
Wilson KJ, Farrow M. [Bayes linear kinematics in a dynamic survival model](#).
International Journal of Approximate Reasoning 2017, 80, 239-356.

Copyright:

© 2016. This manuscript version is made available under the [CC-BY-NC-ND 4.0 license](#)

DOI link to article:

<http://dx.doi.org/10.1016/j.ijar.2016.09.010>

Date deposited:

11/10/2016

Embargo release date:

28 September 2017



This work is licensed under a
[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence](#)

Bayes linear kinematics in a dynamic survival model

K.J. Wilson^{a,*}, M. Farrow^a

^a*School of Mathematics and Statistics, Newcastle University, Newcastle upon Tyne, UK*

Abstract

Bayes linear kinematics and Bayes linear Bayes graphical models provide an extension of Bayes linear methods so that full conditional updates may be combined with Bayes linear belief adjustment. In this paper we investigate the application of this approach to survival analysis with time-dependent covariate effects, a more complicated problem than previous applications. We use a piecewise-constant hazard function with a prior in which covariate effects are correlated over time. The need for computationally intensive methods is avoided and the relatively simple structure facilitates interpretation. Our approach eliminates the problem of non-commutativity which was observed in earlier work by Gamerman. We apply the technique to data on survival times for leukemia patients.

Keywords: Bayes linear kinematics, Bayes linear Bayes graphical model, dynamic model, piecewise constant hazard, survival analysis, time-dependent covariate effects

1. Introduction

A Bayes linear analysis (Goldstein and Wooff, 2007) differs from a full Bayesian analysis in that only first and second order moments are specified in the prior. Posterior (termed *adjusted*) moments are then calculated when data are observed. The introduction of Bayes linear kinematics and Bayes linear Bayes models (Goldstein and Shaw, 2004) extends Bayes linear methods to allow the incorporation of observations of types which are not readily accommodated in a straightforward Bayes linear analysis. For example, beliefs about certain unknown quantities might be updated by full conditional Bayesian inference when observations are made on conditionally Poisson or binomial variables and then information can be propagated between these unknowns, or to other unknowns, via a Bayes linear belief structure. This approach avoids the need for computationally intensive methods such as Markov chain Monte Carlo which are often required in standard Bayesian analyses. Computational time in evaluating posterior distributions can be an important issue in areas such as design of experiments (Müller, 2004), clinical decision rules or evaluation of diagnostic tests. Such analyses may require the repeated evaluation

*Corresponding author

Email addresses: kevin.wilson@newcastle.ac.uk (K.J. Wilson),
malcolm.farrow@newcastle.ac.uk (M. Farrow)

of posterior distributions given large numbers of simulated data sets. In such cases, the Bayes linear kinematic method may provide an effective emulator (Jones et al., 2016).

Wilson and Farrow (2010) introduced the use of a link function to map the range of an unknown, such as the mean of a Poisson distribution, onto the whole real line and improve the linearity of the relationships represented by the Bayes linear structure. In this paper we show how Bayes linear kinematics may be applied to a more complicated problem, specifically in the analysis of survival data, and that this brings appealing advantages over standard techniques. Our analysis uses death and censoring times, in contrast to the relatively simple actuarial methods developed in Wilson and Farrow (2010). We use a piecewise constant hazards model with temporally-dependent hazard priors. We combine fully Bayesian conjugate updating for individuals in intervals and Bayes linear kinematic updating to propagate changes in belief to other individuals and intervals. Our model is related to that of Gamerman (1991) but, using the Bayes linear kinematic approach, we overcome the problem of non-commutativity of updates observed by Gamerman.

We consider Bayesian analysis from a subjectivist perspective (Goldstein, 2006; Lindley, 2006). Therefore we give attention to the appropriate specification of prior beliefs.

The remainder of the paper is structured as follows. Section 2 gives an overview of proportional hazards models and the piecewise constant hazards model and reviews the model of Gamerman (1991). In Section 3 we give a brief introduction to the results of Goldstein and Shaw (2004). In Section 4 we describe our Bayes linear kinematic solution to the survival problem in four stages; the guide relationship, system evolution, use of Bayes linear kinematics and calculation of the expectations and variances. The usefulness of the approach is illustrated with an example involving survival times of leukemia patients in the North-West of England in Section 5. Some conclusions and areas for further work are presented in Section 7.

2. Survival analysis

2.1. Introduction

In this paper we investigate the application of Bayes linear kinematics and Bayes linear Bayes models in survival analysis, specifically a proportional hazards model with piecewise constant hazards. Survival analysis is concerned with modelling the time elapsed, known as the survival time, until some event occurs. For convenience we shall refer to the event as “death”.

The survival time t of an individual is a realisation of a random variable T . Associated with T is a survival function $S(t) = \Pr(T \geq t)$, a probability density function $f(t)$ and a hazard function $h(t) = f(t)/S(t)$. Censoring of observations is a common feature of survival data. In right censoring all that is known is that $t > c$ for some value c , in left censoring this condition is $t < c$ and in interval censoring $c_1 < t < c_2$, for some values c_1, c_2 . In this paper we consider only right censoring, which is the most common type, and assume that the censoring is non-informative. That is, the survival time T is independent of the mechanism which causes an observation to be censored. Further information on Bayesian survival analysis can be found in Klein and Moeschberger (1997) and Ibrahim et al. (2001).

2.2. Proportional hazards models

Suppose we have individuals $i = 1, \dots, p$ and individual i has covariate values $x_i = (x_{i,0}, x_{i,1}, \dots, x_{i,q})'$ where, typically, $x_{i,0} \equiv 1$. Associated with individual i is a hazard function $h_i(t)$. A popular and appealing way to relate the covariate values to the survival distribution for an individual is to make the proportional hazards assumption (Cox, 1972). Then we can write $h_i(t) = \phi_i h_0(t)$, where ϕ_i is a constant with respect to time and $h_0(t)$ is a baseline hazard function. We can relate an individual's hazard function to x_i , the individual's covariate vector, by setting

$$\phi_i = \exp\left(\sum_{k=1}^q x_{i,k} \beta_k\right), \quad (1)$$

for some parameters β_1, \dots, β_q , which, in a simple proportional hazards model, remain constant over time.

2.3. Piecewise constant hazards model

We might be unwilling to assume a particular form for the baseline hazard function $h_0(t)$. A simple and much investigated way to relax this assumption is to use a piecewise constant hazards model (eg. Ibrahim et al., 2001). Time is partitioned into disjoint intervals. In each interval a constant hazard is specified but the hazards are allowed to vary from interval to interval.

Furthermore we may wish to allow the effects of the covariates, represented by the coefficients β_1, \dots, β_q , to vary from one time interval to another. This has led to the development of dynamic survival models in which the coefficients can vary over time (Martinussen and Scheike, 2006). We shall consider a dynamic model $h_i(t) = \exp\{x_i' \beta(t)\}$, where $x_i' = (1, x_{i,1}, \dots, x_{i,q})$ and $\beta'(t) = (\beta_0(t), \beta_1(t), \dots, \beta_q(t))$ with $\beta_0(t) = \log\{h_0(t)\}$, so that we can model changes in the effects of the covariates over time. The static model in (1) is then a special case of this more general model.

We choose fixed time points $\tau_0, \tau_1, \dots, \tau_r$ such that $\tau_0 = 0$ and $\tau_r \rightarrow \infty$. This partitions time into intervals. We say that the j^{th} interval is $R_j = [\tau_{j-1}, \tau_j)$. Then, for $\tau_{j-1} \leq t < \tau_j$, the baseline hazard is $h_0(t) = \lambda_{0,j}$ and the hazard function for individual i is $h_i(t) = \lambda_{i,j} = \phi_{i,j} \lambda_{0,j} = \exp(\eta_{i,j})$, where $\eta_{i,j} = x_i' \beta_j$ is the *linear predictor* and $\beta_j = (\beta_{j,0}, \dots, \beta_{j,q})'$. That is, the hazard for each individual remains constant through each of the time intervals. The integrated hazard $H_i(t) = \int_0^t h_i(u) du$ is then

$$H_i(t) = \sum_{k:\tau_k < t} \lambda_{i,k}(\tau_k - \tau_{k-1}) + \lambda_{i,j}(t - \tau_{j-1}),$$

for $k = 1, \dots, j-1$.

If we condition on $T \geq \tau_j$ then we obtain the conditional survival function and conditional probability density function for individual i at time t . These are, for $\tau_{j-1} \leq t < \tau_j$,

$$f_i(t | T \geq \tau_{j-1}) = \lambda_{i,j} \exp\{-\lambda_{i,j}(t - \tau_{j-1})\}, \quad (2)$$

and

$$S_i(t | T \geq \tau_{j-1}) = \exp\{-\lambda_{i,j}(t - \tau_{j-1})\}. \quad (3)$$

Thus the conditional density takes the form of a shifted exponential distribution.

In our prior distribution, the coefficients could be independent between time intervals (Kalbfleisch, 1978; Ibrahim et al., 2001). However, it would seem sensible that hazards in intervals which are close together are likely to be similar and so it would be beneficial to use a prior distribution in which hazard parameters in neighbouring intervals are correlated. Such a temporally-dependent prior has the effect, if sufficiently large correlations between intervals are used, of both smoothing the resulting posterior mean of the hazard function and providing extra information for later time periods in which there will be fewer individuals left in the study. Examples include Sinha et al. (1999) and Sargent (1997). In the latter case the interval boundaries are the observed event times. See also Sinha and Dey (1997).

Bayesian methods using fully-specified priors, and requiring intense computation, include those proposed by McKeague and Tighiouart (2000) and Kim et al. (2007). McKeague and Tighiouart introduced a method, with temporally-dependent priors using Markov random fields, which gave a prior distribution to the interval boundaries based on a non-homogeneous Poisson process. Updating was fully Bayesian and used the Metropolis-Hastings-Green algorithm. Kim et al. (2007) considered a temporally-dependent proportional hazards model in the context of cure fractions. Their method allowed a random number of intervals and random interval lengths. They used noninformative priors and utilised reversible jump Markov chain Monte Carlo (MCMC) methods for updating.

2.4. The Gamerman model

The temporally-dependent prior for the coefficients $\{\beta_{j,k}\}$ of the linear predictor of the hazard function in successive intervals may be built using a dynamic linear system evolution as in a dynamic linear model. West et al. (1985) introduced dynamic generalised linear models for non-Gaussian time series. Gamerman (1991) applied this idea to a piecewise constant hazards model as follows.

1. Each $\lambda_{i,j}$ was given a gamma prior distribution which was conjugate to the conditional likelihood for that individual in that interval. Within each interval, beliefs about individual hazards could then be updated straightforwardly using full Bayesian conditioning.
2. An evolving system vector, as in a dynamic linear model, was used to specify the prior distribution for the coefficients $\{\beta_{j,k}\}$. The joint distribution of the system vector at times τ_0, \dots, τ_r was not fully specified, just the first and second order moments. The vector β_j for interval j was given a mean vector m_j and a covariance matrix C_j . The covariances between intervals were specified using the system evolution. We call such a partial specification a second order belief specification.
3. A *guide relationship* between the parameters of the conjugate prior and the corresponding elements of the system vector was specified.

Within each interval, beliefs about $\lambda_{i,j}$ were updated when data were observed for that interval. This change in belief was then transmitted to an associated quantity $\eta_{i,j}$ using the guide relationship, denoted here by \approx , as follows,

$$\log(\lambda_{i,j}) \approx \eta_{i,j} = x'_i \beta_j.$$

As a result of the partial specification, the effects of changes in beliefs about $\eta_{i,j}$ were then propagated to other individuals' hazards in the interval and to other intervals using Bayes linear updating rules. Thus inference combined both fully Bayesian and Bayes linear updating. The structure led naturally to the use of a forward-filtering and backward-smoothing algorithm to compute posterior moments of the system vectors.

Gamerman found, however, that the calculated adjusted beliefs depended on the order in which data were included. 'The dependence on the order that the observations are processed is of concern...The results, however, do not differ by much'. Such non-commutativity is clearly a cause for concern. A fully Bayesian analysis with a fully specified prior distribution, for example a multivariate normal distribution in place of the linear Bayes structure, will, of course, give commutativity, but at the expense of greatly increased computation and the necessity of a full distributional specification for the prior. There is a wide range of problems where the idea of combining nonlinear, fully specified, Bayesian updates for separate parts of a model with a linear Bayes structure to connect the parts, in a Bayes linear Bayes graphical model, is attractive. Subject to certain conditions, Bayes linear kinematics (Goldstein and Shaw, 2004; Wilson and Farrow, 2010) provides a method for commutative inference in such structures.

3. Bayes linear kinematics

In a traditional Bayesian analysis a full joint prior distribution is specified for all observables and unknown quantities such as parameters. Prior beliefs are then updated, by conditioning on the observations and using Bayes theorem, and posterior distributions are calculated.

A Bayes linear analysis (Goldstein and Wooff, 2007) differs from a full Bayesian analysis in that only first and second order moments are specified in the prior. Posterior (termed *adjusted*) moments are then calculated. Consider two vector random quantities $A = (a_1, \dots, a_{n_A})'$ and $B = (b_1, \dots, b_{n_B})'$. Suppose that a full second order prior specification has been made for the set $A \cup B$, in which the prior expectations are $E_0(A) = E_{0,A}$ and $E_0(B) = E_{0,B}$ and the prior variances and covariances are $\text{Var}_0(A) = V_{0,A,A}$, $\text{Var}_0(B) = V_{0,B,B}$ and $\text{Cov}_0(A, B) = V_{0,A,B} = V_{0,B,A}'$.

Suppose that we will observe A . Bayes linear methods offer a procedure by which beliefs about B are updated by a process of linear fitting on A . To do this, we minimise $E_0 \{ (B - CA^*)'(B - CA^*) \}$, with respect to the matrix C , where $A^* = (1, a_1, \dots, a_{n_A})'$ and E_0 denotes prior expectation. This gives the Bayes linear updating equations for the adjusted expectation and variance of B given A :

$$\begin{aligned} E(B | A) &= E_{0,B} + V_{0,B,A} V_{0,A,A}^{-1} (A - E_{0,A}), \\ \text{Var}(B | A) &= V_{0,B,B} - V_{0,B,A} V_{0,A,A}^{-1} V_{0,A,B}, \end{aligned} \quad (4)$$

when $V_{0,A,A}$ is invertible. When this matrix is not invertible a suitable generalised inverse such as the Moore-Penrose inverse can be used.

Notice that an alternative representation of the same prior and adjusted beliefs is obtained by writing

$$B = E_{0,B} + M_{B|A}(A - E_{0,A}) + U_{B|A} \quad (5)$$

where $M_{B|A} = V_{0,B,A} V_{0,A,A}^{-1}$ and the random vector $U_{B|A}$ has zero expectation and its variance is $V_{0,B,B} - V_{0,B,A} V_{0,A,A}^{-1} V_{0,A,B}$.

Bayes linear kinematics (Goldstein and Shaw, 2004), named after the probability kinematics of Jeffrey (1965), deals with the case where, instead of observing the value of an unknown, such as A , in the Bayes linear structure, we receive information which causes us to change our beliefs about the unknown. Suppose that we receive information D_A which leads us to revise our moments for A to $E_1(A | D_A) = E_{1,A}$ and $\text{Var}_1(A | D_A) = V_{1,A,A}$. The Bayes linear kinematic update of our beliefs about $A \cup B$ is obtained by assuming that (5) continues to hold so that our adjusted beliefs for B become

$$E_1(B | D_A) = E_{0,B} + V_{0,B,A} V_{0,A,A}^{-1} (E_{1,A} - E_{0,A}), \quad (6)$$

$$\begin{aligned} \text{Var}_1(B | D_A) &= V_{0,B,A} V_{0,A,A}^{-1} V_{1,A,A} V_{0,A,A}^{-1} V_{0,A,B} \\ &\quad + V_{0,B,B} - V_{0,B,A} V_{0,A,A}^{-1} V_{0,A,B}. \end{aligned} \quad (7)$$

Consider the following simple example. We wish to perform a survival analysis for two groups of patients. Suppose that within each group the survival times of patients follow an exponential distribution with a common hazard rate and let this rate have a gamma prior distribution. Further suppose that the two hazard rates are correlated. That is, learning about the hazard rate in one group informs us about the hazard rate in the other group. Then, when we observe the survival times of a sample of patients from Group 1, this will allow us to update the hazard rate of that group using Bayes Theorem. Equations (6) and (7) then allow us to propagate this update through to the hazard rate in Group 2.

Now consider the case where we have J sets of random quantities $\mathcal{X}_1, \dots, \mathcal{X}_J$, where, for $j = 1, \dots, J$, the elements of \mathcal{X}_j are arranged as a vector $X_j = (X_{j,1}, \dots, X_{j,n_j})'$. The sets \mathcal{X}_j need not be disjoint. Suppose that a full second order prior specification has been made for $\mathcal{X} = \mathcal{X}_1 \cup \dots \cup \mathcal{X}_J$ and that the elements of \mathcal{X} are arranged in a vector X . We denote this prior specification $S_0(X) = [E_0(X), \text{Var}_0(X)]$. Let the adjustment implied by (4), if X_k is observed, be denoted $E_{(0)}(X | X_k)$ for the expectation and $\text{Var}_{(0)}(X | X_k)$ for the variance. Now suppose that information D_k is received which causes the beliefs about X_k to be updated to $S_1(X_k | D_k) = [E_1(X_k), \text{Var}_1(X_k)]$.

Suppose that, in this new situation, the adjustment, given an observation of X_k , would be $E_{(1)}(X | X_k)$ for the expectation and $\text{Var}_{(1)}(X | X_k)$ for the variance. Then the Bayes linear kinematic update for X is found by setting

$$E_{(0)}(X | X_k) = E_{(1)}(X | X_k), \quad \text{Var}_{(0)}(X | X_k) = \text{Var}_{(1)}(X | X_k),$$

that is, setting the adjusted expectations and variances using specifications $S_0(X)$ and $S_1(X | D_k)$ equal to each other. Specifically we can use (5) with A replaced by X_k and B replaced by X .

Now suppose that, for each j ($j = 1, \dots, J$), information D_j is received once and beliefs are changed for X_j . A Bayes linear kinematic update can be made for X each time. However, successive Bayes linear kinematic updates are not necessarily commutative. Once (6) and (7) have been used for an update given D_1 , the moments of the Bayes linear structure are changed and so the update by D_2 is changed. The Bayes linear kinematic method depends on the assumption that the updating formulae do not change, for example that (5) continues to hold. By straightforward repeated application of (6) and (7) we violate this assumption and it turns out that commutativity does not hold.

It is necessary to define Bayes linear kinematic formulae for updating by the whole of the data, based on an assumption analogous to the assumption that (5) continues to hold, but which will apply commutatively to intermediate steps, whatever the order in which we enter the data. Goldstein and Shaw (2004) derived the conditions under which a commutative Bayes linear kinematic update exists and under which this update is unique. When a unique commutative update exists, it is given by

$$P(X | D) = \sum_{j=1}^J P(X | D_j) - (J - 1)P(X), \quad (8)$$

$$P(X | D)E(X | D) = \sum_{j=1}^J P(X | D_j)E(X | D_j) - (J - 1)P(X)E(X), \quad (9)$$

where $D = (D_1, \dots, D_J)$ and $P(X) = \text{Var}(X)^{-1}$ is the precision matrix.

Goldstein and Shaw give formal proofs in the general case. We derive the results which we need for our purposes in the appendix.

Goldstein and Shaw (2004) introduced a structure which they called a Bayes linear Bayes graphical model. This is constructed as follows. We have a collection \mathcal{X} of unknowns which are given a full second-order prior specification so that Bayes linear belief adjustments may be made. There are certain subsets $\mathcal{X}_1, \dots, \mathcal{X}_J$ of \mathcal{X} where, for each j , the subset \mathcal{X}_j is also part of a model which contains other quantities \mathcal{Z}_j . A full joint probabilistic prior specification is made over $\mathcal{X}_j \cup \mathcal{Z}_j$. The sets $\mathcal{Z}_1, \dots, \mathcal{Z}_J$ are disjoint and every element of \mathcal{Z}_j is conditionally independent of every element of $\mathcal{X} \setminus \mathcal{X}_j$ and of every element of \mathcal{Z}_k , $k \neq j$, given \mathcal{X}_j . Observing elements of \mathcal{Z}_j will cause us to revise our beliefs about \mathcal{X}_j and such changes can then be propagated to other elements of \mathcal{X} using (8) and (9).

3.1. Relevance to survival analysis

The relevance of this to our survival analysis is that we regard the survival model, with second order prior specification for the hazard parameters, as a Bayes linear Bayes graphical model. Information is received, by observation of the deaths or survival of individuals in the time intervals, which changes our beliefs about components of the underlying Bayes linear structure.

Gamerman (1991) simply used equations (6) and (7) repeatedly to compute adjusted moments when making multiple observations. Clearly, if we were to adjust some parameters, B , based on information $D_{A,1}$ and $D_{A,2}$ and if we used (6) and (7) twice, first adjusting by $D_{A,1}$, followed by $D_{A,2}$ and then by $D_{A,2}$ followed by $D_{A,1}$, there is no reason to suppose that the final adjusted expectations and variances would be the same, and indeed in general they are not. In contrast, in equations (8) and (9) the adjustments by $D_{A,1}$ and $D_{A,2}$ are summed and would, therefore, always result in the same adjusted expectations and variances, regardless of the order of updating.

Returning to our simple example, now suppose that we observe the survival times of patients in Group 1 one at a time. If we update the hazard rate for Group 1 sequentially for these patients using Bayes Theorem then the posterior distribution for the rate would not depend on the order in which we observed the survival time of the patients. However, if we propagated these changes one at a time using (6) and (7), then the final adjusted

expectation and variance of the hazard rate in Group 2 would. In contrast, the updates would be commutative if (8) and (9) were used to calculate the adjusted expectation and variance.

Other applications of Bayes linear kinematics are described by Wilson and Farrow (2010) and Gosling et al. (2013).

4. Bayes linear kinematics in survival analysis

4.1. Introduction

In Section 2.4 we saw that the method of Gamerman (1991) combined Bayesian and Bayes linear updates in a dynamic survival model and used the structure of a Bayes linear Bayes model. This avoided the intensive numerical computations of a full Bayesian approach, but was not coherent in the sense that, if the order of the updating using the data was altered, the inferences would change. The same is true of dynamic generalised linear models (West et al., 1985). In this section we outline our approach to this problem which retains the ease of calculation of the Gamerman approach by utilising a combination of full Bayesian and Bayes linear steps while solving the issue of commutativity by using Bayes linear kinematics to provide inferences which remain unchanged under permutations of the order of the observations.

In the following subsections we discuss in more detail the observational model and Bayesian updating, the specification and use of the guide relationship, the covariance structure defined by the system evolution and Bayes linear kinematic updating. In particular, in Section 4.3, we show how the conditions are satisfied for the existence of unique commutative Bayes linear kinematic updates.

4.2. Observations, likelihood and Bayesian updating

In each interval R_j we give each of the $\lambda_{i,j}$ a gamma prior distribution which is conjugate to the conditional density and survival functions given in (2) and (3). The distribution of the hazard $\lambda_{i,j}$ for individual i in interval R_j is $\lambda_{i,j} \sim \text{Ga}(\alpha_{i,j}, \theta_{i,j})$.

The observations on different individuals i, i' are conditionally independent given $\{\lambda_{i,j}\}$ and $\{\lambda_{i',j}\}$. If individual i is alive and uncensored at time τ_{j-1} then this individual can die in interval R_j , can survive interval R_j or can be right-censored during interval R_j . The likelihood contribution from individual i in interval R_j is then

$$L_{i,j} = \lambda_{i,j}^{\delta_{i,j}} \exp\{-\lambda_{i,j}(t_{i,j} - \tau_{j-1})\}, \quad (10)$$

where, if individual i dies in R_j , we have $\delta_{i,j} = 1$ and $t_{i,j} = t_i$, if individual i survives R_j , we have $\delta_{i,j} = 0$ and $t_{i,j} = \tau_j$, if individual i is censored in R_j , we have $\delta_{i,j} = 0$ and $t_{i,j} = t_i$ and if $t_i < \tau_{j-1}$, then $\delta_{i,j} = 0$ and $t_{i,j} = \tau_{j-1}$. Thus non-informative right censoring can be introduced into the model in a straightforward way.

The complete likelihood contribution for individual i is

$$L_i = \prod_{j \in \mathcal{A}_i} L_{i,j},$$

where $\mathcal{A}_i = \{j : \tau_{j-1} < t_i\}$. This is proportional to the likelihood resulting from making observations $D_{i,j}$ for $j = 1, \dots, J(i)$ where $J(i) = \max\{j : \tau_{j-1} < t_i\}$ if, given $\{\eta_{i,j}\}$,

$D_{i,1}, \dots, D_{i,J(i)}$ were conditionally independent and, if $\delta_{i,j} = 1$, $D_{i,j}$ is an observation with value $t_{i,j} - \tau_{j-1}$ from an exponential distribution with mean $\lambda_{i,j}^{-1}$ and, if $\delta_{i,j} = 0$, $D_{i,j}$ is a right-censored observation from the same distribution with censoring time $t_{i,j} - \tau_{j-1}$. A Bayesian update with a fully-specified prior distribution and this equivalent likelihood could be done by conditioning on these exponential observations in any order. Since the likelihood is equivalent, this would also apply in our case, given a fully-specified prior distribution. This is convenient as it allows the posterior to be computed using the temporal structure of the dynamic model. Of course we wish the Bayes linear kinematic updates to have the same property and indeed they do.

The update for $\lambda_{i,j}$ is conjugate and so the posterior distribution for $\lambda_{i,j}$, based on the single observation $D_{i,j}$ at time $t_{i,j}$, is $\lambda_{i,j} \sim \text{Ga}(\alpha_{i,j,1}, \theta_{i,j,1})$, where

$$\alpha_{i,j,1} = \alpha_{i,j,0} + \delta_{i,j}, \quad \theta_{i,j,1} = \theta_{i,j,0} + t_{i,j} - \tau_{j-1}.$$

In terms of the notation of Section 3, the elements of \mathcal{X} are $\eta_{i,j}$, for $i = 1, \dots, p$, and $\beta_{j,k}$, for $k = 0, \dots, q$, both for $j = 1, \dots, J$. The information received, which causes us to change our beliefs about $\eta_{i,j}$ is an observation of $\mathcal{Z}_{i,j} = (\delta_{i,j}, t_{i,j})$.

4.3. Expectation and variance of $\eta_{i,j}$

Goldstein and Shaw (2004) used the mean and variance of a parameter indexing the conditional distribution of an observation X to convey the information when X is observed. For example, if $X | \lambda \sim \text{Po}(\lambda)$, a Poisson distribution with mean λ , then the mean and variance of λ could be used. However we require $\lambda > 0$, we might typically expect a positive association between the means and variances of the various $\lambda_{i,j}$ and an observation with $\delta_{i,j} = 1$ and $t_{i,j} - \tau_{j-1}$ small can lead to an increase in the variance of $\lambda_{i,j}$. These features make Bayes linear propagation of information between the different $\lambda_{i,j}$ less appealing. To overcome them we can introduce a monotonic link function g and use instead the mean and variance of $\eta = g(\lambda)$, which, of course, also indexes the conditional distribution of X since $X | \lambda \sim \text{Po}(g^{-1}(\eta))$. Like West et al. (1985) and Gamerman (1991), we propose $\eta = g(\lambda) = \log(\lambda)$, so that η is unbounded, the idea of the variance of η not depending on the mean seems more reasonable and, as we shall see, the variance of η can not be increased by an observation.

In fact, following (West et al., 1985), we regard the link function as a *guide relationship*, written $\log(\lambda_{i,j}) \approx \eta_{i,j}$, which guides how information is passed between the individual hazards $\lambda_{i,j}$ and the underlying parameters in the model. As part of our prior judgement we specify the conditional first and second moments of $\{\eta_{i,j}\}$, given the possible observations, using the guide relationship in conjunction with the conjugate update of the gamma distributions for $\{\lambda_{i,j}\}$.

The expectation and variance of $\eta_{i,j}$ are then

$$\begin{aligned} \text{E}_0(\eta_{i,j}) &= g_1(\alpha_{i,j}, \theta_{i,j}) = f_{i,j}, \\ \text{Var}_0(\eta_{i,j}) &= g_2(\alpha_{i,j}, \theta_{i,j}) = q_{i,j}, \end{aligned} \tag{11}$$

for some functions $g_1(\cdot)$ and $g_2(\cdot)$, based on our guide relationship and providing a 1 – 1 transformation between $(\alpha_{i,j}, \theta_{i,j})$ and $(f_{i,j}, q_{i,j})$. In the prior specification we set $\alpha_{i,j} = \alpha_{i,j,0}$ and $\theta_{i,j} = \theta_{i,j,0}$, giving $f_{i,j} = f_{i,j,0}$ and $q_{i,j} = q_{i,j,0}$.

This gives a posterior mean $f_{i,j,1}$ and variance $q_{i,j,1}$ for $\eta_{i,j}$, based on this single observation, $D_{i,j}$, using (11) but with the new parameter values.

For Poisson observations, West et al. (1985) suggest two possible choices for g_1 and g_2 . These are (i) use of the mean and variance of $\log \lambda_{i,j}$ and (ii) use of the mode and the curvature at the mode of the log density of $\log \lambda_{i,j}$. A third possibility would be to equate the mean and variance of $\lambda_{i,j}$ to the mean and variance of a lognormal distribution with parameters $f_{i,j}, q_{i,j}$. We will refer to these as the “log-moment”, “log-mode” and “lognormal” methods respectively. In each case we find that

$$\begin{aligned} f_{i,j} = g_1(\alpha_{i,j}, \theta_{i,j}) &= h_1(\alpha_{i,j}) - \log(\theta_{i,j}), \\ \text{and } q_{i,j} = g_2(\alpha_{i,j}, \theta_{i,j}) &= h_2(\alpha_{i,j}) \end{aligned}$$

for some functions h_1, h_2 . Observation of the survival, censoring or death of individual i in interval j gives the posterior expectation and variance

$$f_{i,j,1} = h_1(\alpha_{i,j,0} + \delta_{i,j}) - \log(\theta_{i,j,0} + t_{i,j} - \tau_{j-1}) \quad \text{and} \quad q_{i,j,1} = h_2(\alpha_{i,j,0} + \delta_{i,j}).$$

Provided that $h_2(z)$ is a strictly decreasing function of z for $z > 0$, $h_2(\alpha_{i,j,0} + 1) < h_2(\alpha_{i,j,0})$ and the variance of $\eta_{i,j}$ decreases when $\delta_{i,j} = 1$ and remains the same when $\delta_{i,j} = 0$. This condition is satisfied by all three methods.

The inverse transformation is $\alpha_{i,j} = h_2^{-1}(q_{i,j})$, where h_2^{-1} is the inverse function of h_2 , and $\theta_{i,j} = \exp(-f_{i,j}) \exp\{h_1(\alpha_{i,j})\}$.

For the log-moment method, calculating the mean and variance of $\eta_{i,j}$ as those of $\log \lambda_{i,j}$ gives

$$h_1(\alpha_{i,j}) = \psi(\alpha_{i,j}) \quad \text{and} \quad h_2(\alpha_{i,j}) = \psi_1(\alpha_{i,j}),$$

where $\psi(\cdot)$ is the digamma function and $\psi_1(\cdot)$ is the trigamma function. We can solve $q_{i,j} = \psi_1(\alpha_{i,j})$ numerically for $\alpha_{i,j}$ if required.

For the log-mode method, we set $\mu_{i,j} = \log(\lambda_{i,j})$ but, rather than using the mean and variance of $\mu_{i,j}$ directly, we say that $\eta_{i,j}$ is such that it has mean and variance given by

$$f_{i,j} = m_{i,j}, \quad \text{and} \quad q_{i,j} = - \left\{ \frac{d^2 l_{i,j}(\mu_{i,j})}{d\mu_{i,j}^2} \right\}_{m_{i,j}}^{-1},$$

where $m_{i,j}$ is the mode and $l_{i,j}(\mu_{i,j})$ is the log-density of $\mu_{i,j}$. We find

$$h_1(\alpha_{i,j}) = \log(\alpha_{i,j}) \quad \text{and} \quad h_2(\alpha_{i,j}) = \alpha_{i,j}^{-1}.$$

Furthermore

$$\alpha_{i,j} = \frac{1}{q_{i,j}} \quad \text{and} \quad \theta_{i,j} = \frac{1}{q_{i,j}} e^{-f_{i,j}}.$$

For the lognormal method we set $\alpha_{i,j}/\theta_{i,j} = \exp(f_{i,j} + q_{i,j}/2)$ and $\alpha_{i,j}/\theta_{i,j}^2 = \exp(2f_{i,j} + q_{i,j})[\exp(q_{i,j}) - 1]$ giving

$$h_1(\alpha_{i,j}) = \log \left[\alpha \sqrt{\alpha/(\alpha + 1)} \right] \quad \text{and} \quad h_2(\alpha_{i,j}) = \log(1 + \alpha^{-1}).$$

Furthermore

$$\alpha_{i,j} = [\exp(q_{i,j}) - 1]^{-1} \quad \text{and} \quad \theta_{i,j} = [\exp(q_{i,j}) - 1]^{-1} e^{-q_{i,j}/2} e^{-f_{i,j}}.$$

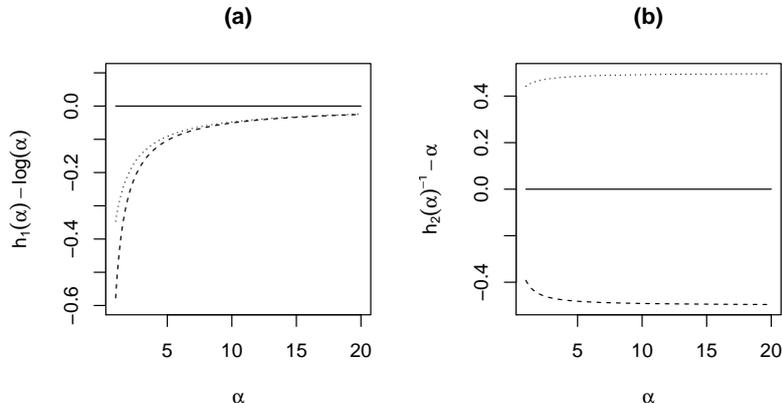


Figure 1: Differences between the functions (a) $h_1(\alpha)$ and (b) $h_2(\alpha)^{-1}$ (ie the precision of η) plotted against $\alpha^{(0)}$ for the log-moment method (dashes), the log-mode method (solid) and the lognormal method (dots). The value for the log-mode method is subtracted as a reference.

While we are free to choose any of these methods and we will use the log-mode method in the example in Section 5, a brief comparison is useful to investigate the sensitivity of the results to this choice. Figure 1 (a) shows the differences between $h_1(\alpha)$ for the three methods with the log-mode method as a reference. We see that the curves for the log-moment and lognormal methods are very similar, lying slightly below the curve for the log-mode method and converging towards it as α increases. Figure 1 (b) shows the differences in the reciprocal of $h_2(\alpha)$, that is the precision of η , again with the log-mode method, where $h_2(\alpha)^{-1} = \alpha$, as a reference. The precision with this method would increase by 1 when a death is observed. The precision using the log-mode method lies between those for the other two methods with the log-moment and lognormal methods giving respectively smaller and larger values. The differences between adjacent curves are slightly less than 0.5.

As a second comparison, let us return to our simple example. We have two hazards, λ_1, λ_2 . For $i = 1, 2$, λ_i has a gamma $\text{Ga}(\alpha_i^{(0)}, \theta_i^{(0)})$ distribution. We give η_1 and η_2 a correlation ρ . We make a single observation with hazard λ_1 and compare the resulting adjusted mean and standard deviation for λ_2 . We make this comparison for (i) an observed death and (ii) a censored observation, both at time $t_1 - \tau = t$ since the beginning of the interval. The adjusted distribution of λ_i is $\text{Ga}(\alpha_i^{(1)}, \theta_i^{(1)})$. The prior mean and variance of λ_i are $m_i^{(0)} = \alpha_i^{(0)}/\theta_i^{(0)}$ and $v_i^{(0)} = \alpha_i^{(0)}/(\theta_i^{(0)})^2$ respectively and the corresponding adjusted values are $m_i^{(1)}$ and $v_i^{(1)}$. For the censored observation, $\alpha_1^{(1)} = \alpha_1^{(0)}$ and $\theta_1^{(1)} = \theta_1^{(0)} + t$ and it is simple to show that, for all three of our Bayes linear kinematic methods, $\alpha_2^{(1)} = \alpha_2^{(0)}$ and $\theta_2^{(1)} = \theta_2^{(0)}k_c$ where $k_c = \{\theta_1^{(1)}/\theta_1^{(0)}\}^\rho$ and therefore $m_2^{(1)} = m_2^{(0)}/k_c$ and $v_2^{(1)} = v_2^{(0)}/k_c^2$. For the observed death, $\alpha_1^{(1)} = \alpha_1^{(0)} + 1$ and the results depend on the choice of transformation.

Figure 2 shows an example where we set $\alpha_1^{(0)} = \alpha_2^{(0)}$ and $\theta_i^{(0)} = \alpha_i^{(0)}$ and the corre-

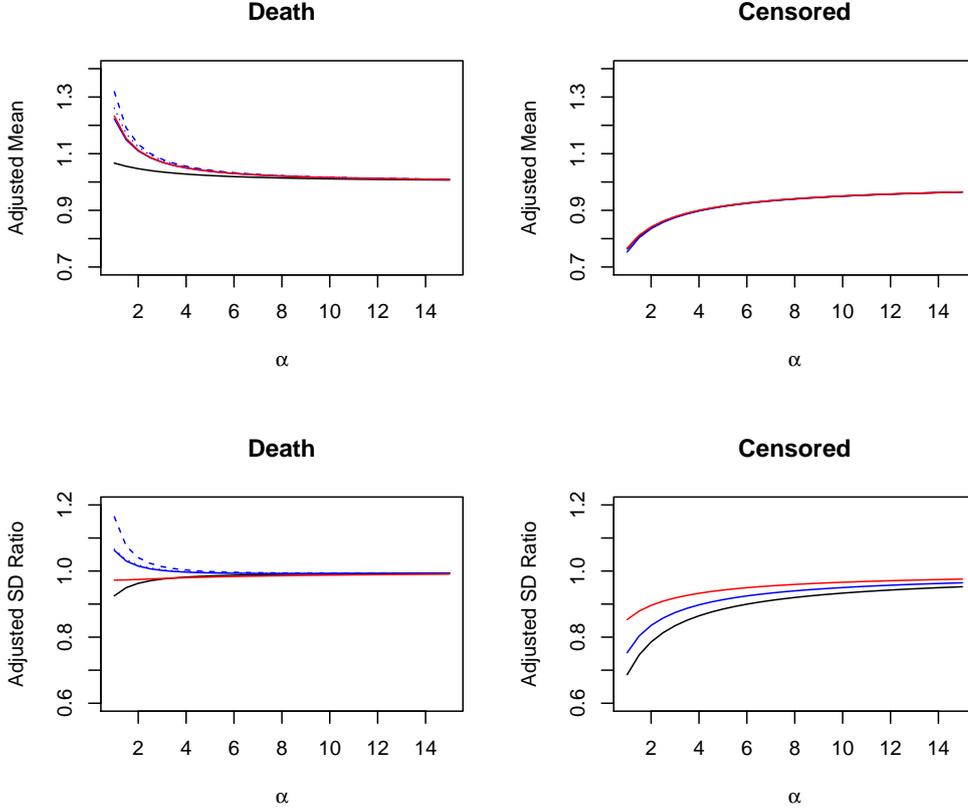


Figure 2: Adjusted means and scaled standard deviations of λ_2 given a single observation with hazard λ_1 . Results for the three Bayes linear kinematic methods with log link are shown in blue with dashes for the log-moment method, solid for the log-mode method and dots for the lognormal method. The black line is for a full-Bayes posterior and the red line is for Bayes linear kinematics with identity link.

lation between $\eta_1^{(0)}$ and $\eta_2^{(0)}$ is 0.7. The prior mean of λ_i is 1 for all $\alpha_i^{(0)}$ but the prior standard deviation depends on $\alpha_i^{(0)}$ so the adjusted standard deviations have been scaled by dividing by the prior standard deviations. For comparison, the posterior mean and standard deviation resulting from a full-Bayes analysis are shown. A bivariate normal prior was used for η_1 and η_2 , giving λ_i a lognormal prior but the parameters were chosen to match the mean and variance of the prior gamma distribution for λ_i in the Bayes linear kinematic case. Also shown is the result of a Bayes linear kinematic update directly based on λ_i rather than *via* η_i , with a correlation of 0.7 between λ_1 and λ_2 , which we term the “identity-link” method. In every case we set $t = 1/\lambda^*$ where λ^* is one prior standard deviation greater than the prior mean of λ_i .

In the case of an observed death, we see that the Bayes linear kinematic methods all give very similar adjusted means which are slightly greater than the full-Bayes posterior

mean for small $\alpha = \alpha_1^{(0)} = \alpha_2^{(0)}$ but quickly converge towards it as α increases. As α increases the shape of a gamma distribution becomes more similar to that of a lognormal distribution. Of these three curves, log-mode is the lowest and log-moment the highest. The adjusted standard deviations give a similar picture except that the identity link method gives results closest to full-Bayes.

In the case of a censored observation, the three Bayes linear kinematics (BLK) with log link methods give identical results. In fact, in the adjusted mean, the other two methods give results which are almost identical to these three. The means for the identity link and log link BLK methods are respectively slightly greater and slightly less than the full-Bayes means. There is more of a difference in the standard deviations with the log link BLK method being closer to full Bayes than the identity link BLK method.

We prefer to use a log link method rather than an identity link because of the guarantee of positive λ , the non-increasing variance and the variance stabilisation. Of the three log link BLK methods, the log-mode method is the simplest, comes closer to full Bayes than the log-moment method and, while its results are very similar to those of the lognormal method, they are very slightly closer to full Bayes. On these grounds, the log-mode method is our preferred method.

4.4. System evolution

We relate the log hazards $\eta_{i,j}$ to model parameters β_j ($j = 1, \dots, r$) using $\eta_{i,j} = x'_{i,j}\beta_j$ where $x'_i = (1, x_{i,1}, \dots, x_{i,q})$, $\beta'_j = (\beta_{j,0}, \dots, \beta_{j,q})$ and $\beta_{j,0} = \log \lambda_{0,j}$.

To build the prior specification for the parameter vector, Gamerman (1991) utilised the system evolution of a dynamic generalised linear model (West et al., 1985). Here we extend this approach and incorporate some ideas from Farrow (2003) to give a more general form.

We write

$$\beta_j - B_j = G_j(\beta_{j-1} - B_{j-1}) + \epsilon_j,$$

where the vector ϵ_j is the cumulative innovation over R_j which has mean zero and covariance matrix E_j . Gamerman (1991) suggests allowing both the system evolution matrix G_j and the innovation covariance matrix E_j to depend on the length b_j of the interval R_j , with $E_j = b_j \bar{E}_j$, where \bar{E}_j is a “unit” covariance matrix. We introduce $B_j = m_j + U$, where m_j is a specified mean vector and the vector U , with mean zero and covariance matrix C_0 , allows us to specify a global component in our uncertainty about β_1, \dots, β_r .

Beliefs about the parameter vector β_j are not given a full distributional form but are simply specified in terms of a mean vector m_j and a covariance matrix C_j . We write $\beta_j \sim [m_j, C_j]$. Thus if our prior beliefs for the parameter vector in interval R_1 , $\beta_1 = (\beta_{1,0}, \dots, \beta_{1,q})$, are given by $\beta_1 \sim [m_1, C_1]$, then we can calculate the prior specification for the parameters in interval R_j as $\beta_j \sim [m_j, C_j]$, where the mean vector is m_j and the covariance matrix is

$$C_j = C_0 + G_j(C_{j-1} - C_0)G'_j + E_j,$$

as β_{j-1} and ϵ_j are independent. The covariance matrix between parameter vectors in different intervals is given by

$$\text{Cov}(\beta_j, \beta_{j+l}) = C_0 + (C_j - C_0) \prod_{m=j+1}^{j+l} G_m = C_{j,j+l}.$$

For example the covariance matrix between β_{j-1} and β_j is $\text{Cov}(\beta_{j-1}, \beta_j) = C_0 + (C_{j-1} - C_0)G_j$.

Using the guide relationship, the prior expectation and variance of $\eta_{i,j}$ are

$$\begin{aligned} f_{i,j} &= \text{E}_0[\eta_{i,j}] &= x_i' m_j, \\ q_{i,j} &= \text{Var}_0(\eta_{i,j}) &= x_i' C_j x_i. \end{aligned}$$

The prior covariances between the transformed quantities $\eta_{i,j}$ are given by

$$\text{Cov}_0(\eta_{i,j}, \eta_{k,l}) = x_i' C_{j,l} x_k = q_{(i,k)(j,l)}$$

where $C_{j,j} = C_j$ and $q_{(i,i)(j,j)} = q_{i,j}$. Finally we need the covariances between the transformed quantities and the parameter values. These are

$$\text{Cov}_0(\eta_{i,j}, \beta_l) = x_i' C_{j,l} = s_{i(j,l)}.$$

We define $\Omega = (\eta', \beta)'$, where $\eta = (\eta_{1,1}, \dots, \eta_{p,r})'$ and $\beta = (\beta_{1,0}, \dots, \beta_{r,q})'$ to be the set of all quantities of interest. Prior specifications for this set are given by $\text{E}_0(\Omega) = L$ and

$$\text{Var}_0(\Omega) = \begin{pmatrix} \text{Var}_0(\eta) & \text{Cov}_0(\eta, \beta) \\ \text{Cov}_0(\beta, \eta) & \text{Var}_0(\beta) \end{pmatrix} = Z,$$

and each of the components of the matrix can be calculated in terms of the quantities found above.

4.5. Bayes linear kinematics

Having made observations, we propagate changes in belief about individuals through to the other individuals, other intervals and model parameters, through the Bayes linear structure, using Bayes linear kinematics. Adjusting our beliefs about Ω having observed individual i in interval R_j gives an adjusted expectation and variance for Ω of

$$\text{E}_{1(i,j)}(\Omega) = L + \text{Cov}_0(\Omega, \eta_{i,j}) \frac{[f_{i,j,1} - f_{i,j,0}]}{q_{i,j,0}},$$

and

$$\text{Var}_{1(i,j)}(\Omega) = Z - \text{Cov}_0(\Omega, \eta_{i,j}) \text{Cov}_0(\eta_{i,j}, \Omega) \left(\frac{1}{q_{i,j,0}} - \frac{q_{i,j,1}}{q_{i,j,0}^2} \right).$$

As discussed in Section 3, we need to define a Bayes linear kinematic update for multiple observations. If we simply performed the updates given in the equations above sequentially, for each patient in each interval, as in Gamerman (1991), then the final values obtained for the adjusted moments of the unknown quantities would change depending on the order in which the updates are performed. The order of patients within an interval should be irrelevant and, in fact, since the analysis is retrospective in contrast to a sequential forecasting algorithm in time-series analysis, the order in which intervals are entered should be irrelevant. When a unique commutative Bayes linear kinematic update exists the adjusted expectation and variance are

$$\text{Var}_{p^*}(\Omega) = \left\{ \sum_{j=1}^r \sum_{i=1}^{p_j} \text{Var}_{1(i,j)}^{-1}(\Omega) - (p^* - 1) \text{Var}_0^{-1}(\Omega) \right\}^{-1}$$

and

$$E_{p^*}(\Omega) = \text{Var}_{p^*}(\Omega) \left\{ \sum_{j=1}^r \sum_{i=1}^{p_j} \text{Var}_{1(i,j)}^{-1}(\Omega) E_{1(i,j)}(\Omega) - (p^* - 1) \text{Var}_0^{-1}(\Omega) E_0(\Omega) \right\},$$

where $p^* = \sum_{j=1}^r p_j$ and p_j is the number of patients known to be alive at time τ_{j-1} . From these we can obtain the adjusted means and variances of the parameters.

With any of the log link BLK methods, $q_{i,j}$, the variance of $\eta_{i,j}$, decreases if individual i dies in interval R_j and stays unchanged by any other observation on individual i in R_j . We can therefore use the commutative formulae (8) and (9). Furthermore, propagation of a decrease in the variance of $\eta_{i,j}$ through the Bayes linear structure can not increase, but can decrease, the variance of other $\eta_{i',j'}$. It follows that, provided that we observe at least one death, however we partition the information received, at least one contribution causes a decrease in at least some of the variances and no contribution can cause an increase in any variance. The update given by (8) and (9) is therefore the unique commutative update. See Goldstein and Shaw (2004) and Wilson and Farrow (2010).

4.6. Prior Robustness

There remain questions around robustness with respect to prior parameters. We now investigate this in the context of our simple example, previously described, in which we are interested in the survival times of two groups of patients, each with a constant hazard rate and a gamma prior distribution. The prior parameters in this example are $(\alpha_1^{(0)}, \theta_1^{(0)}, \alpha_2^{(0)}, \theta_2^{(0)}, \rho)$. To investigate robustness with respect to the prior specifications we perform a simulation exercise in which we simulate observations $t_{1,j} \sim \text{Exp}(\lambda_1)$, $j = 1, \dots, N$ with $\lambda_1 = \alpha/\theta$ for chosen (α, θ) . We assume that the prior parameters are of the form $\alpha_1^{(0)} = \alpha_2^{(0)} = k\alpha$ and $\theta_1^{(0)} = \theta_2^{(0)} = \sqrt{k}\theta$, so that k changes the prior means but not the prior variances of λ_1, λ_2 . We vary k to observe the effect of changes in the prior means of Groups 1 and 2. The observations are made on Group 1 and we consider the effect on the posterior mean of Group 2. We consider different numbers of observed survival times (N) between 2 and 1000. For each of these sample sizes the simulation is run 1000 times.

We choose $\alpha = 12, \theta = 5$ and $\rho = 0.7$. In Figure 3, we see the sample size plotted against the empirical mean over the 1000 runs of the posterior mean of λ_2 . Also given are 95% empirical bounds for the posterior mean from the 1000 runs. We compare the three transformations identified earlier; the log-moment (red), log-mode (green) and lognormal (blue) methods. In addition we show the results from a full-Bayes analysis, using numerical integration, with a lognormal model in black. The plots, from the top, show the cases $k = 0.75, k = 1$ and $k = 1.25$.

Between the three plots we see a pattern. The log-moment, lognormal and full-Bayes methods give results which are almost identical. However the log-mode method results in posterior means which are lower than those of the other methods.

The simulated value for λ_2 is 2.4. When $k = 1$, all methods give posterior means which are close to this value for sample size above 100. Even with smaller samples, however, the posterior mean is on average close to 2.4. The variation in the posterior means decreases as sample size increases. When $k = 1.25$, the prior mean is higher than the hazard rate and all of the methods tend to give posterior means greater than 2.4

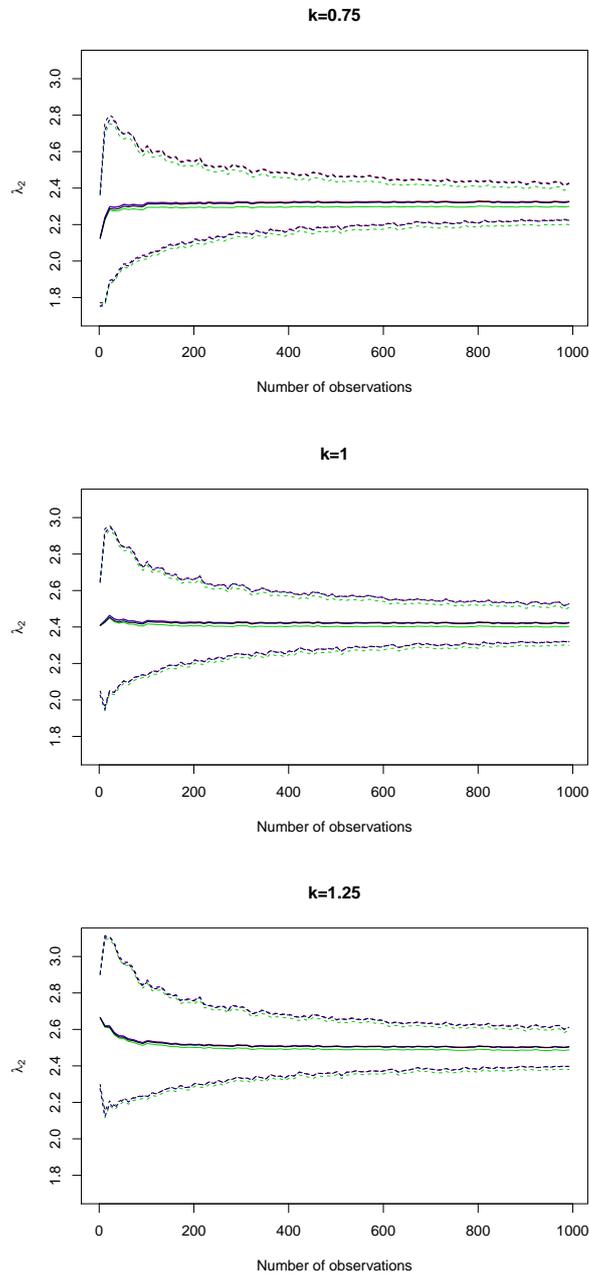


Figure 3: The sample size plotted against the empirical mean over 1000 simulations of the posterior mean of λ_2 . Also given are 95% empirical bounds for the posterior mean from the 1000 runs (dashed lines). We compare three transformations; the log-moment (red), log-mode (green) and lognormal (blue) methods and also a full-Bayes analysis (black).

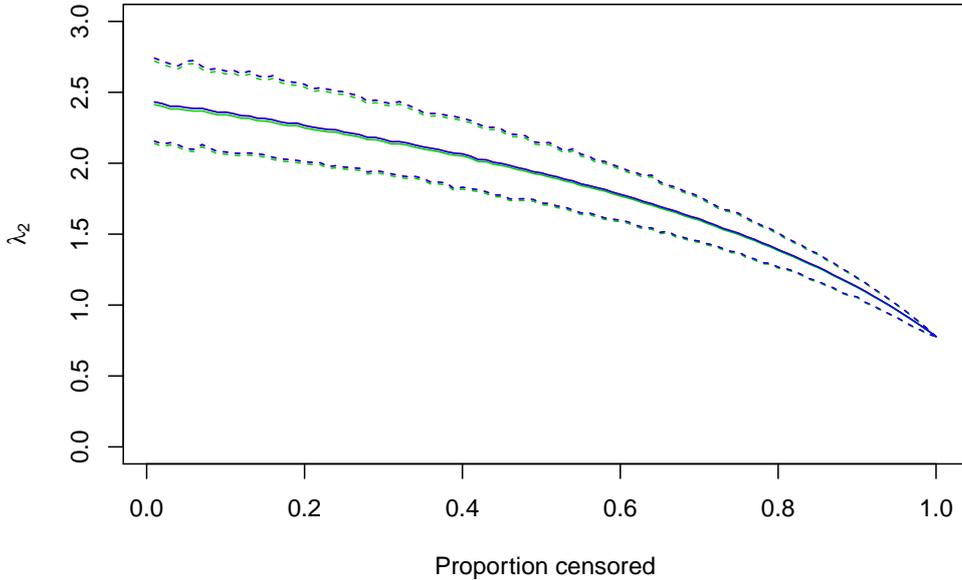


Figure 4: The proportion of right censored observations plotted against the empirical mean over the posterior means of λ_2 from 1000 samples. Also given are the 2.5% and 97.5% quantiles for the posterior mean from the 1000 runs (dashed lines). We compare three transformations; the log-moment (red), log-mode (green) and lognormal (blue) methods.

for all sample sizes but they all give posterior means closer to 2.4 as the sample size increases. The converse is true for $k = 0.75$, with all of the methods tending to give posterior means smaller than 2.4. Note, however, the closeness to the results from the full Bayesian analysis.

We can also investigate the impact of censoring on the posterior mean for this example. In this case, we use the same prior as for Figure 3 but hold the number of observations constant at 100 and vary the proportion of those observations which are right censored. All of the right censoring occurs at $t = 0.2$ which is a little less than half of the predictive mean lifetime of $2.4^{-1} = 0.417$. Each simulated data set was formed by creating the required number, say m , of observations censored at $t = 0.2$ and then simulating $100 - m$ uncensored observations. The fraction of censored observations was therefore fixed at $m/100$. In Figure 4 we compare the three transformation methods and plot the average and 95% symmetric intervals for the posterior means from 1000 samples. On the x-axis is the proportion of right censored values.

The posterior means of λ_2 decrease as the censoring proportion increases because the posterior means of λ_1 decrease. This is because the number of observed deaths decreases and the proportion of cases surviving beyond $t = 0.2$ increases. We see that there are small differences between the log-mode method and the other two when there is very

little censoring in the data but as censoring increases this method converges to the other two. We also see less variation in the posterior means across the runs for all methods as the proportion of censoring increases. This is because the variation in the data decreases.

5. Application: Leukemia survival rates

5.1. Background

Henderson et al. (2002) investigated leukemia survival rates for patients based on data obtained from the North-West Leukemia Register in the UK. In their analysis they considered adult patients with acute myeloid leukemia, the most common type, of which there are 1043 cases in the database between 1982 and 1998. Of these 1043 patients, 879 died and 164 were right censored. The response variable in the study is the time T in days until the death of a patient.

The database also includes the values, for each patient, of a number of covariates which are thought to affect the survival rates of patients with leukemia. As this is an observational study, inferred effects are interpreted as predictive and not necessarily causal. The covariates are

1. The age A_i in years of patient i . We use $x_{i,1} = A_i - 60$ so that the intercept refers to a typical patient.
2. The sex of the patient. That is, $x_{i,2} = -1$ if the patient is female and $x_{i,2} = 1$ if the patient is male.
3. White blood cell count (WBC) at the time of diagnosis. This is truncated at 500 units with 1 unit = $50 \times 10^9/L$. We use $x_{i,3} = WBC - 8$.
4. The Townsend score, a measure of the deprivation of an area of residence on a scale from -7 to 10. The larger this is, the less affluent the area (Townsend et al., 1988). The Townsend score is used directly as the covariate $x_{i,4}$.

The hazard for patient i in interval R_j is then $\exp\{\beta_{j,0} + \sum_{k=1}^4 \beta_{j,k} x_{i,k}\}$. The hazards $\exp\{\beta_{j,0} + \beta_{j,2}\}$ and $\exp\{\beta_{j,0} - \beta_{j,2}\}$ refer respectively to a male patient and a female patient each aged 60, with a white blood cell count of 8 units and a deprivation score of 0.

5.2. Elicitation of prior information

We need to choose the time intervals to be used. In a case like this, with a relatively small proportion of censoring so that most of the recorded times are deaths, the times become much less frequent later. We might suppose *a priori* that the times will have roughly an exponential distribution. Thus it does not seem appropriate to use equal time intervals. Instead, suppose we define $\tau_j = -\nu \log(1 - \kappa j)$, for some values of ν and κ . We choose $\kappa = 0.1$ to give ten intervals and $\nu = 500$, an approximate prior expectation of the marginal mean lifetime, with the aim of having approximately 10% of the events in each interval.

Specification of prior beliefs is required in the form of prior means, variances and covariances for the collection of parameter vectors $(\beta_1, \dots, \beta_r)$ where $\beta_j = (\beta_{j,0}, \beta_{j,1}, \beta_{j,2}, \beta_{j,3}, \beta_{j,4})$. Within our formulation this is done by specifying the moments for β_1 , the system evolution matrices $G_j(b_j)$ and the covariance matrices, E_j , for the innovations ϵ_j .

Table 1: Prior moments for each of the effects.

Effect		Mean	Std.Dev.	Variance
Baseline	$\beta_{1,0}$	-6.000	0.80	0.64
Age	$\beta_{1,1}$	0.020	0.02	0.0004
Sex	$\beta_{1,2}$	0.000	0.35	0.1225
WBC	$\beta_{1,3}$	0.005	0.005	0.000025
Deprivation score	$\beta_{1,4}$	0.000	0.1	0.01

We can obtain suitable prior moments for $\beta_{1,0}$ by assuming a constant hazard and considering a plausible range for the mean lifetime for “baseline” patients. A mean of -6 and a standard deviation of 0.8 for $\beta_{1,0}$ gives a ± 2 standard deviations range corresponding to a range for the mean lifetime from 81 to 1998 days. In order to specify values for the moments of the coefficients of the covariates we utilise the proportional hazards assumption. If hypothetical individuals i and i' have covariate vectors which are equal except that $x_{i,k} \neq x_{i',k}$ and their hazards are $h_i(t)$ and $h_{i'}(t)$ respectively for $\tau_{j-1} \leq t < \tau_j$, then their ratio is

$$r_{i,i'}(t) = \frac{h_i(t)}{h_{i'}(t)} = \exp\{\beta_{j,k}(x_{i,k} - x_{i',k})\}.$$

That is, we can specify values for the hyperparameters by eliciting information about ratios of hazards between individuals. This can be done by considering the situation where both individuals are alive at time τ_{j-1} and we are told that one of them dies during the interval. The conditional probability that it is individual i who dies first is then $L_{C,i,i'}(t) = h_i(t)/\{h_i(t) + h_{i'}(t)\}$, the Cox partial likelihood given just these two individuals (Cox, 1972) and $r_{i,i'}(t) = L_{C,i,i'}(t)/\{1 + L_{C,i,i'}(t)\}$ so quantiles for $L_{C,i,i'}(t)$ are easily converted to quantiles for $r_{i,i'}(t)$ and for $\beta_{j,k}$.

For example, consider $\beta_{1,1}$, the coefficient of the age of the patient. Ages in the investigation range from 14 to 92 and we might expect that increasing age would increase the hazard. We elicit the mean and variance of $\beta_{1,1}$ by supposing that patient i is 10 years older than patient i' . Then the ratio of hazard functions for the two patients is

$$\frac{h_i(t)}{h_{i'}(t)} = \exp\{10\beta_{1,1}\},$$

as long as individuals i and i' have identical covariates otherwise. A 20% decrease in the hazard suggests $\beta_{1,1} \approx 0.005$ and an 80% increase in the hazard suggests $\beta_{1,1} \approx 0.06$. If we assume that these two values give an approximate 95% interval for a normal prior distribution over $\beta_{1,1}$, then we obtain

$$E_0(\beta_{1,1}) = 0.02, \quad \text{Var}_0(\beta_{1,1}) = 0.0004.$$

We can perform a similar process for $\beta_{1,2}$, $\beta_{1,3}$ and $\beta_{1,4}$. Details are omitted. The results of the elicitation process are given in Table 1.

Revie et al. (2010) discuss the problem of assessing prior covariances in Bayes linear structures. In our case, we centred the covariates by subtracting typical values so that

the baseline hazard refers to a central, “typical”, case. Having done this, it is reasonable to have zero prior covariance between the coefficients of covariates in an interval and the baseline log hazard in that interval. Moreover, we saw no reason to suppose that learning that the coefficient of one covariate was larger than our prior expectation of it would cause us to revise our expectation for the coefficient of another covariate so we set the prior covariances between coefficients within a time interval to zero. It still remains to make specifications of the forms of G_j and E_j . With equal-length intervals, a prior in which E_j decreases with j is likely to be appropriate as we may well expect effect sizes to settle towards steady values as time increases. However, with our lengthening intervals, we propose a stationary prior. We set $B_j = B$ with $E(B) = m_0$, $\text{Var}(B) = C_0$, $E_j = E$ and $G_j = \phi I_5$ for all j , where ϕ is a scalar with $0 < \phi < 1$ and I_5 is the (5×5) identity matrix. We achieve stationarity by setting $C_j = C = C_0 + E/(1 - \phi^2)$. Once we have chosen values for C , C_0 and ϕ , this allows us to determine the value of E .

The choices of C_0 , E and ϕ govern the prior covariances between coefficient values in different intervals and hence the degree of smoothing over time. Let Γ_k be the symmetric covariance matrix between β_j and β_{j+k} . Then $\Gamma_k = C_0 + \phi^k(C - C_0)$. In particular, in the special case where C_0 and E are diagonal and $C_0 = c_0 C$ for some scalar $c_0 \in [0, 1]$, then $\Gamma_k = \phi_k C$ where $\phi_k = c_0(1 - \phi^2) + \phi^k$. If we were to learn the value of β_j then our covariance matrix for β_{j+k} would be reduced from C to $C - \Gamma_k C^{-1} \Gamma_k$ or $C(1 - \phi_k^2)$ in the special case. By considering such reductions in variance, suitable values may be elicited. Note that it may be preferred to consider, rather than complete learning of the value of β_j , the effect of gaining information which reduces the variance of β_j from C to $(1 - \delta)C$ for some $\delta \in [0, 1]$. Then, using (7), the covariance matrix of β_{j+k} becomes $C - \delta \Gamma_k C^{-1} \Gamma_k$ or, in the special case, $C(1 - \delta \phi_k^2)$. In this example we choose the special case with $c_0 = 0$ and $\phi = 0.92$ to give moderately strong correlation between neighbouring intervals, falling to quite weak correlation between the first and last. Specifically, this gives $\phi_1^2 = 0.846$, $\phi_5^2 = 0.434$ and $\phi_9^2 = 0.223$.

5.3. Results

For the reasons given in Section 4.3, we shall use the log-mode method here for all calculations.

Having updated the four parameters of interest using the data we can plot the effects of the covariates over time. The posterior means for the effects of age and sex are shown by circles in Figure 5. The posterior parameter means for interval j are plotted at $\tau_j^* = -\nu \log\{1 - \kappa(j - 0.5)\}$ but would remain constant within that interval. The horizontal scale is proportional to $1 - \exp(-t/\nu)$. Also plotted are ± 2 standard deviation intervals.

The effect of age is marked. There is also a strong suggestion of a temporal component to the effect of age. The posterior means for the coefficients of age are positive in all intervals, as are most of those for sex, indicating that increased age and being male both increase the hazard associated with death from acute myeloid leukemia.

The age effect is strongest initially and appears to decrease over time then finally increase again. Only one of the ± 2 standard deviation intervals includes zero, however, so even in intervals where the effect of age is not as strong, there is still an effect. In contrast, the effect associated with sex appears to increase over time. The effect of this covariate is less clear, however, as all but two of the ± 2 standard deviation intervals contain zero.

Table 2: Posterior means and standard deviations for each of the parameters in each interval.

j	τ_j	$\beta_{j,1} \times 10^2$	$\beta_{j,2} \times 10$	$\beta_{j,3} \times 10^3$	$\beta_{j,4} \times 10^2$
1	52.6	3.647 (0.401)	-0.047 (0.554)	6.453 (0.584)	3.492 (1.557)
2	111.6	3.743 (0.454)	0.694 (0.678)	3.541 (0.975)	5.472 (1.891)
3	178.3	2.689 (0.468)	0.370 (0.833)	3.040 (1.136)	2.172 (2.493)
4	255.4	2.424 (0.480)	1.170 (0.861)	4.086 (1.104)	3.940 (2.339)
5	346.6	2.126 (0.505)	0.749 (0.896)	3.272 (1.421)	2.300 (2.505)
6	458.1	1.107 (0.480)	0.545 (0.877)	1.803 (1.598)	2.080 (2.431)
7	602.0	1.230 (0.557)	-0.103 (0.923)	3.783 (1.468)	1.729 (2.480)
8	804.7	0.719 (0.585)	1.555 (0.998)	2.090 (1.580)	-3.246 (2.921)
9	1151.3	1.821 (0.674)	2.748 (1.113)	-4.029 (2.099)	-8.808 (3.388)
10	∞	4.640 (0.711)	4.428 (1.200)	-7.147 (2.426)	-10.519 (3.685)

The posterior means for the effects of all covariates in all intervals are given in Table 2. Posterior standard deviations are given in brackets.

We see that the covariates, on the whole, do indeed appear to have an effect on the survival times of patients with leukemia. Age has the most pronounced overall effect, with all of the posterior means of the coefficients being positive. Deprivation score and white blood cell count have positive posterior means for all but the later time periods, indicating that they too could be associated with the hazard.

The negative posterior means in the later intervals could be an indication that the effects of these covariates decrease over time. This could be related to the phenomenon of dynamic selection of the population. For example, if, in the later intervals, few older patients are still observed then this could be the reason for a smaller effect of age at later time-points. In our case the mean age of patients in each interval are (60.7, 57.1, 54.5, 53.5, 52.2, 51.1, 50.6, 49.0, 48.5, 48.3) and so dynamic selection may explain the smaller effects in later intervals.

5.4. Comparison with full Bayesian analysis

We can compare the results of our Bayes linear Bayes analysis with a full Bayesian approach. The full Bayesian model we use for the comparison has piecewise constant hazards over the same intervals as the Bayes linear Bayes model. The prior first and second moments of the coefficients are the same as for the Bayes linear Bayes model but now we give the parameters a multivariate normal prior distribution. We computed the posterior distribution using a Gibbs sampler with two parallel chains. The first 2000 iterations were discarded and, after this, convergence appeared satisfactory although mixing was not particularly good. The posterior distribution was calculated using a further 10000 iterations. This took approximately twenty minutes to run on a desktop computer in comparison with about two seconds for the Bayes linear Bayes model. The posterior means and 95% symmetric posterior intervals for $\beta_{j,1}$ and $\beta_{j,2}$ are given in Figure 5 alongside the means and ± 2 standard deviations intervals previously given for the Bayes linear Bayes method.

We see that, in most cases, the results using the two methods are similar, but not always. In particular, in the last time period, the posterior means using the full Bayes

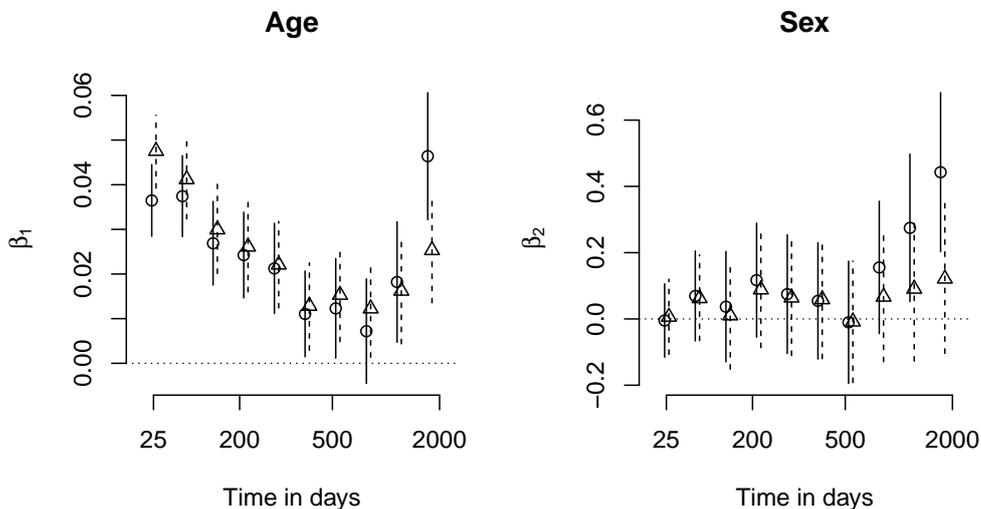


Figure 5: The effects of age and sex on the hazard functions of patients with leukemia. The posterior means for the Bayes linear Bayes method are given by circles and those for the full Bayesian method with triangles. Both are plotted with posterior intervals.

method are noticeably less than those obtained using the Bayes linear Bayes analysis. Considering Figure 2, we can conclude that the largest discrepancies between the full-Bayes and Bayes linear Bayes results are likely to occur when α is small and the patient dies in the interval in question. In fact, in our example, α is always small. In the case of the log mode method, we have $\alpha_{i,j} = q_{i,j}^{-1}$ where $q_{i,j}$ is the variance of $\eta_{i,j}$. From Table 1 we can see that even a baseline patient would have $q_{i,j} = 0.64 + 0.1225 = 0.7625$ giving $\alpha_{i,j} = 1.31$. Other patients will have larger values of $q_{i,j}$ and therefore smaller $\alpha_{i,j}$. A prior variance of 0.8 for a log hazard is really quite large since a ± 2 standard deviation interval for $\eta_{i,j}$ would correspond to an interval for the hazard $\lambda_{i,j}$ with a ratio of 35.8 between the upper and lower limits. Thus, where possible, the use of less diffuse and realistically informative priors is likely to lead to results which are closer to those of the corresponding full Bayes analysis.

Table 3 shows the differences between the full Bayes posterior means and the Bayes linear Bayes posterior means, standardised by dividing by the full Bayes posterior standard deviation. Positive differences imply that the Bayesian method gives larger values for the posterior mean. For comparison, the differences between the full Bayes posterior means and the prior means divided by the prior standard deviations are shown in brackets.

The differences in the case of $\beta_{j,1}$ and $\beta_{j,2}$ can be seen here. There are also relatively large discrepancies in the case of $\beta_{j,3}$, the coefficient of white blood cell count, and in the log baseline hazard $\beta_{j,0}$ (not shown here), although the general patterns of the results given by the two methods are the same.

Table 3: Standardised differences between the posterior means, full Bayes minus Bayes linear Bayes, for each of the parameters in each interval. Standardised differences between the full Bayes posterior means and prior means are shown in brackets.

Interval j	$\beta_{j,1}$		$\beta_{j,2}$		$\beta_{j,3}$		$\beta_{j,4}$	
1	2.63	(1.38)	0.19	(0.02)	-3.72	(-0.14)	1.52	(0.58)
2	0.84	(1.06)	-0.11	(0.18)	-1.60	(-0.58)	0.19	(0.58)
3	0.61	(0.50)	-0.33	(0.03)	-0.68	(-0.54)	-0.04	(0.21)
4	0.36	(0.30)	-0.32	(0.25)	-0.57	(-0.32)	-0.55	(0.27)
5	0.16	(0.10)	-0.12	(0.18)	-0.41	(-0.46)	-0.51	(0.11)
6	0.35	(-0.36)	0.05	(0.17)	-0.22	(-0.70)	-0.35	(0.12)
7	0.57	(-0.23)	0.02	(-0.02)	-0.58	(-0.42)	-0.16	(0.13)
8	0.92	(-0.39)	-0.88	(0.19)	0.06	(-0.56)	0.79	(-0.10)
9	-0.32	(-0.19)	-1.65	(0.26)	2.00	(-0.96)	1.76	(-0.32)
10	-3.45	(0.27)	-2.78	(0.35)	2.69	(-1.20)	1.78	(-0.47)

This was a severe test of the Bayes linear kinematic approach since each observation on an individual in an interval typically contributed little information and thus had relatively little effect on the gamma prior distribution for the corresponding hazard. These gamma prior distributions had small shape parameters α and Figure 2 shows that this leads to the greatest difference from the full Bayes analysis. An observed death would increase α by 1 while a censoring or the observation that the individual survived the interval would not increase α at all. In the Bayes linear kinematic approach, each of these α values corresponding to observations would only be changed, if at all, by its own observation and subsequent observations also act on prior gamma distributions with small α .

6. Simulation

As a further examination of the behaviour of the Bayes linear Bayes method, we carried out some simulation experiments. We generated simulated data sets and then, in each case, computed posterior means of the parameters using the model and prior specification as used in the leukaemia example. To generate the data sets we used the same number of patients and the same covariate values as in the leukaemia example but we randomly generated survival times and censoring. First we generated a random survival time t for each patient. Then a censoring time t_{cens} for the patient was drawn from an exponential distribution with rate 0.0001. If $t > t_{\text{cens}}$ then that patient was labelled as censored at time t_{cens} .

Figure 6 shows an example of the results. This is based on 2000 simulated data sets. The survival times were drawn from a Weibull distribution with shape parameter 0.8. Thus the baseline hazard decreases over time. The other parameters were set at $\beta_0 = -5.5$, $\beta_1 = 0.025$, $\beta_2 = 0.1$, $\beta_3 = 0.004$, $\beta_4 = 0.04$. Note that these coefficients did not change over time but the baseline hazard did because of the Weibull shape parameter. Thus, in terms of the piecewise constant hazard model, β_0 changes over time and the other coefficients do not. Figure 6 shows results for β_0 and β_1 . For each time

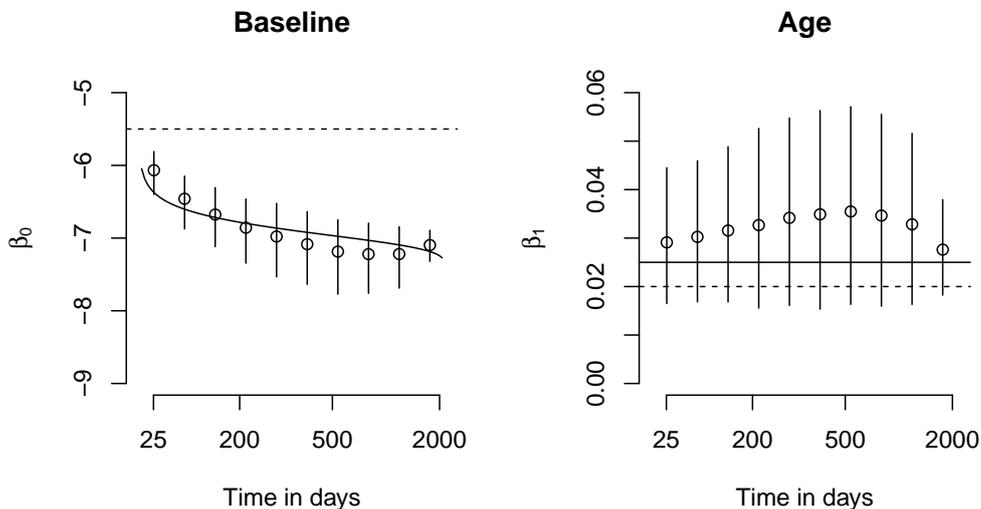


Figure 6: Simulation results where the true distribution is Weibull. Empirical means and 95% intervals of Bayes linear kinematic posterior means for 2000 simulated samples. True values are shown by the solid line and prior means by the dashed line.

interval, the 2000 posterior means are summarised by their empirical mean and empirical 95% interval. The true parameter value is shown by a solid line and the prior mean is shown by a dashed line. It can be seen that, in every case, the true value is inside the 95% interval. Of particular interest is the way that the posterior means of the baseline hazard successfully track the true value even though the Bayes linear Bayes model used for inference contains no information about the form which this would take.

7. Discussion and Conclusions

In this paper we have investigated the application of a Bayes linear Bayes model and Bayes linear kinematics to survival analysis using a piecewise constant hazards model with a temporally dependent prior. The approach taken involved fully Bayesian conjugate updates for individual hazards within an interval. These changes in belief were propagated through to hazards for other individuals and intervals using Bayes linear kinematics.

The approach is similar to that of Gamerman (1991). However, by using Bayes linear kinematics as opposed to standard Bayes linear updating, we are able to produce a solution which is commutative unlike the analysis of Gamerman (1991). Unlike other published work on Bayes linear Bayes methods, we used a nonlinear link function between the parameters of the observational distributions, in this case the hazards, and the quantities in the Bayes linear structure.

We applied our approach to an example involving patients with acute myeloid leukemia in the North-West of England. There were 1043 patients in the analysis and, due to the

analytic nature of all of the updating, posterior means and variances were calculated efficiently. No intensive numerical methods were necessary even though our method allows covariate effects to change over time in a flexible way. We compared the results from our Bayes linear Bayes approach to those from a typical full Bayesian approach. Posteriors in this case were found using Markov chain Monte Carlo methods. The Bayes linear Bayes approach produced solutions far more quickly. Further research on producing Bayes linear Bayes analyses which more closely match the corresponding full Bayes results is planned.

We have examined a Bayes linear kinematic approach to survival analysis for a piecewise constant hazards model. This does, of course, require a choice of a set of change points. A possible future direction would be to apply our method without arbitrary change-points, in an analysis analogous to a semi-parametric Cox proportional hazards regression. This would be comparable to the reversible jump MCMC work of Kim et al. (2007). Frailties (eg. Henderson and Oman, 1999) can produce an apparent decrease over time in covariate effects but the piecewise constant hazards model is more flexible in the form of changes over time. If required, frailties can be introduced in a straightforward way in our approach by adding a random patient effect Z_i in $\eta_{i,j}$.

Acknowledgement

Our thanks go to Professor Rob Henderson for kindly giving us access to the data for the application. We are grateful to referees for helpful comments on earlier drafts of this paper.

References

- D. R. Cox. Regression models and life tables. *Journal of the Royal Statistical Society. Series B*, 34: 187–220, 1972.
- M. Farrow. Building of subjective covariance structures for large complicated systems. *The Statistician*, 52:553–573, 2003.
- D. Gamerman. Dynamic Bayesian models for survival data. *Applied Statistics*, 40:63–79, 1991.
- M. Goldstein. Subjective bayesian analysis: Principles and practice. *Bayesian Analysis*, 1:403–420, 2006.
- M. Goldstein and S. Shaw. Bayes linear kinematics and Bayes linear Bayes graphical models. *Biometrika*, 91:425–446, 2004.
- M. Goldstein and D. Wooff. *Bayes Linear Statistics*. Wiley, 2007.
- J. P. Gosling, A. Hart, H. Owen, M. Davies, J. Li, and C. MacKay. A Bayes linear approach to weight-of-evidence risk assessment for skin allergy. *Bayesian Analysis*, 8:169–186, 2013.
- R. Henderson and P. Oman. Effect of frailty on marginal regression estimates in survival analysis. *Journal of the Royal Statistical Society. Series B*, 61:367–379, 1999.
- R. Henderson, S. Shimakura, and D. Gorst. Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association*, 97:965–972, 2002.
- J. G. Ibrahim, M. H. Chen, and D. Sinha. *Bayesian Survival Analysis*. Springer, 2001.
- R. C. Jeffrey. *The logic of decisions*. McGraw-Hill, 1965.
- M. Jones, M. Goldstein, P. Jonathan, and D. Randell. Bayes linear analysis for Bayesian optimal experimental design. *Journal of Statistical Planning and Inference*, 171:115–129, 2016.
- J. D. Kalbfleisch. Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society. Series B*, 40:214–221, 1978.
- S. Kim, M. H. Chen, K. Dey, and D. Gamerman. Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Analysis*, 13:17–35, 2007.
- J. P. Klein and M. L. Moeschberger. *Survival Techniques for Censored and Truncated Data*. Springer, 1997.
- D. Lindley. *Understanding Uncertainty*. Wiley, 2006.
- T. Martinussen and T. H. Scheike. *Dynamic Regression Models for Survival Data*. Springer, 2006.

- I. W. McKeague and M. Tighiouart. Bayesian estimators for conditional hazard functions. *Biometrics*, 56:1007–1015, 2000.
- P. Müller. Optimal sample size for multiple testing: the case of gene expression microarrays. *Journal of the American Statistical Association*, 99:990–1001, 2004.
- M. Revie, T. Bedford, and L. Walls. Evaluation of elicitation methods to quantify Bayes linear models. *Journal of Risk and Reliability*, 224:322–332, 2010.
- D.J. Sargent. A flexible approach to time-varying coefficients in the Cox regression setting. *Lifetime Data Analysis*, 3:13–25, 1997.
- D. Sinha and D.K. Dey. Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, 92:1195–1212, 1997.
- D. Sinha, M-H. Chen, and S.K. Ghosh. Bayesian analysis and model selection for interval-censored survival data. *Biometrics*, 55:585–590, 1999.
- P. Townsend, P. Phillimore, and A. Beattie. *Health and Deprivation: Inequality and the North*. Croom Helm, 1988.
- M. West, P. J. Harrison, and H. S. Migon. Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, 80:73–83, 1985.
- K. J. Wilson and M. Farrow. Bayes linear kinematics in the analysis of failure rates and failure time distributions. *Journal of Risk and Reliability*, 224:309–321, 2010.

Appendix

Suppose we observe the values of unknowns within a Bayes linear structure. Suppose that, as above, we have $\mathcal{X} = \bigcup_{j=1}^J \mathcal{X}_j$. Let $\text{Var}_0(X_j) = W_j$. In addition, we have other vector random quantities Y_1, \dots, Y_J such that the dimension of Y_j is n_j and the elements of Y_j are conditionally uncorrelated with those of $\mathcal{X} \setminus \mathcal{X}_j$ given X_j . Moreover we can represent the relationship between Y_j and X_j by writing

$$Y_j = m_{Y,j} + M_j(X_j - m_{X,j}) + U_j$$

where $m_{X,j}$ and $m_{Y,j}$ are specified prior mean vectors for X_j and Y_j , M_j is a specified matrix and U_j is a zero-mean random vector with $\text{Var}(U_j) = V_j$. Hence $E_0(Y_j) = m_{y,j}$, $\text{Var}_0(Y_j) = V_j + M_j W_j M_j'$ and $\text{Cov}_0(Y_j, X_j) = M_j W_j$.

Let the dimension of X be n_X , let $n_Y = \sum_{j=1}^J n_j$ and let M be a $n_Y \times n_X$ matrix consisting of zeroes except that, in rows $1 + \sum_{j=1}^{k-1} n_j$ to $\sum_{j=1}^k n_j$ we have the elements of the rows of M_j placed in the columns corresponding to the elements of \mathcal{X}_j . Let $Y = (Y_1', \dots, Y_J')'$, $m_Y = (m_{Y,1}', \dots, m_{Y,J}')'$, $m_X = E_0(X)$, $U = (U_1', \dots, U_J')'$ and let $\text{Var}_0(U) = V_U$, which is block-diagonal, and $\text{Var}_0(X) = V_X$. Then $\text{Var}_0(Y) = M V_X M' + V_U$ and $\text{Cov}_0(X, Y) = V_X M'$ so, given an observation y of the whole of Y , our adjusted expectation and variance are

$$\begin{aligned} E(X | Y) &= m_X + V_X M' (M V_X M' + V_U)^{-1} (y - m_Y), \\ \text{Var}(X | Y) &= V_X - V_X M' (M V_X M' + V_U)^{-1} M V_X. \end{aligned}$$

It is easily verified that the precision is

$$P(X | Y) = \text{Var}^{-1}(X | Y) = P_X + M' P_U M$$

where $P_X = V_X^{-1}$ and $P_U = V_U^{-1}$ (where the necessary inverses exist). Furthermore, because of the block-diagonal structure of V_U and P_U ,

$$P(X | Y) = P_X + \sum_{j=1}^J \tilde{M}_j' P_j \tilde{M}_j, \quad (12)$$

where $P_j = V_j^{-1}$ and \tilde{M}_j is a $n_j \times n_X$ matrix composed of rows $1 + \sum_{j=1}^{k-1} n_j$ to $\sum_{j=1}^k n_j$ of M .

Similarly

$$E(X | Y) = P^{-1}(X | Y)\{P_X m_X + M' P_U (y - m_Y)\}$$

and, because of the structure of M and P_U ,

$$E(X | Y) = P^{-1}(X | Y)\{P_X m_X + \sum_{j=1}^J \tilde{M}_j' P_j (y_j - m_{Y,j})\}. \quad (13)$$

Now consider what happens if we only observe $Y_j = y_j$. The corresponding results are

$$P(X | Y_j) = P_X + \tilde{M}_j' P_j \tilde{M}_j \quad (14)$$

and

$$E(X | Y_j) = P^{-1}(X | Y_j)\{P_X m_X + \tilde{M}_j' P_j (y_j - m_{Y,j})\}. \quad (15)$$

From (12) and (14) we see that

$$P(X | Y) = \sum_{j=1}^J P(X | Y_j) - (J - 1)P_X. \quad (16)$$

From (13) and (15) we see that

$$\begin{aligned} P(X | Y)E(X | Y) &= P_X m_X + \sum_{j=1}^J \tilde{M}_j' P_j (y_j - m_{Y,j}) \\ &= \sum_{j=1}^J \{P_X m_X + \tilde{M}_j' P_j (y_j - m_{Y,j})\} - (J - 1)P_X m_X \\ &= \sum_{j=1}^J P(X | Y_j)P^{-1}(X | Y_j)\{P_X m_X + \tilde{M}_j' P_j (y_j - m_{Y,j})\} \\ &\quad - (J - 1)P_X m_X \\ &= \sum_{j=1}^J P(X | Y_j)E(X | Y_j) - (J - 1)P_X m_X. \end{aligned} \quad (17)$$

Now we return to the more general situation of Bayes linear kinematics and suppose that, instead of observing Y_1, \dots, Y_J , which would cause us to adjust our beliefs about X , we gain other pieces of information, D_1, \dots, D_J , where receiving D_j causes us to revise our beliefs about X_j directly but only affects our beliefs about $\mathcal{X} \setminus \mathcal{X}_j$ through this effect on X_j .

Goldstein and Shaw (2004) derived the conditions under which a commutative Bayes linear kinematic update exists and under which this update is unique. When a unique commutative update exists, it is given by replacing Y_j and Y with D_j and D in (16) and (17), giving (8) and (9).