our application, we can get the likelihood by taking partial derivatives (up to third order) of (5.27) with respect to (y_1, \dots, y_6) . Unfortunately, there does not seem to be an obvious recursive form for the likelihood as in the single frailty case. The derivatives required are given in Appendix E. The results from fitting the model are given in Table 5.9.

Note that the two frailty model generalizes the single frailty with stable distribution model in the following sense. The net frailty associated with each individual, $\theta\phi_i$, follows a positive stable distribution (with parameter $\lambda_0\lambda_6$). So for example, considering groups 1 and 2 only in (5.26) reduces to the two group model with a single frailty following a stable distribution. So an analogous generalization of the gamma frailty model could be achieved by taking any joint distribution on (θ, ϕ_i) such that the product $\theta\phi_i$ has a gamma distribution. However, for physical reasons, one may wish to impose other conditions such as independence between θ and ϕ_i , or perhaps certain marginal distributions. A convenient specification yielding gamma distributed net frailty was not found. For a general discussion of these types of mixture models, from the point of view of specified marginals, see Joe (1991).

5.8 Including Covariates

We propose to include the data on male litters by means of a regression function acting multiplicatively on an animal's conditional hazard, in the manner of Cox (1972). Since there are only two observed tumours in the male animals, we choose to model the other cause hazard only. Because of the binary nature of the covariate, we can generalize the model of section 5.5 by introducing a multiplicative factor λ_4 in the conditional hazards

Table 5.9: Estimation with rat-litter data : competing risks, two frailties

positive stable frailty			posterior mode (MLE)	approx. posterior mean
frailty parameters		λ_0	.938 (.062)	.897 (.064)
		λ_6	.945 (.071)	.942 (.071)
baseline hazard parameters	treatment	λ_1	.56 (.14)	.60 (.15)
	control	λ_2	.216 (.056)	.231 (.059)
	common shape	λ_3	4.29 (.61)	4.46 (.63)
other cause baseline		λ_4	.424 (.076)	.445 (.082)
		λ_5	5.66 (.71)	5.79 (.73)

of male rats only. The estimation results from this model are given in Table 5.10.

5.9 Assessing Model Choice

To compare choice of model and frailty distribution, we propose to use predictive methods. Following the cross-validation idea of Geisser (1980a), for each uncensored observation, we construct a predictive density conditional on all the other observations. For example, to do this for the 40 tumour appearances observed would require 40 different posteriors, the i^{th} posterior being based on all the data except the i^{th} tumour appearance time. In practice we use the posterior based on all the data to compute each predictor, as deleting one of 150 observations should not greatly influence the posterior. To compute the predictive densities, we use importance sampling with multinormal samples of size 1000, as described in the previous chapter. Using these densities to assess model choice is an application of the idea of predictive sample reuse, introduced independently by Geisser (1975) and Stone (1974).

As a measure of the quality of prediction, we take the log of the product of the predictive densities evaluated at the observed values. This choice of measure of prediction is inspired by the idea of comparisons based on likelihood ratios. It is also suggested in

Table 5.10: Estimation with rat-litter data : single risk (other cause), proportional hazards model

gamma frailty			posterior mode (MLE)	approx. posterior mean
frailty parameter		λ_0	.75 (.13)	.71 (.13)
baseline hazard parameters	treatment	λ_1	.48 (.10)	.51 (.11)
	control	λ_2	.489 (.087)	.511 (.085)
	common shape	λ_3	4.87 (.40)	4.94 (.42)
regression parameter		λ_4	1.47 (.31)	1.50 (.34)
positive stable frailty				
frailty parameter		λ_0	.895 (.050)	.878 (.053)
baseline hazard parameters	treatment	λ_1	.401 (.089)	.413 (.093)
	control	λ_2	.415 (.077)	.422 (.079)
	common shape	λ_3	5.14 (.43)	5.20 (.44)
regression parameter		λ_4	1.48 (.33)	1.53 (.36)

Geisser & Eddy (1979). Their interpretation is that of the models considered, the one maximizing this predictive score best explains the data. These predictive scores for all the models considered are given in Table 5.11. Note that for the purposes of comparison, the score for the model including males is based on prediction of the female survival times only.

So, the predictive scores indicate superior performance by the stable frailty distribution in all but the proportional hazards model. To investigate this phenomenon, we give histograms of the individual ratios of stable predictive density to gamma predictive density evaluated at the observed failure times. These are given for the single risk models and the competing risk single frailty model in Figure 5.6. Note that in three of the four models, the superior performance of stable frailty can be attributed to one or two relatively large predictive density ratios, while the remaining observations show relatively even performance between the two frailty distributions, with any edge in fact favouring

Figure 5.6: Histograms of predictive density ratios : positive stable frailty versus gamma frailty

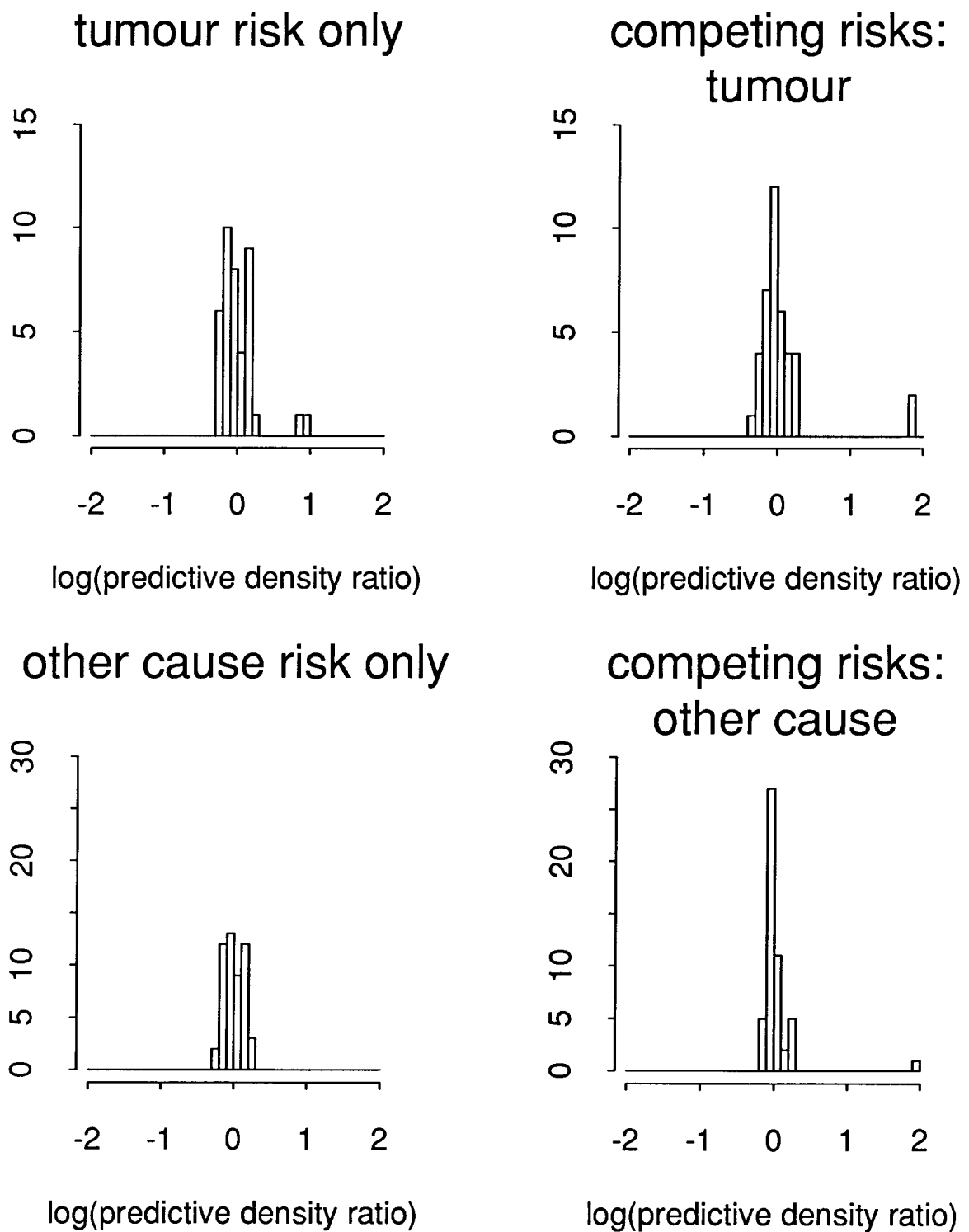


Table 5.11: Predictive scores

model	tumour appearance			other cause	
	gamma frailty	stable frailty		gamma frailty	stable frailty
tumour risk only	-24.71	-24.18			
other cause risk only				-8.93	-8.71
other cause risk only including males				-10.16	-10.26
Competing risks single frailty	-25.49	-22.80		-9.44	-7.51
Competing risks two frailties		-22.92			-8.04

the gamma distribution. Again it seems the stable frailty choice is robust, relative to the gamma frailty specification.

Also it is clear that on the basis of these predictive scores, the single frailty, competing risks model with stable frailty is the most suitable, as it gives the best prediction of both tumour appearance and other cause failure times. Note that while this model with stable frailty improves upon the single risk models, the same model with gamma frailty is worse than the single risk models. This underscores the importance of the frailty distribution selection—different choices can lead to different conclusions about the manner in which frailty acts upon individuals hazards.

The double frailty model exhibits slightly poorer performance than the best model, even though it is a generalization. Perhaps the ad-hoc assumptions concerning frailty distributions in this model are inappropriate at least in this situation. The assumption of exchangeable risk-specific frailties is perhaps too restrictive. A model with more flexible inclusion of general and risk-specific frailties might exhibit better performance; however,

such a model would be computationally difficult.

Including the male rats in the other risk only model via proportional hazards does not improve the prediction of the female other cause times. This certainly calls into question the assumption of conditional proportional hazards between male and female animals.

Computing a predictive score for each model provides a very easy way to choose between several models for given data (the model with the highest score is most suitable). Of course there are other model-choosing criteria. Some of these are mentioned in the next chapter.

Chapter 6

Discussion

6.1 Approaches to Frailty Models

As noted in (2.1), the hierarchical model can be collapsed to a one-stage model. When the n_i are small (the case with which we have been concerned), there seems to be little point in trying to estimate the θ_i 's. In a sense the two-stage model is only motivation for the one-stage likelihood. However, it is important to keep in mind the notion of underlying frailties. As we saw in Chapter 4, the knowledge about an individual's frailty is fundamental to predicting that individual's failure time. Moreover, in Chapter 5, we saw that the concept of frailty led to more sophisticated models, with competing risks, and several frailties. Of course, if the Laplace transform of the frailty density does not have a nice form, we could not straightforwardly collapse to a one-stage model. In this instance the hierarchical nature of the model would be more visible (though no more important).

We also saw that these models could be approached from the perspective of say bivariate models, with given marginals. Here obviously model specification would be in terms of marginal rather than conditional (baseline) distributions. This has some disadvantages. Specifying marginals is specifying an effect averaged over a population. In analogy to the Cox (1972) model, it would be akin to specifying a form for the baseline hazard averaged with respect to the covariate vectors of the individuals. Conversely, specifying conditionals corresponds to choosing a form for the baseline hazards. The

latter approach seems more sensible. Additionally, specifying arbitrary marginals implies that the conditional (baseline) hazards depend on the frailty parameter(s). This can also be construed as undesirable.

Lindley & Singpurwalla (1986) motivate a model with time-varying frailties. In particular they suggest treating an individual frailty as a stochastic process. They point out that assuming frailty is a purely random process with a stationary distribution results in a model analogous to ours, but with different interpretation. Apparently, this approach has not been pursued in the literature.

6.2 Specification of Distributions

In practice, we have restricted ourselves to the use of Weibull distributions for the first stage of our model. Although this family of distributions is flexible, and commonly used for failure times, this is a somewhat arbitrary specification. For more generality, one could use the generalized gamma distribution of Stacy (1962). This three parameter family includes the Weibull and gamma families, and has a log-normal distribution as a limiting case. Unfortunately, an additional parameter will have ramifications in terms of computational practicality.

In light of the existing literature, we have considered gamma and positive stable distributions as second-stage distributions. These families have the advantage of possessing nice Laplace transforms of their densities. In our data analysis, the stable frailty outperforms the gamma frailty, at least in terms of our predictive sample reuse criteria. As mentioned previously, this may be attributable to the robustness of the stable specification rather than a higher degree of “truthfulness.” This robustness in turn comes from the long right tail of positive stable distributions. In any case, at least according to our criteria, the stable frailty specification offers a better explanation of the data than the

gamma frailty. This is consistent with Hougaard's theoretical preference for the positive stable frailty.

There is however, a drawback to using stable frailty in terms of implementing the Bayesian paradigm. With gamma frailty, a specification of a prior on λ and particularly on λ_0 , describes belief in a simple function of the variance of the frailty distribution. With stable frailty, λ_0 is the index of the variate, and no such straightforward interpretations exist (the mean and variance of the frailty distribution do not exist). In short, with stable frailty, it is harder to encapsulate knowledge about the heterogeneity of a population.

Again, one wonders about other frailty distribution specifications. Hougaard (1986a) introduces a three-parameter family generalizing both the positive stable and gamma distributions. However, using this entire family would present computational difficulties, and possibly problems with identifiability. An alternate family with the property of closed form Laplace and inverse Laplace transforms is the inverse Gaussian densities. However, inverse Gaussian frailty models cannot attain the Fréchet upper bound. In fact, the limiting distribution corresponds to positive stable frailty with parameter $\lambda_0 = 1/2$ (Oakes, 1989). Hence this frailty distribution specification would not encompass as wide a range of dependence/association as gamma or positive stable frailty.

There are some nonparameteric and semiparametric procedures which can be used for multivariate frailty models. Oakes (1986,1989) discusses such procedures. However, they are presently limited to bivariate situations with identical groupings within each family. Also, they do not provide for the specification of prior information.

6.3 Techniques

To estimate parameters, we used both maximum likelihood estimation and Laplace's method with diffuse priors. In the data analysis, there were no cases of these two methods providing greatly differing results. In the simulation studies, the overall performance of the MLE was perhaps superior; however, at least in one instance Laplace's method provided the better results. Of course, if one has non-diffuse prior information then Laplace's method would provide reliable estimates of functions of parameters. The additional advantage of Laplace's method is its simplicity—it rivals maximum likelihood estimation in terms of ease of implementation.

We also witnessed the dependence of Laplace's method on parameterization. A fixed posterior function can be approximated via any parameterization of the model. Due to the nature of Laplace's method, the approximation will be better with a parameterization in which the posterior distribution is in some sense closer to a normal distribution. In fact if the parameter space were 1 or 2-dimensional, one could visually inspect the (non-normalized) posterior under a variety of parameterizations. The parameterization which induces the most “normal looking” posterior would be a good choice to use for Laplace approximations. Of course in our situation (and most other realistic situations) the parameter space is of higher dimension. Achcar & Smith (1990) give examples of the sensitivity of Laplace approximations to reparameterization in a number of situations.

The Laplace approximation was not so attractive for approximating predictive densities. For this purpose, importance sampling gave us normalization essentially for free. This was due to the nice form of the conditional densities (PEB predictive densities), which in turn was due to the nice form of the Laplace transform of the frailty density.

Another Monte Carlo technique which is potentially useful for approximating posterior quantities in hierarchical models is Gibbs sampling. For a general discussion of

Gibbs sampling, see Geman & Geman (1984), or Gelfand & Smith (1988). In CIHM's, Gibbs sampling would require simulating variates from the distributions of $(\theta_i \mid \lambda, \text{data})$, $i = 1, \dots, k$. and $(\lambda_j \mid \lambda_{(j)}, \theta, \text{data})$, $j = 1, \dots, r$. In the Weibull model with gamma frailty, $(\theta_i \mid \lambda, \text{data})$ will also have a gamma distribution (gamma frailty is effectively a conjugate prior specification). However, $(\lambda_j \mid \lambda_{(j)}, \theta, \text{data})$ would have a complicated distribution function. Thus in our particular hierarchical model, Gibbs sampling would be very difficult to implement.

6.4 Assessing Models

Chapter 5 illustrated several ways of modelling a data set. Of particular interest was the difference between the single risk models and the competing risk models. According to our criteria, (predictive sample reuse), the competing risk models fared better, in terms of predicting both tumour appearance times and other cause times. Put another way, with data on families, even if say tumour appearance is the response of interest, it is advantageous to model the censoring times (in this case the other cause times).

The criteria used to assess models is also of interest. The predictive sample reuse idea is conceptually appealing. In a sense it favours the model which offers the best explanation of the data. There is not such a strong emphasis on the truth of various models, which is an approximate notion at best. There is also not the reliance on asymptotics which is used for likelihood assessment with nested models.

There are many other criteria available to assess the different models. Even within the framework of predictive sample reuse, there are many ways to implement a model-selection scheme. Geisser (1980a) has suggested making point predictions and using some discrepancy function to measure the distance between predicted and realized values. An entirely different approach would be to use Bayes factors. When comparing two models

(each consisting of a likelihood L and a prior Π), the Bayes factor is given by

$$\frac{\int L_1(\lambda)\Pi_1(\lambda)d\lambda}{\int L_2(\lambda)\Pi_2(\lambda)d\lambda}$$

with a large Bayes factor favouring model 1 (Berger, 1985, p. 146). In our models, computing Bayes factors would be difficult, requiring the integration of the unconditional likelihood (2.5) with respect to the prior measure Π on λ . Also, with diffuse priors, we expect Bayes factors to be very similar to likelihood ratios. Because of the behaviour with nested models, we wish to avoid likelihood ratio criteria.

When likelihood ratio is applied to nested models, the more complex model will necessarily have a larger likelihood, or score. We wish to avoid the appeal to asymptotics usually made to resolve this problem. Predictive sample reuse and Bayes factors will not necessarily favour a more complex model. Other approaches involve modifying the likelihood ratio. Akaike's information criterion (Akaike, 1973) uses

$$\log L(\hat{\lambda}) - [\# \text{ estimable parameters}]$$

as a score with which to compare models. ($\hat{\lambda}$ is a maximum likelihood estimate.) The approach of Schwarz (1978) modifies the criterion to

$$\log L(\hat{\lambda}) - [(1/2) \log n \times (\# \text{ estimable parameters})]$$

where n is the number of independent observations. Of course, these latter two criteria do not admit prior information.

These criteria are not entirely dissimilar. Smith & Spiegelhalter (1980) show a connection between Bayes factors and the Schwarz criterion in certain contexts. Also, Stone (1977) has shown an asymptotic equivalence between a variant of the predictive sample reuse criterion and the Akaike information criterion. The variant in question is similar to the one employed by us, but replaces the full predictive density with the PEB predictive density.

We note that in our data analysis, the single frailty competing risks model is a special case ($\lambda_6 = 1$) of the double frailty model. Thus we have an example of a predictive sample reuse criteria directly favouring the simpler of two nested models.

Our treatment of censoring in terms of predictive sample reuse bears discussion. We have based our criteria on prediction of the uncensored observations only. This allows for direct comparison between single risk and competing risk models. For methods of incorporating the censored values, see Geisser (1980b).

Appendix A

Quadrant Dependence of Frailty Models

Following (2.11), let p denote the Laplace transform of the frailty distribution. We know p is convex. We will show that if $\log p$ is convex, then for arbitrary first stage distributions, the bivariate distribution is positively quadrant dependent. We will need the facts $p(0) = 1$, $p \geq 0$, and $p' < 0$; these are elementary properties of Laplace transforms.

Note that to show (2.12) holds, it would suffice to show

$$p(a+b) \geq p(a)p(b) \quad \forall a, b > 0,$$

or equivalently

$$\log p(a+b) \geq \log p(a) + \log p(b).$$

But

$$\log p(a+b) = \log p(a) + \int_a^{a+b} \frac{p'(s)}{p(s)} ds.$$

Now we know that the integrand is negative, and by assumption it is increasing, hence

$$\log p(a+b) \geq \log p(a) + \int_0^b \frac{p'(s)}{p(s)} ds;$$

that is

$$\log p(a+b) \geq \log p(a) + \log p(b) - \log p(0).$$

But $\log p(0) = 0$, giving the desired result.

Appendix B

Female Rat Litter Data

Female Rat Litters		
Drug-treated	Control-1	Control-2
101-D	49-T	104-D
104-D	102-D	104-D
104-D	104-D	104-D
77-D	97-D	79-D
89-D	104-D	104-D
88-T	96-T	104-D
104-D	94-D	77-T
96-T	104-D	104-D
82-D	77-D	104-D
70-T	104-D	7-D
89-T	91-D	90-D
91-D	70-D	92-D
39-T	45-D	50-T
103-T	69-D	91-D
93-D	104-D	103-D
85-D	72-D	104-D
104-D	63-D	104-D
104-D	104-D	74-D
81-D	104-D	69-D
67-T	104-D	68-T
104-D	104-D	104-D
104-D	104-D	104-D
104-D	83-D	40-T
87-D	104-D	104-D
104-D	104-D	104-D

Female Rat Litters		
Drug-treated	Control-1	Control-2
89-D	104-D	104-D
78-D	104-D	104-D
104-D	81-T	64-T
86-T	55-T	94-D
34-T	104-D	54-T
76-D	87-D	74-D
103-T	73-T	84-T
102-T	104-D	80-D
80-T	104-D	73-D
45-T	79-D	104-D
94-T	104-D	104-D
104-D	104-D	104-D
104-D	101-T	94-D
76-D	84-T	78-T
80-T	81-T	76-D
72-T	95-D	104-D
73-T	104-D	66-T
92-T	104-D	102-T
104-D	98-D	73-D
55-D	104-D	104-D
49-D	83-D	77-D
89-T	104-D	104-D
88-D	79-D	99-D
103-T	91-D	104-D
104-D	104-D	79-T

Appendix C

Male Rat Litter Data

Male rat litters		
Drug-treated	Control-1	Control-2
91-D	104-D	102-D
91-D	104-D	102-D
98-D	62-D	77-D
91-D	98-D	76-D
104-D	104-D	98-D
96-D	71-D	91-D
79-D	104-D	99-D
61-D	88-D	85-D
63-D	104-D	102-D
104-D	104-D	102-D
104-D	80-D	92-D
104-D	104-D	101-D
104-D	53-D	102-D
65-D	104-D	91-D
86-D	104-D	75-D
104-D	100-D	102-D
95-D	104-D	95-D
104-D	104-D	102-D
104-D	93-D	80-D
92-D	98-D	83-D
104-D	89-D	89-D
63-D	32-D	51-D
104-D	98-D	78-D
104-D	104-D	102-D
87-D	104-D	94-D

Male rat litters		
Drug-treated	Control-1	Control-2
104-D	104-D	102-D
104-D	91-D	102-D
90-D	104-D	55-D
91-D	104-D	102-D
104-D	104-D	102-D
23-D	104-D	102-D
104-D	71-T	90-D
87-D	51-D	102-D
104-D	83-D	102-D
104-D	104-D	96-D
67-D	84-D	94-D
104-D	104-D	99-D
104-D	94-D	102-D
104-D	103-D	102-D
51-D	104-D	91-D
102-D	98-D	102-D
88-D	54-D	39-D
67-D	84-D	54-D
104-D	104-D	87-D
81-D	82-D	102-D
94-D	104-D	102-D
104-D	89-D	77-D
104-D	69-D	102-D
104-D	75-T	64-D
92-D	104-D	102-D

Appendix D

Performance of Weibull/Gamma Model on Diagnostic Plot

Letting $y = \log[-\log S(t)]$ and $x = \log t$, from the Weibull/gamma marginal hazard (5.23), we see

$$y = \log \alpha + \log(\log g(x))$$

where

$$g(x) = k + \exp(\lambda_3 x), \quad k = \alpha/\lambda_1$$

We wish to show that for large x , dy/dx is positive, while d^2y/dx^2 is negative.

Noting that $g'(x) = \lambda_3[g(x) - k]$, we can write

$$\frac{dy}{dx} = \lambda_3 \frac{1 - (k/g)}{\log g}$$

which will be positive when $g > 1$. Since g is increasing this will occur when x is sufficiently large.

Differentiating again, we can write

$$\frac{d^2y}{dx^2} = \frac{\lambda_3^2}{[g \log g]^2} \left\{ k[g \log g] - k^2[\log g] - [(g - k)^2] \right\}$$

which we can see will be negative when g (and hence x) is sufficiently large.

Appendix E

Derivatives of Survivor Function in Two-Frailty Model

We want the first, second, and third-order derivatives of (5.27) with respect to (y_1, \dots, y_6) . Initially, we will express these in terms of the derivatives of $H^* = H_a^{\lambda_6} + H_b^{\lambda_6}$. We denote (5.27) by $A = \exp(-(H^*)^{\lambda_0})$ and write $B = -\lambda_0(H^*)^{\lambda_0-1}$. Then the first-order derivative is

$$A_i = ABH_i^*$$

The second-order derivatives have the form

$$A_{ij} = A(H_i^* H_j^*) \left[B^2 + B^{(1)} + \frac{H_{ij}^*}{H_i^* H_j^*} B \right]$$

where $B^{(d)}$ is the d^{th} derivative of B with respect to H^* .

The third-order derivatives can be expressed as a sum of three terms:

$$\begin{aligned} A_{ijk} &= H_i^* H_j^* H_k^* AB[B^2 + B^{(1)} + \frac{H_{ij}^*}{H_i^* H_j^*} B] \\ &\quad + A(H_{ik}^* H_j^* + H_{jk}^* H_i^*)[B^2 + B^{(1)} + \frac{H_{ij}^*}{H_i^* H_j^*} B] \\ &\quad + A(H_i H_j)[2BB^{(1)}H_k^* + B^{(2)}H_k^* + \frac{d}{dy_k} \frac{H_{ij}^*}{H_i^* H_j^*} B] \end{aligned}$$

where

$$\frac{d}{dy_k} \frac{H_{ij}^*}{H_i^* H_j^*} B = \frac{H_{ijk}^*}{H_i^* H_j^*} B - \frac{H_{ij}^* H_{ik}^*}{(H_i^*)^2 H_j^*} B - \frac{H_{ij}^* H_{jk}^*}{(H_j^*)^2 H_i^*} B + \frac{H_{ij}^* H_k^*}{H_i^* H_j^*} B^{(1)}$$

The derivatives of H^* are given by

$$H_i^* = \lambda_6 H_{(i)}^{\lambda_6-1} h_i$$

$$H_{ij}^* = \begin{cases} \lambda_6(\lambda_6 - 1)H_{(i)}^{\lambda_6-2}h_i h_j & (i) = (j) \\ 0 & \text{otherwise} \end{cases}$$

and

$$H_{ijk}^* = \begin{cases} \lambda_6(\lambda_6 - 1)(\lambda_6 - 2)H_{(i)}^{\lambda_6-3}h_i h_j h_k & (i) = (j) = (k) \\ 0 & \text{otherwise} \end{cases}$$

where (i) indicates a or b depending on whether Y_i is a tumour response or an other cause response respectively.

Bibliography

- [1] ACHCAR, J.A. & SMITH, A.F.M.(1990). Aspects of reparameterization in approximate Bayesian inference. *Bayesian and Likelihood Methods in Statistics and Econometrics* 439-52. eds. S. Geisser, J.S. Hodges, S.J. Press, A. Zellner. Amsterdam: North-Holland.
- [2] AKAIKE, H.(1973). Information theory and an extension of the maximum likelihood principle. *2nd. International Symposium on Information Theory* 267-81. Budapest: Akademia Kiado.
- [3] BERGER, J.O.(1985). *Statistical Decision Theory and Bayesian Analysis* New York: Springer-Verlag.
- [4] BLOMQVIST, N.(1950). On a measure of dependence between two random variables. *Ann. Math. Statist.* **21**, 593-600.
- [5] CHAMBERS, J.M., MALLOWS, C.L. & STUCK, B.W.(1966). A method for simulating stable random variables. *J. Am. Statist. Assoc.* **71**, 340-44.
- [6] CLAYTON, D.(1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141-51.
- [7] CORNFIELD, J. & DETRE, K.(1977). Bayesian life table analysis. *J. R. Statist. Soc. B* **39**, 86-94.
- [8] COX, D.R.(1972). Regression models and life tables (with discussion). *J. R. Statist. Soc. B* **34**, 187-202.
- [9] COX, D.R. & OAKES, D.(1984). *Analysis of Survival Data* Chapman & Hall.
- [10] FELLER, W.(1971). *An Introduction to Probability Theory and Its Applications*, **2**, 2nd ed. New York: Wiley.
- [11] GAMMERMAN, D. & WEST, M. (1987). An application of dynamic survival models in unemployment. *Statistician* **36**, 269-74.
- [12] GEISSER, S.(1975). The predictive sample reuse method with applications. *J. Am. Statist. Assoc.* **70**, 320-38,350.

- [13] GEISSER, S.(1980a). A predictivist primer. *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*. 363-81. ed. A. Zellner. Amsterdam: North-Holland.
- [14] GEISSER, S.(1980b). Predictive sample reuse techniques for censored data. *Bayesian Statistics*. 433-68. eds. J.M. Bernardo et. al., Valencia: University Press
- [15] GEISSER, S. & EDDY, W.F.(1979). A predictive approach to model selection. *J. Am. Statist. Assoc.* **74**, 153-160.
- [16] GELFAND, A.E. & SMITH, A.F.M.(1988). Sampling based approaches to calculating marginal densities. *Technical report: Nottingham Statistics Group*. University of Nottingham.
- [17] GEMAN, S. & GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-41.
- [18] GENEST, C. & MACKAY, R.J.(1986a). Copules Archimediennes et familles de lois marges sont donnees. *Can. J. Stat.* **14**, 145-59.
- [19] GENEST, C. & MACKAY, R.J.(1986b). The joy of copulas: bivariate distributions with given marginals. *The American Statistician* **40** 280-83.
- [20] HOLT, J.D. & PRENTICE, R.L.(1974). Survival analysis in twin studies and matched-pair experiments. *Biometrika* **61** 17-30.
- [21] HOUGAARD, P.(1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* **71**, 75-83.
- [22] HOUGAARD, P.(1986a). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387-96.
- [23] HOUGAARD, P.(1986b). A class of multivariate failure time distributions. *Biometrika* **73**, 671-78.
- [24] JOE, H.(1991). Parametric families of multivariate distributions with given margins. *Technical report: Department of Statistics*. University of British Columbia.
- [25] KALBFLEISH, J.(1978). Non-parametric Bayesian analysis of survival time data. *J. R. Statist. Soc. B* **40**, 214-21.
- [26] KALBFLEISH, J. & PRENTICE, R.L.(1980). *The Statistical Analysis of Failure Time Data* New York: Wiley.

- [27] KANTER, M.(1975). Stable densities under change of scale and total variation inequalities. *Annals of Prob.* **3**, 697-707.
- [28] KASS, R.E. & STEFFEY, D.(1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J. Am. Statist. Assoc.* **84**, 717-26.
- [29] KENDALL, M.G.(1938). A new measure of rank correlation. *Biometrika* **30**, 81-93.
- [30] LAWLESS, J.F.(1982). *Statistical Models and Methods for Lifetime Data* New York: Wiley.
- [31] LEHMANN, E.L.(1966). Some concepts on dependence. *Ann. Math. Statist.* **37**, 1137-53.
- [32] LINDLEY, D.V. & SINGPURWALLA, N.D.(1986). Multivariate distributions for the life lengths of components of a system sharing a common environment. *J. Applied Prob.* **23**, 418-31.
- [33] MANTEL, N., BOHIDAR, N.R. & CIMINERA, J.L.(1977). Mantel-Haenszel analyses of litter matched time-to-response data, with modifications for recovery of interlitter information. *Cancer Res.* **37**, 3863-68.
- [34] MARSHALL, A.W. & OLKIN, I.(1988). Families of multivariate distributions. *J. Am. Statist. Assoc.* **83**, 834-41.
- [35] MARTZ, H.F. & WALLER, R.A.(1982). *Bayesian Reliability Analysis* New York: Wiley.
- [36] OAKES, D.(1986). A model for bivariate survival data. *Modern Statistical Methods in Chronic Disease Epidemiology* eds. R.L. Prentice and S.H. Moolgavkar. New York, Wiley.
- [37] OAKES, D.(1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487-93.
- [38] SCHWARZ, G.(1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461-64.
- [39] SMITH, A.F.M. & SPIEGELHALTER, D.J.(1980). Bayes factors and choice criteria for linear models. *J. R. Statist. Soc. B* **42**, 213-20.
- [40] SPEARMAN, C.(1904). The proof and measurement of association between two things. *Am. J. of Psychology* **15**, 72-101.
- [41] STACY, E.W.(1962). A generalization of the gamma distribution. *Ann. Math. Statist.* **33**, 1187-92.

- [42] STONE, M.(1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *J. R. Statist. Soc. B* **36**, 111-47.
- [43] STONE, M.(1977). An asymptotic equivalence of choice of model by cross-validation and Aikake's criterion. *J. R. Statist. Soc. B* **39**, 44-47.
- [44] TIERNEY, L. & KADANE, J.B.(1986). Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Assoc.* **81**, 82-86.
- [45] VAUPEL, J.W., MANTON, K.G. & STALLARD, E.(1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439-54.
- [46] WILD, C.J.(1983). Failure time models with matched data. *Biometrika* **70**, 633-41.