

Bayes Linear Bayes Networks with an Application to Prognostic Indices

Wael A. J. Al-Taie[†] and Malcolm Farrow^{*}

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. The use of Bayes linear kinematics eliminates the problem of non-commutativity which was observed in earlier work involving moment-based belief updates. In this paper we describe this approach and investigate its application to the rapid computation of prognostic index values in survival when a patient’s values may only be available for a subset of covariates. We consider the use of covariates of various kinds and introduce the use of non-conjugate marginal updates. We apply the technique to an example concerning patients with non-Hodgkin’s lymphoma, in which we treat the linear predictor of the lifetime distribution as a latent variable and use its expectation, given whatever covariates are available, as a prognostic index.

Keywords: Bayes linear kinematics, Bayes linear Bayes graphical model, survival analysis, prognostic index, missing covariates.

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.

In this paper, we investigate the application of Bayes linear Bayes models and Bayes linear kinematics to the calculation of prognostic indices in medicine, particularly in survival or time-to-event problems. Many prognostic indices used in practice involve only a small number of covariates but other information may be available. We wish to be able to construct a prognostic index which can use information from a larger number of variables. An obstacle to doing this with conventional methods is that the extra

^{*}School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK, malcolm.farrow@newcastle.ac.uk

[†]Northern Technical University, Mosul, Iraq, wael9460@yahoo.co.uk

variables might not always be observed. We wish to overcome this difficulty by being able to compute an index value even when some variables are not observed. This can be done in the case where all variables are jointly multivariate normal. It can also be done, in a Bayesian network, when all variables are categorical with finite sets of states (eg Lauritzen and Spiegelhalter, 1988; Jensen and Nielsen, 2007). Other cases are more computationally demanding and typically require the use of methods such as Markov chain Monte Carlo (MCMC) algorithms. We develop an approach which allows for a large number of potential covariates but which allows the calculation of a prognostic index value when only a subset of the covariates are observed. The Bayes linear kinematic approach makes this possible in a straightforward way, with deterministic calculations, and, indeed, allows the index to be adjusted in a straightforward and commutative way if new covariate information is received. This opens the possibility of a fast and easy-to-use Web-based prognostic index calculator which might be used in different parts of the world where different subsets of covariates may be available. Covariates used in prognostic indices can take many forms so we investigate a new approach to dealing with non-conjugate relationships in the marginal full-Bayes parts of the model. This extends the applicability of the Bayes linear Bayes approach.

In Section 2 we discuss prognostic indices and the case where some covariates might not be observed. In Section 3 we briefly outline Bayes linear methods, Bayes linear kinematics and Bayes linear Bayes graphical models. In Section 4 we extend previous work on Bayes linear Bayes graphical models by developing the case where the marginal full-Bayesian sub-models are non-conjugate. In Section 5 we discuss the application of these methods to the routine computation of prognostic index values for patients with terminal illness where only partial information may be available. We discuss the stages of constructing the Bayes linear Bayes network which is used to compute the index values. In Section 6 we illustrate this with an example involving patients with Non-Hodgkin's lymphoma.

2 Prognostic Indices

2.1 Prognostic Indices in Survival

Prognostic indices are used to help to predict outcomes in patients, for example survival times in patients with a terminal illness. The value taken by the prognostic index depends on the clinical information about patients. A prognostic index can be useful to make a decision about the appropriate treatment for the patients. For example, Henderson et al. (2001) described the significance of using prognostic indices in terms of the impact on the selection of various treatments and the importance for the patients and their families to know and think about the future to give the best support to the patients in the remaining years of their lives.

Typically, a prognostic index is a linear predictor based on the explanatory variables. In a proportional hazards model, the hazard function for patient i is $\lambda_i(t) = \lambda_0(t)e^{\eta_i}$, where $\lambda_0(t)e^{\beta_0}$ is the baseline hazard, and the prognostic index of patient i is $\eta_i =$

$\log[\lambda_i(t)/\lambda_0(t)]$. A greater value corresponds to worse prognosis. Typically we write

$$\eta_i = \beta_0 + \sum_{j=1}^J \beta_j x_{j,i} \tag{1}$$

where $(\beta_0, \beta_1, \dots, \beta_J)$ are the regression coefficients and $x_{1,i}, \dots, x_{J,i}$, are the covariate values for the patient of interest. Thus the baseline hazard applies to a patient for whom all of the covariate values are zero. Usually we would determine values for the coefficients by fitting a suitable survival model to historical data. For example, if T_i is the survival time of patient i in the historical data then we might use a Weibull model with $T_i \sim \text{Weibull}(\alpha, \lambda_i)$ where, α is the shape parameter in the model and $\lambda_i = \exp\{\eta_i\}$ is the scale parameter. So, $\eta_i = \log(\lambda_i)$ is the prognostic index for patient i .

Note that the predictive mean for the linear predictor for a new patient with covariate values $x_{1,i}, \dots, x_{J,i}$ is obtained by substituting the posterior means of β_0, \dots, β_J into (1), provided, of course, that all of the covariates are observed. This predictive mean is then given as the prognostic index value. Our method provides a means of dealing with the case where not all of the covariates are observed.

Once the prognostic index value for a patient has been computed, it might be transformed to help with interpretation. For example, we might use the distribution of prognostic index values in the historical data and determine the percentile of this distribution corresponding to a new index value. Thus we would give a score on a 0-100 scale.

2.2 Prognostic Networks and Unobserved Covariates

If we have a fixed list of covariates $\mathcal{X} = \{X_1, \dots, X_J\}$ then the index must be given by a function of the values x_1, \dots, x_J taken by these covariates. Conventionally, to construct a prognostic index, we try to choose a suitable function $g(x_1, \dots, x_J)$. Suppose now that we have a list $\mathcal{X}_{\max} = \{X_1, \dots, X_J\}$ of covariates which can be observed but that we might not always observe all of X_1, \dots, X_J but rather we observe some subset of \mathcal{X}_{\max} . Let the possible observed subsets be $\mathcal{X}_1, \dots, \mathcal{X}_M$. We need a different function g_m for each possible subset \mathcal{X}_m . To do this in a coherent and principled way, we introduce the idea of a latent variable Z_T . When we supply an index value to a user, we will give our current expectation of Z_T , given the information available to us. We can choose Z_T to be a quantity on which the lifetime distribution depends, as in a traditional prognostic index. For example, in a Weibull model with survival function $\exp\{-\lambda_i t^\alpha\}$ for subject i , we can use $Z_T = \eta_i = \log(\lambda_i)$ as the prognostic index. If $\eta_i = \beta_0 + \sum_{j=1}^J \beta_j X_{j,i}$, where $X_{j,i}$ is the value of covariate j for subject i , but not all of $X_{1,i}, \dots, X_{J,i}$ are observed for subject i , then we use $G_i = E(\eta_i | \mathcal{X}_i)$ where \mathcal{X}_i is the subset of observations made for patient i . We refer to G_i as the *predicted* prognostic index value. Alternatively we can regard Z_T as the result of a transformation of the lifetime, T , itself, such that $E(Z_{T,i} | \eta_i) = \eta_i$ where $Z_{T,i}$ is the value of Z_T for patient i . Again, in this case, we compute the current expectation of η_i as our predicted prognostic index value.

This allows us to compute a (predicted) prognostic index value given observations of any subset of the possible covariates, for example when some values are missing or when some variables are only measured in certain cases.

Additional flexibility is provided by modelling the joint distribution of Z_T and the covariates, often through latent variables Z_1, \dots, Z_J associated with the covariates, so that Z_T is not known precisely even when all of the covariates are observed. In this way we always use an expectation of Z_T as our declared index value.

Bayes linear methods applied to $\{Z_1, \dots, Z_J, Z_T\}$ require the specification of only the first and second moments of this collection and, when given the values of some of these variables, it is straightforward to compute the adjusted expectation of Z_T . We outline Bayes linear methods in Section 3.1. In many cases, the observation of a covariate X_j does not determine the value of the corresponding Z_j but only changes its moments. The use of Bayes linear kinematics and a Bayes linear Bayes model allows us to deal with this situation in a way which is quick and efficient and, importantly, commutative so that covariate information may be introduced in any order without affecting the resulting index value. We outline Bayes linear kinematics and Bayes linear Bayes models in Sections 3.3 and 3.4.

3 Bayes Linear Kinematics and Bayes Linear Bayes Graphical Models

3.1 Bayes Linear Methods

In the standard Bayesian paradigm, we specify the full joint prior distribution for all unknown quantities. Using Bayes' theorem, we update our prior beliefs by conditioning on the observations and then calculating the posterior distributions. A Bayes linear analysis (Goldstein and Wooff, 2007) is distinct from the full Bayesian approach in that we only need to specify the first and second-order moments for the prior and then calculate posterior moments. For instance, if we have a vector random quantity \underline{Z} , then we specify the prior expectation and variance of \underline{Z} respectively as $E_0(\underline{Z})$ and $\text{Var}_0(\underline{Z}) = E_0[\{\underline{Z} - E_0(\underline{Z})\}\{\underline{Z} - E_0(\underline{Z})\}']$. Furthermore, for two quantities \underline{Z}_1 and \underline{Z}_2 , we also specify a prior covariance $\text{Cov}_0(\underline{Z}_1, \underline{Z}_2) = E_0[\{\underline{Z}_1 - E_0(\underline{Z}_1)\}\{\underline{Z}_2 - E_0(\underline{Z}_2)\}']$.

Suppose that we have two vectors $\underline{A} = (A_1, \dots, A_p)'$ and $\underline{B} = (B_1, \dots, B_q)'$, where \underline{A} is the observed quantities and \underline{B} is the inferential quantities, and that we have made a full second-order prior specification for the elements of $\underline{Z} = (\underline{A}', \underline{B}')$. Bayes linear methods provide a way to update beliefs about \underline{B} by a linear fitting on \underline{A} using the Bayes linear updating equations for \underline{B} adjusted by \underline{A} as

$$E_1(\underline{B}) = E_0(\underline{B}) + H_{B|A} [\underline{A} - E_0(\underline{A})], \quad (2)$$

$$\text{Var}_1(\underline{B}) = \text{Var}_0(\underline{B}) - H_{B|A} \text{Cov}_0(\underline{A}, \underline{B}), \quad (3)$$

where $H_{B|A} = \text{Cov}_0(\underline{B}, \underline{A}) \text{Var}_0^{-1}(\underline{A})$, provided that $\text{Var}_0(\underline{A})$ is invertible.

3.2 Probability Kinematics

Bayes linear kinematics was introduced by Goldstein and Shaw (2004). It is named after the probability kinematics proposed by Jeffrey (1965) since the relationship of Bayes

linear kinematics to Bayes linear methods can be seen as analogous to the relationship of probability kinematics to probabilistic conditioning. So, before discussing Bayes linear kinematics, we first briefly introduce probability kinematics.

Probability kinematics is a method for revising a probability specification which depends on new probabilities over a partition. Suppose that we have a partition $A = (a_1, \dots, a_m)$ and corresponding probabilities $\Pr_0(a_k)$ and $\sum_{k=1}^m \Pr_0(a_k) = 1$. Suppose that we have obtained some information \mathcal{I} which causes us to update our probabilities of these events to $\Pr_1(a_1 | \mathcal{I}), \dots, \Pr_1(a_m | \mathcal{I})$. Then we can impose the so-called *rigidity*, or sufficiency, condition that, for any future event B ,

$$\Pr_0(B | a_k) = \Pr_1(B | a_k), \quad \forall k.$$

Therefore, the new marginal probability of B can be found by probability kinematics on $\Pr_1(a_1 | \mathcal{I}), \dots, \Pr_1(a_m | \mathcal{I})$ as

$$\Pr_1(B) = \sum_{k=1}^m \Pr_0(B | a_k) \Pr_1(a_k | \mathcal{I}).$$

An important special case is where \mathcal{I} consists of the observation of the value x of a random variable X where X is judged to be probabilistically conditionally independent of B given A . See, for example, Diaconis and Zabell (1982). In this case, $\Pr_1(B) = \Pr_0(B | X = x)$.

However, suppose that we have two different partitions, $A_1 = (a_{1,1}, \dots, a_{1,m_1})$ and $A_2 = (a_{2,1}, \dots, a_{2,m_2})$. Then successive probability kinematic revisions based on these two partitions might not be commutative if the order is reversed. A simple example of non-commutativity is given by Wilson (2011). The conditions for commutativity are discussed by Field (1978) and Diaconis and Zabell (1982).

3.3 Bayes Linear Kinematics

Bayes linear kinematics was introduced by Goldstein and Shaw (2004). It is a special form of Bayes linear analysis where, instead of observing \underline{A} , as in Section 3.1, we simply update our beliefs about this set of quantities by obtaining some information. Then those changes in our beliefs can be propagated through other unknown quantities within a Bayes linear structure. The idea is that, instead of observing \underline{A} directly and using (2) and (3), we observe some relevant information \mathcal{I} that changes our beliefs about \underline{A} so that our mean and variance become $E(\underline{A} | \mathcal{I})$ and $\text{Var}(\underline{A} | \mathcal{I})$. Then we wish to propagate these updates to \underline{B} . In particular, the information \mathcal{I} , may be observations on some related quantities \underline{X} .

In order to propagate these changes in our beliefs, we could use a full-Bayes analysis which requires a full probabilistic specification and more intensive calculations such as Markov chain Monte Carlo (MCMC) methods. However, by imposing the sufficiency conditions that (2) and (3) continue to hold when we have not observed \underline{A} but merely

changed its mean and variance, we can use Bayes linear kinematics by adjusting the expectation vector and variance matrix based on (2) and (3) and therefore we can adjust our mean and variance of \underline{Z} . Applying (2) and (3) to \underline{B} we obtain

$$E_{A|\mathcal{I}}(\underline{B}) = E_0(\underline{B}) + H_{B|A}[E(\underline{A} | \mathcal{I}) - E_0(\underline{A})], \quad (4)$$

$$\begin{aligned} \text{Var}_{A|\mathcal{I}}(\underline{B}) &= H_{B|A} \text{Var}(\underline{A} | \mathcal{I}) H'_{B|A} + \text{Var}_0(\underline{B}) - H_{B|A} \text{Var}_0(\underline{A}) H'_{B|A} \\ &= \text{Var}_0(\underline{B}) - H_{B|A} \{ \text{Var}_0(\underline{A}) - \text{Var}(\underline{A} | \mathcal{I}) \} H'_{B|A} \end{aligned} \quad (5)$$

$$\text{and} \quad \text{Cov}_{A|\mathcal{I}}(\underline{B}, \underline{A}) = H_{B|A} \text{Var}(\underline{A} | \mathcal{I}).$$

We can combine these into two updating equations for the whole of \underline{Z} by writing

$$H_{Z|A} = \text{Cov}_0(\underline{Z}, \underline{A}) \text{Var}_0^{-1}(\underline{A}) = \begin{pmatrix} I_p \\ H_{B|A} \end{pmatrix},$$

where I_p is a $p \times p$ identity matrix. Then

$$E_{A|\mathcal{I}}(\underline{Z}) = E_0(\underline{Z}) + H_{Z|A}[E(\underline{A} | \mathcal{I}) - E_0(\underline{A})] \quad (6)$$

$$\text{and} \quad \text{Var}_{\underline{A}|\mathcal{I}}(\underline{Z}) = \text{Var}_0(\underline{Z}) - H_{Z|A}[\text{Var}_0(\underline{A}) - \text{Var}(\underline{A} | \mathcal{I})] H'_{Z|A}. \quad (7)$$

Now suppose we wish to make multiple updates. As with probability kinematics, we need to consider commutativity. Suppose that we have J sets of random quantities $\mathcal{Z}_1, \dots, \mathcal{Z}_J$, where, for $j = 1, \dots, J$, the elements of \mathcal{Z}_j are arranged as a vector $\underline{Z}_j = (Z_{j,1}, \dots, Z_{j,n_j})'$. The sets \mathcal{Z}_j need not be disjoint. Suppose that a full second order prior specification, $S_0(\underline{Z}) = [E_0(\underline{Z}), \text{Var}_0(\underline{Z})]$, has been made for $\mathcal{Z} = \mathcal{Z}_1 \cup \dots \cup \mathcal{Z}_J$ and that the elements of \mathcal{Z} are arranged in a vector \underline{Z} . If we observed \underline{Z}_k then (2) and (3) would lead to an adjusted expectation $E_{(0)}(\underline{Z} | \underline{Z}_k)$ and variance $\text{Var}_{(0)}(\underline{Z} | \underline{Z}_k)$. Now suppose that, instead of observing \underline{Z}_k , information \mathcal{I}_k is received which causes our beliefs about \underline{Z}_k to be updated to $S_1(\underline{Z}_k | \mathcal{I}_k) = [E_1(\underline{Z}_k), \text{Var}_1(\underline{Z}_k)]$.

Suppose that, in this new situation, if we now observed \underline{Z}_k , our expectation would be $E_{(1)}(\underline{Z} | \underline{Z}_k)$ and our variance would be $\text{Var}_{(1)}(\underline{Z} | \underline{Z}_k)$. Then the Bayes linear kinematic update for \underline{Z} , if we gain information \mathcal{I}_k , is found by setting the adjustments to expectations and variances using specifications $S_0(\underline{Z})$ and $S_1(\underline{Z} | \mathcal{I}_k)$ equal to each other. That is we can use (2) and (3) with \underline{A} replaced by \underline{Z}_k and \underline{B} replaced by \underline{Z} . So $E_{(0)}(\underline{Z} | \underline{Z}_k) = E_{(1)}(\underline{Z} | \underline{Z}_k)$ and $\text{Var}_{(0)}(\underline{Z} | \underline{Z}_k) = \text{Var}_{(1)}(\underline{Z} | \underline{Z}_k)$.

Now suppose that, for each j ($j = 1, \dots, J$), information \mathcal{I}_j is received once and beliefs are changed for \underline{Z}_j . A Bayes linear kinematic update can be made for \underline{Z} each time. However, successive Bayes linear kinematic updates are not necessarily commutative. Making an update given \mathcal{I}_1 changes the moments of the Bayes linear structure and so the update using \mathcal{I}_2 is changed. The Bayes linear kinematic method depends on the assumption that the updating formulae do not change. 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 (2) and (3) continue to hold, but which will apply commutatively to intermediate steps, whatever

the order in which we enter the information. 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, the adjusted variance $[P(\underline{Z} | \mathcal{I})]^{-1}$ and expectation $E(\underline{Z} | \mathcal{I})$ are given by

$$P(\underline{Z} | \mathcal{I}) = \sum_{j=1}^J P(\underline{Z} | \mathcal{I}_j) - (J-1)P(\underline{Z}), \quad (8)$$

$$P(\underline{Z} | \mathcal{I})E(\underline{Z} | \mathcal{I}) = \sum_{j=1}^J P(\underline{Z} | \mathcal{I}_j)E(\underline{Z} | \mathcal{I}_j) - (J-1)P(\underline{Z})E(\underline{Z}), \quad (9)$$

where $\mathcal{I} = (\mathcal{I}_1, \dots, \mathcal{I}_J)$ and $P(\underline{Z}) = \text{Var}(\underline{Z})^{-1}$ is the prior precision matrix. Goldstein and Shaw give formal proofs in the general case.

3.4 Bayes Linear Bayes Graphical Models

In complicated models, experts often make full marginal probabilistic specifications, but they are not able to assess the full joint probability distribution for all unknown quantities in the model. Goldstein and Shaw (2004) developed a formalism for updating beliefs about these quantities which depends on Bayes linear kinematics. They introduced Bayes linear Bayes graphical models with directed graph $G = (V, E)$ where V is a collection of nodes representing quantities $(X_1, \dots, X_J, Z_1, \dots, Z_K)$, each of which may be a vector, and E is a collection of edges. A Bayes graphical model is a model where the generalised conditional independence relationship is probabilistic conditional independence (Lauritzen, 1996; Cowell et al., 1999) and by taking second-order belief as the generalised conditional independence, we obtain a Bayes linear graphical model. See Goldstein and Wilkinson (2000). A Bayes linear Bayes graphical model is a combination of fully Bayesian and Bayes linear graphical models allowing conditioning on marginal distributions of any form and taking advantage of Bayes linear kinematics to involve full conditional updates within Bayes linear adjustments. The quantities Z_1, \dots, Z_K are connected in a Bayes linear structure. Each X_j , $j = 1, \dots, J$ depends on an associated $Z_{k(j)}$, which is one of Z_1, \dots, Z_K , through a conditional probability distribution and is conditionally independent of $X_{j'}$ for $j' \neq j$ and of Z_k for $k \neq k(j)$ given $Z_{k(j)}$. See Goldstein and Shaw (2004). In the rest of this paper we will assume that Z_1, \dots, Z_K are labelled so that $k(j) = j$.

Apart from Goldstein and Shaw (2004), published work on Bayes linear Bayes graphical models includes Wilson and Farrow (2010); Wilson et al. (2013); Gosling et al. (2013); Quigley et al. (2013) and Wilson and Farrow (2017).

Example 1 Figure 1 shows a very simple Bayes linear Bayes graphical model. We have two observable quantities, X_1 and X_2 . Each X_j depends, through a conditional probability distribution, on a corresponding underlying quantity Z_j . So, beliefs about the pair (Z_j, X_j) are represented by a probability model. The relationship between Z_1 and Z_2 is described by a Bayes linear structure.

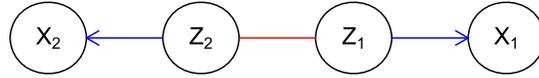


Figure 1: Bayes linear Bayes graphical model with two variables. The red undirected edge linking the pair (Z_1, Z_2) represents a Bayes linear structure (we could use a directed edge in either direction) and the directed blue arrows refer to the conditional distributions which allow full-Bayes analysis.

For example, consider the following prognostic problem. Patients suffering from chronic obstructive pulmonary disease (COPD) suffer exacerbations from time to time. Current research includes the development of systems to monitor patients and predict when the next exacerbation might occur in each patient. See, for example, Guerra et al. (2017). Many covariates might be used but suppose, for the purpose of this example, that we just use the number, X_1 , of exacerbations suffered by the patient in an earlier year. We are interested in X_2 , the time, in days, until the next exacerbation. In a very simple model we might suppose that, given the value of λ_1 , X_1 has a Poisson distribution with mean $365\lambda_1$ and, given the value of λ_2 , X_2 has an exponential distribution with mean λ_2^{-1} . Although we might expect λ_1 and λ_2 to be similar, we do not expect them to be equal since the condition of the patient changes over time. We have marginal prior distributions for λ_1 and λ_2 and we express our belief in the relationship between them by a Bayes linear structure involving quantities Z_1 and Z_2 which are related to λ_1 and λ_2 in some specified way. So we specify means and variances for Z_1 and Z_2 and a covariance between them. Then, given an observation of X_1 for a particular patient, our beliefs about Z_1 change. This changes our beliefs about Z_2 , through the Bayes linear relationship, and therefore our beliefs about X_2 change, although it may be sufficient for us to have a revised expectation of Z_2 .

Example 2 Figure 2 shows a larger Bayes linear Bayes graphical model, with $J = K = 5$. Here X_5 is denoted T to represent survival time as this model is intended for use in survival prognosis. The $m = J - 1 = 4$ quantities X_1, \dots, X_4 represent covariates. We give $\underline{Z} = (Z_1, \dots, Z_K)'$ a second-order prior specification. For generality, this Bayes linear structure is shown as a fully connected undirected graph in Figure 2. In particular applications, directed edges can be used and conditional independence might lead to the omission of some edges. If we observe X_j , then Bayes theorem is used to calculate the adjusted expectation $E(Z_j | X_j)$ and variance $\text{Var}(Z_j | X_j)$. These changes are passed through the rest of the network using Bayes linear kinematics. When $m > 1$ we use (8) and (9).

Early work on Bayes linear Bayes models used conjugate prior distributions in the full-Bayes components and applied Bayes linear kinematics to the usual parameters of these distributions. For example, if the conditional distribution of X_j given $Z_{k(j)}$ were Poisson with mean $\lambda = Z_{k(j)}$, then a gamma prior distribution would be given to λ . Observing X_j would lead to a posterior gamma distribution and the revised mean and

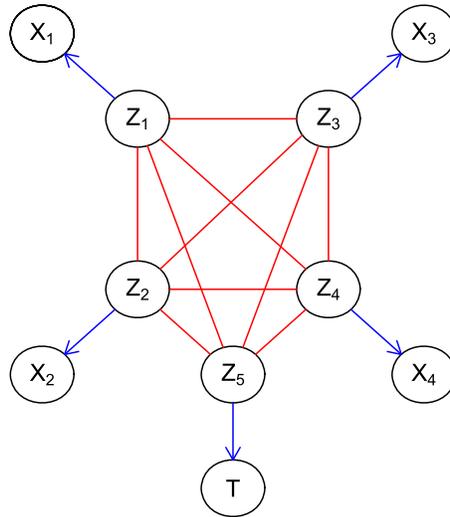


Figure 2: Bayes linear Bayes graphical model for use in survival prognosis. The red undirected edges linking Z_1, \dots, Z_5 represent a Bayes linear structure and the directed blue arrows refer to the conditional distributions which allow full-Bayes analysis.

variance would lead to changes in the moments in the Bayes linear structure which would be propagated by Bayes linear kinematics. Wilson and Farrow (2010) introduced the idea of using a transformation, or link function. This can make the use of a Bayes linear structure more reasonable by eliminating bounds on the values of parameters and removing mean-variance relationships. It can also, in certain cases, ensure that the conditions for (8) and (9) to be the unique commutative Bayes linear kinematic update are met. For example, in the Poisson case, we would use $Z_{k(j)} = \eta = \log \lambda$ instead of λ and compute the adjusted mean and variance of η given X_j , while still giving λ a gamma prior distribution. Wilson and Farrow (2017) retained the conjugate prior distribution but suggested using a “guide relationship”, as used, for example, by West et al. (1985) and Gamerman (1991). The flexibility of Bayes linear kinematics allows us to use as the adjusted mean and variance of $Z_{k(j)}$, after observing X_j , values which are approximations to the posterior mean and variance given X_j , rather than the exact values. In particular, in the Poisson case, instead of setting $Z_{k(j)} = \eta = \log \lambda$ and giving $Z_{k(j)}$ the mean and variance of the resulting log-gamma distribution, we could give $Z_{k(j)}$ the mean and variance of the normal distribution such that the corresponding lognormal distribution has the same mean and variance as the gamma prior for λ , or we could give $Z_{k(j)}$ a mean equal to the mode of the distribution of η and a variance given by $-[dl(\eta)/d\eta]^{-1}$, evaluated at the mode, where $l(\eta)$ is the log density of η . These options, in the Poisson case, are discussed by Wilson and Farrow (2017) who examine the effects of these choices on the adjustments propagated through a simple network.

Example 1 We illustrate three variations on the method using Example 1. For each of λ_1 and λ_2 , we assess the prior mean to be 0.01 and the prior upper quartile to be

about three times the prior lower quartile. This suggests a gamma $\text{Ga}(1.8, 180)$ prior distribution for each, which has a variance of $0.01/180 = 5.56 \times 10^{-5}$. Suppose that, in each case, we use the same relationship for (λ_1, Z_1) as for (λ_2, Z_2) so $E_0(Z_1) = E_0(Z_2)$ and $\text{Var}_0(Z_1) = \text{Var}_0(Z_2)$, and, in each case, we give Z_1 and Z_2 a prior correlation of 0.9. Therefore $H_{Z_2|Z_1} = 0.9$. Suppose for a particular patient, we observe six exacerbations in the earlier year. The posterior distribution for λ_1 is therefore $\text{Ga}(7.8, 545)$, with a mean of 0.0143 and a variance of 2.62×10^{-5} .

Using (4) and (5), we have, in each case,

$$\begin{aligned} E_{Z_1|X_1}(Z_2) &= E_0(Z_2) + 0.9[E(Z_1 | X_1) - E_0(Z_1)] \\ \text{and} \quad \text{Var}_{Z_1|X_1}(Z_2) &= \text{Var}_0(Z_2) - 0.9^2\{\text{Var}_0(Z_1) - \text{Var}(Z_1 | X_1)\}. \end{aligned}$$

Working directly in terms of λ_1 and λ_2 , so $Z_j = \lambda_j$, we obtain the moments in the rows labelled ‘‘Direct’’ in Table 1.

Using the method of Wilson and Farrow (2010), we obtain the moments in the rows labelled ‘‘WF10’’ in Table 1. The adjusted moments of Z_2 correspond to a $\text{Ga}(4.466, 314.9)$ distribution for $\lambda_2 | X_1$.

Now consider the ‘‘lognormal’’ method of Wilson and Farrow (2017). If we give Z_j a normal $N(-4.826, 0.4418)$ prior distribution, then $\exp(Z_j)$ has the same mean and variance as the gamma prior for λ_j . Similarly, if Z_1 has a normal $N(-4.307, 0.1206)$ distribution, then $\exp(Z_j)$ has the same mean and variance as the gamma posterior for λ_1 . We obtain the moments in the rows of Table 1 labelled ‘‘WF17’’. Using the lognormal relationship the adjusted moments correspond to the moments given for $\lambda_2 | X_1$.

For comparison, we can use a full-Bayes analysis of a model in which $\lambda_j = \exp(Z_j)$ and $Z_j \sim N(m, v)$, $j = 1, 2$. We show the results of this with two different prior moment specifications. Firstly we set $m = -4.826$ and $v = 0.4418$, the prior moments used to illustrate the Wilson and Farrow (2017) method. The results are labelled ‘‘Full Bayes 1’’ in Table 1. Secondly we set $m = -4.937$ and $v = 0.6635$ which gives the specified mean and ratio of quartiles for λ_j . We obtain the moments labelled ‘‘Full Bayes 2’’ in Table 1.

In this paper we consider a broad range of different kinds of covariates. Conjugate priors might not be available. We investigate the use of non-conjugate priors, typically using a one-dimensional numerical integration for each marginal update. For example, in the Poisson case, we set $Z_{k(j)} = \eta = \log \lambda$ but give η a normal prior and update its moments numerically when X_j is observed. We see that the resulting revised mean for a prognostic index is similar to that obtained by a full-Bayes calculation. This is described in Section 4.

	Z_j	$E_0(Z_j)$	$E(Z_1 X_1)$	$E_{Z_1 X_1}(Z_2)$	$E_{Z_1 X_1}(\lambda_2)$
Direct	λ_j	0.01	0.0143	0.0139	0.0139
WF10	η_j	-4.908	-4.312	-4.372	0.0142
WF17	$\tilde{\eta}_j$	-4.826	-4.307	-4.359	0.0140
Full Bayes 1	η_j	-4.826	-4.369	-4.415	0.0133
Nonconjugate	η_j	-4.937	-4.346	-4.405	0.0139
Full-Bayes 2	η_j	-4.937	-4.345	-4.404	0.0139
	Z_j	$\text{Var}_0(Z_j)$	$\text{Var}(Z_1 X_1)$	$\text{Var}_{Z_1 X_1}(Z_2)$	$\text{Var}_{Z_1 X_1}(\lambda_2)$
Direct	λ_j	5.56×10^{-5}	2.62×10^{-5}	3.18×10^{-5}	3.18×10^{-5}
WF10	η_j	0.7370	0.1368	0.2508	4.50×10^{-5}
WF17	$\tilde{\eta}_j$	0.4418	0.1206	0.1816	3.90×10^{-5}
Full Bayes 1	η_j	0.4418	0.1426	0.1995	3.74×10^{-5}
Nonconjugate	η_j	0.6635	0.1580	0.2541	5.57×10^{-5}
Full-Bayes 2	η_j	0.6635	0.1582	0.2547	5.25×10^{-5}

Table 1: Moments using different methods in Example 1. WF10: Wilson and Farrow (2010). WF17: Wilson and Farrow (2017) “lognormal”. Here $\eta_j = \log(\lambda_j)$ and $\tilde{\eta}_j \approx \eta_j$.

4 Non-Conjugate Marginal Updates

4.1 Introduction

While there are advantages in using a link function, as in Wilson and Farrow (2010, 2017), the use of a conjugate prior and then calculation of the change in mean and variance of a transformed parameter is a somewhat restrictive arrangement. In addition, a normal prior distribution is symmetric, has unbounded support and no mean-variance dependence. These features fit well with the use of Bayes linear updating between nodes in the network. A direct non-conjugate relationship to such a prior is closer to the usual structure in a full-Bayes analysis and, as we shall see, the results tend to be closer to those of the full-Bayes analogue. Furthermore our approach provides a general method which allows many different kinds of observational distributions. The price to be paid for this is that we need to use numerical integration to find the adjusted mean and variance of Z_j given $X_j = x_j$. However this is typically a one-dimensional integration and suitable fast approximation methods can often be used. Suppose that we give Z_j a prior distribution with density $f_j(z)$ and that the likelihood from observing $X_j = x_j$ is $L_j(z; x_j)$. Then the posterior density of Z_j is proportional to $g_j(z) = f_j(z)L_j(z; x_j)$. For example, in the Poisson case mentioned above, we might give a normal prior distribution to $\eta = \log \lambda$ and the likelihood is proportional to $\exp(-\lambda)\lambda^{x_j}$.

In suitable cases, particularly where the support is unbounded, we might use a simple normal approximation to obtain posterior moments. Let $G_j(z) = \log g_j(z)$. Then we can find the maximum m of $G_j(z)$ as an approximation to the posterior mean and use $v = -[\partial^2 G_j(z)/\partial z^2]^{-1}$ evaluated at $z = m$ as an approximation to the posterior

variance. Alternatively, we write the posterior mean as

$$E_1(Z_j | X_j = x_j) = \frac{\int_{-\infty}^{\infty} z g_j(z) dz}{\int_{-\infty}^{\infty} g_j(z) dz} = \frac{\int_{-\infty}^{\infty} \exp\{\log(z)G_j(z)\} dz}{\int_{-\infty}^{\infty} \exp\{G_j(z)\} dz}$$

and then use Laplace approximations (Tierney and Kadane, 1986) for the integrals in the numerator and denominator. Another possibility would be to use Gauss-Hermite quadrature (eg Naylor and Smith, 1982).

Example 1 To apply this method in Example 1, we use the same model as in the full-Bayes analysis except that we do not relate Z_1 and Z_2 through a bivariate normal distribution but simply through a second-order moment specification. We use the Laplace approximation to find the moments of $Z_1 | X_1$. The effect of this change is then passed to Z_2 using (4) and (5) as in the other Bayes linear kinematic methods. The results are shown in the rows labelled “Nonconjugate” in Table 1. We can see that these results are the closest of all of the Bayes linear kinematic results to the corresponding full-Bayes moments (“Full Bayes 2”). The “Nonconjugate” moments are closer to the “Full Bayes 2” moments than the “WF17” moments are to either set of full-Bayes results. This is achieved with a very small increase in computation compared to the conjugate-update methods. The greatest discrepancy between the “Nonconjugate” and “Full Bayes 2” moments is in $\text{Var}_{Z_1|X_1}(\lambda_2)$ where the full-Bayes posterior variance is obtained by integrating over the posterior distribution of Z_2 to obtain the posterior variance of $\lambda_2 = \exp(Z_2)$ but the Bayes linear kinematic method has available only the first two moments of Z_2 .

In the rest of Section 4 we consider some special types of observational variables, especially those which may be relevant to our prognostic index application.

4.2 Binary Variables

In the case of a binary variable X , a number of possibilities arise. One is simply to have separate conditional models for the two possible values of the variable. We will see an example of this in Section 5. Another is to let $X = 1$ if the corresponding $Z \geq 0$ and $X = 0$ if $Z < 0$ and, for this purpose, we assign to Z a normal prior distribution. In this case the support of the posterior distribution is bounded at zero and, for this reason, a quadrature method rather than the normal or Laplace approximation would be used. We treat binary variables as a special case of ordinal variables. See Section 4.3.

4.3 Ordinal Variables

Suppose that we have ordered categories labelled $1, \dots, S$ and X takes the value of the category label. Then $X = s$ if and only if $c_{s-1} \leq Z < c_s$ for a set of thresholds c_1, \dots, c_{S-1} where $c_0 \rightarrow -\infty$ and $c_S \rightarrow \infty$. In this case the posterior support is bounded,

often both below and above, and we use a quadrature method to find the posterior moments.

Suppose that X_j is an ordinal variable and that the conditional mean of Z_j , given $\bar{\underline{Z}}_j = \bar{z}_j$, is $m_j + \underline{b}'_j(\bar{z}_j - \bar{\underline{m}}_j)$ where $\bar{\underline{Z}}_j = (Z_1, \dots, Z_{j-1}, Z_{j+1}, \dots, Z_J)'$ and $\bar{\underline{m}}_j$ is the corresponding mean vector, and the conditional variance is w_j . Then, if $Z_j \mid \bar{\underline{Z}}_j = \bar{z}_j$ has a conditional normal distribution,

$$\Pr(X_j \leq s \mid \bar{\underline{Z}}_j = \bar{z}_j) = \Phi([c_s - m_j - \underline{b}'_j(\bar{z}_j - \bar{\underline{m}}_j)]/\sqrt{w_j}) = \Phi(k_s - \underline{d}'_j \bar{z}_j),$$

where $k_s = [c_s - m_j + \underline{b}'_j \bar{\underline{m}}_j]/\sqrt{w_j}$, $\underline{d}_j = \underline{b}_j/\sqrt{w_j}$ and $\Phi(\cdot)$ is the standard normal distribution function. So the marginal model corresponds to a conventional ordinal regression with a probit link. For identifiability, when $S > 2$, we fix two thresholds, usually $c_1 = 0$ and $c_{s-1} = 1$. In the binary case $S = 2$ and we fix $c_1 = 0$ and $w_j = 1$.

To do the numerical integration we proceed as follows. Suppose that the prior distribution of Z is $N(m, v)$. Let $u = g(z) = \Phi([z - m]/\sqrt{v})$. Then $U = g(Z)$ has a uniform $U(0, 1)$ prior distribution. Let $\tilde{c}_s = g(c_s)$ for $s = 1, \dots, S - 1$, $\tilde{s}_0 = 0$ and $\tilde{s}_S = 1$. Then the posterior distribution of U , given $X = s$, is $U(\tilde{c}_{s-1}, \tilde{c}_s)$. Then the posterior mean and variance of Z are $E_1(Z \mid X = s) = M_1$ and $\text{Var}_1(Z \mid X = s) = M_2 - M_1^2$, where

$$M_r = \frac{1}{\tilde{c}_s - \tilde{c}_{s-1}} \int_{\tilde{c}_{s-1}}^{\tilde{c}_s} [g^{-1}(u)]^r du$$

where $g^{-1}(u) = m + \sqrt{v}\Phi^{-1}(u)$. We approximate the integral M_r using $M_r \approx H^{-1} \sum_{h=1}^H [g^{-1}(u_h)]^r$, where $u_h = \tilde{c}_{s-1} + (2h - 1)(\tilde{c}_s - \tilde{c}_{s-1})/2H$ for $h = 1, \dots, H$.

4.4 Unordered Categorical Variables

Dealing with a categorical variable with $S > 2$ unordered categories, other than by conditioning the whole model on the categories, requires the use of an underlying vector variable with dimension $S - 1$. This can be handled within the general Bayes linear kinematic and Bayes linear Bayes framework since an observation X_j can be associated with a subset $\mathcal{Z}_j = \{Z_{j,1}, \dots, Z_{j,S}\}$ containing more than one of the elements of \underline{Z} . For example, we can set

$$\Pr(X_j = s \mid Z_{j,1} = z_{j,1}, \dots, Z_{j,S} = z_{j,s}) = \frac{\exp(z_{j,k})}{\sum_{s=1}^S \exp(z_{j,s})}$$

and this provides the likelihood. A constraint, such as $Z_{j,1} = 0$ or $\sum_{s=1}^S Z_{j,s} = 0$ is applied for identifiability. The elements of \mathcal{Z}_j which are not fixed can then be given a multivariate normal prior distribution with marginal variances of 1.

A numerical integration of dimension $S - 1$ is required. This presents a difficulty if S is large but, in certain special cases, such as when the categories have a hierarchical structure, the $(S - 1)$ -dimensional integration can be replaced with a sequence of integrations of lower dimension. We have not come across an application to prognostic indices where anything other than a low dimensional integration was required.

4.5 Interval Censored Variables

A variable subject to interval censoring may be handled by methods similar to those in the case of ordinal variables. If X is not censored, then $Z = X$ and we make a direct observation. If the observation is censored then we observe that $c_1 \leq Z < c_2$ for lower and upper bounds c_1 and c_2 and the posterior support is bounded.

5 Application to Bayes Linear Bayes Prognostic Networks

5.1 Introduction

Consider Figure 2. The nodes X_1, \dots, X_4 represent covariates. Just as X_j depends on Z_j , we have a lifetime T which depends on $Z_5 \equiv Z_T$. In general we can have covariates X_1, \dots, X_J depending on Z_1, \dots, Z_J respectively and then an additional dependent node T depending on an element Z_{J+1} of the Bayes linear structure. Then Z_{J+1} represents our prognostic index. When we observe some or all of the covariates this changes our expectation of Z_{J+1} and therefore the index value which we report.

In routine use, with a new patient, i , we need a mean vector and a variance-covariance matrix for the elements, $Z_{1,i}, \dots, Z_{J,i}, Z_{J+1,i}$ of the Bayes linear structure. Some or all of these moments might differ between patients because of other information about individual patients. Specifically, we might select certain important variables which are always observed, such as Age and Sex, and condition the rest of the model on these. We will refer to these as *permanent covariates*. We propose to allow the mean vector \underline{m}_i of $\underline{Z}_i = (Z_{1,i}, \dots, Z_{J,i}, Z_{J+1,i})'$ for patient i to depend on the values of the permanent covariates for this patient, typically through a linear model. However we propose that the variance-covariance matrix, V , should not depend on the permanent covariates.

The variance-covariance matrix might be developed in a general, unstructured, way, as suggested by Figure 2. Alternatively we might impose some structure and exploit conditional independences, perhaps by introducing mediating nodes which induce correlation between related covariates. This might be done by expert judgement. A subjective covariance structure might be developed using an approach similar to methods described in Farrow (2003). On the other hand we might use analysis of historical data to help determine a suitable network structure. Methods for structure learning for Bayesian networks are discussed in, for example, Heckerman et al. (1995); Neapolitan (2003); Margaritis (2003); Wang (2015). In our Bayes linear Bayes structure, the dependencies between variables are specified by the variance-covariance matrix, V , of \underline{Z}_i . Given a value for this matrix, it is straightforward to compute, for example, conditional correlations given the values of some elements and this can help in the search for potential conditional independences.

Once we have chosen a qualitative structure, we need to quantify it with values

for means, variances and covariances. These might be chosen subjectively. More likely we will use historical data and an *offline learning* phase in which we fit an analogous model, with a fully specified prior distribution, using, for example, Markov chain Monte Carlo (MCMC) methods to compute posterior summaries for model parameters. We construct a model for the joint distribution of $Z_{1,i}, \dots, Z_{J,i}, Z_{J+1,i}$, the corresponding covariates, $X_{1,i}, \dots, X_{J,i}$, and the lifetime, T . This provides a missing data model for any missing covariate values in the historical data. The model for offline learning is the same as the Bayes linear Bayes model except that we specify a full joint probability distribution and a prior distribution for the unknown model parameters, including the thresholds for ordinal covariates and the parameters involved in the means and the variance-covariance matrix in the Bayes linear structure.

Currently we use the posterior means from the offline learning as values for model parameters in our Bayes linear Bayes model. The historical data are, however, independent of future patients, given the model parameters. This raises the possibility, that we can avoid any such compromise and obtain exactly the expectations which we need. For example, in (8) and (9), clearly we can obtain the posterior expectations of $P(\underline{Z})$ and $P(\underline{Z})E(\underline{Z})$ directly from the MCMC computations in the learning phase. However further work is required to address the problem of parameter uncertainty in the adjusted expectations and precisions. Nevertheless, with a large historical data set, such effects are likely to be small.

A possible alternative to using MCMC in the offline-learning phase is the integrated nested Laplace approximation (INLA) (eg Rue et al., 2009; Martino et al., 2011). Abdul Jalal (2020) discusses the use of INLA for survival data with missing covariate values.

Once we have a fully specified model, in routine use with new patients we compute adjusted expectations of the prognostic index given observations of some or all of the covariates. Because we can do the calculation when only a subset of the covariates are observed, we can include a greater number of potential covariates in our model and therefore use more information when it is available.

5.2 The Latent Variables \underline{Z}

In the offline-learning phase, the vector $\underline{Z}_i = (Z_{1,i}, \dots, Z_{J,i}, Z_{J+1,i})'$, where $Z_{J+1,i} \equiv Z_{T,i}$, is given a multivariate normal distribution, conditional on the values for patient i of any permanent covariates $\underline{X}_i^{(p)}$, such as Age and Sex. The mean vector is $\underline{m}_i = E(\underline{Z}_i | \underline{X}_i^{(p)})$ and the variance matrix is $V = \text{Var}(\underline{Z}_i | \underline{X}_i^{(p)})$. To facilitate specification of a prior distribution for the $(J+1)(J+2)/2$ unique elements of V , we factorize the distribution of $\underline{Z}_i | \underline{X}_i^{(p)}$ into a distribution for $Z_{1,i} | \underline{X}_i^{(p)}$ and, for $j > 1$, conditional distributions of $Z_{j,i}$ given $Z_{1,i}, \dots, Z_{j-1,i}$ and $\underline{X}_i^{(p)}$. That is, for $j > 1$, we regress $Z_{j,i}$ on its predecessors. Thus

$$Z_{1,i} = m_{1,i} + \varepsilon_{1,i}, \quad \text{and} \quad Z_{j,i} = m_{j,i} + \sum_{k=1}^{j-1} \phi_{j,k}(Z_{k,i} - m_{k,i}) + \varepsilon_{j,i}, \quad (j > 1) \quad (10)$$

where $\phi_{j,k}$ is the regression coefficient of $Z_{j,i}$ on $Z_{k,i}$, $\varepsilon_{j,i} \sim N(0, \tau_j^{-1})$ and $\varepsilon_{j,i}$ is independent of $\varepsilon_{j',i'}$ unless $j = j'$ and $i = i'$. The means are given by

$$m_{j,i} = \mu_j + \sum_{h=1}^H \gamma_{j,h} X_{h,i}^{(p)}$$

where the values of the permanent covariates for patient i are $X_{1,i}^{(p)}, \dots, X_{H,i}^{(p)}$ and $\gamma_{j,h}$ is the regression coefficient of $Z_{j,i}$ on $X_{h,i}^{(p)}$. The variables $\{X_{h,i}^{(p)}\}$ might include indicator variables for any categorical permanent covariates, such as Sex.

Our model parameters thus include the $J + 1$ means, μ_1, \dots, μ_{J+1} , the $(J + 1)H$ regression coefficients, $\gamma_{1,1}, \dots, \gamma_{J+1,H}$, on the permanent covariates and our new parameters for V which are the $J + 1$ conditional precisions $\tau_1, \dots, \tau_{J+1}$ and the $J(J + 1)/2$ coefficients $\phi_{j,k}$. The form (10) gives the possibility of introducing conditional independence structure, if required, by setting some of the regression coefficients to zero and thus omitting some edges in the directed graph. The regression parameters $\phi_{j,k}$ are unrestricted and the conditional precisions τ_j just need to be positive. Expressing prior beliefs in terms of these parameters allows greater flexibility in the specification of a prior distribution than using, for example, an inverse-Wishart prior. Furthermore, especially if the order of the variables is chosen to reflect prior perceptions of the direction of causality or possible conditional independences, we believe that prior specification in terms of these regression parameters is much easier than directly in terms of the elements of V .

The use of this reparameterisation of a variance matrix was introduced, in the context of models for longitudinal data, by Pourahmadi (1999), who explained its relationship to a square-root-free Cholesky decomposition of the precision matrix V^{-1} . An explanation of this relationship is given in the Supplementary Materials (Al-Taie and Farrow, 2022). See also Daniels and Pourahmadi (2002). Ibrahim et al. (2001) use a similar structure, applied directly to the covariates, as a missing-data model in survival and Zhao (2010) discusses this and other approaches to missing data in survival.

If we wish to use the historical data to select a more sparse structure, with some edges omitted, then one possibility is to use a shrinkage prior for the parameters. See, for example, Wang (2015); Carvalho et al. (2009).

5.3 Marginal Models for Covariates

We need to specify the conditional distributions of the covariates X_j given the corresponding variables Z_j , and of T given Z_T . Some such distributions involve parameters and we use the offline learning to choose values for these. We will discuss the case of the lifetime T in Section 5.4.

In the case of an ordinal covariate, it is necessary to learn about the values of the thresholds, c_1, \dots, c_{S-1} . However, since the mean and variance of Z are both unknown, for identifiability we fix two thresholds. Thus, if $S = 3$, we can fix $c_1 = 0$ and $c_2 = 1$. If $S = 4$, then we can fix $c_1 = 0$ and $c_3 = 1$ and then give c_2 a beta prior distribution.

For $S > 4$, we can fix $c_1 = 0$ and $c_{S-1} = 1$ and then give u_1, \dots, u_{S-2} a Dirichlet prior distribution and, for $s = 2, \dots, S - 2$, let $c_s = \sum_{i=1}^{s-1} u_i$.

Some continuous covariates may be suitable for direct inclusion in the Bayes linear structure, perhaps transformed through a suitable function $g(\cdot)$ so that $Z_j = g(X_j)$. In this case Z_j might be observed. If we have more than one such *directly observed* element of \underline{Z} , the covariance structure of \underline{Z} might be such that the conditional independence conditions for the use of (8) and (9) are violated. To overcome this, when updating, we collect all such directly observed elements into a single vector observation, where there may be nonzero conditional covariances between the elements, given the non-directly observed elements of \underline{Z} . The directly observed elements can include interval-censored variables when they are not censored. In other cases we might prefer a less direct relationship. For example we might set $Z_j = E[g(X_j)]$. We might then structure the graph so that each such X_j has a single parent $Z_{k(j)}$ so that the conditional independence conditions given in Section 3.4 are satisfied. Then we can include $g(X_j)$ in updating in the same way as other covariates except that we update the mean and variance of $Z_{k(j)}$ using Bayes linear adjustment.

5.4 Lifetime Distribution

In the routine use of our Bayes linear Bayes prognostic index with new patients, we do not observe T or Z_T . We simply need to compute the expectation of a quantity related to T , given the available information. We order the variables in (10) so that Z_T comes last. Then, to compute the expectation of Z_T given any information on the covariates, that is given any current values of the expectations of Z_1, \dots, Z_J , we do not need to specify τ_T , where $\tau_T^{-1} = \sigma_T^2 = \text{Var}(\varepsilon_T)$, the conditional variance of Z_T given $\underline{Z}_c = (Z_1, \dots, Z_J)'$. To see this, first consider the dependence of Z_T on \underline{Z}_c . The variance-covariance matrix of \underline{Z} can be written as

$$\text{Var}(\underline{Z}) = \begin{pmatrix} V_c & \underline{c} \\ \underline{c}' & V_T \end{pmatrix},$$

where V_c is the variance-covariance matrix of \underline{Z}_c , \underline{c} is the covariance between \underline{Z}_c and Z_T and V_T is the marginal variance of Z_T . Therefore we can write

$$E(Z_T | \underline{Z}_c) = m_T + \underline{c}' V_c^{-1} [\underline{Z}_c - \underline{m}_c], \tag{11}$$

where m_T is the prior mean of Z_T and \underline{m}_c is the prior mean of \underline{Z}_c . Moreover (11) continues to hold when we gain information, I , which changes our beliefs about some or all of the elements of \underline{Z}_c . In particular

$$E(Z_T | I) = m_T + \underline{c}' V_c^{-1} [E(\underline{Z}_c | I) - \underline{m}_c]$$

which does not involve τ_T . Further details are given in the Supplementary Materials (Al-Taie and Farrow, 2022).

In the offline-learning phase, we use observed lifetimes and we do need to specify a model for the conditional lifetime distribution. An approach which is analogous

to traditional survival models is to use a parametric model, such as a Weibull distribution, for the conditional lifetime distribution given \underline{Z}_c . In the case of a Weibull distribution, for patient i , we use a Weibull(α, λ_i) distribution for patient i , where $\eta_i = \log(\rho_i) = \alpha^{-1} \log(\lambda_i)$ and we set $\eta_i = E(Z_{T,i} | \underline{Z}_c)$. We do not really need to spell out the relationship between T_i and $Z_{T,i}$, merely to suppose that it exists and specify the relationship between the conditional mean of $Z_{T,i}$ and the scale parameter of the lifetime distribution. However we can describe the relationship as follows, conditioning throughout on the values of $m_{1,i}, \dots, m_{J,i}$ and $\phi_{J+1,1}, \dots, \phi_{J+1,J}$.

Let $m_{T,i}^* = m_{J+1,i} + \sum_{j=1}^J \phi_{J+1,j}(Z_{j,i} - m_{j,i})$ be the conditional mean of $Z_{T,i}$ given $\underline{Z}_{c,i} = (Z_{1,i}, \dots, Z_{J,i})'$. Let $Z_{T,i} | m_{T,i}^* \sim N(m_{T,i}^*, \sigma_T^2)$. Then $T_i | m_{T,i}^* \sim \text{Weibull}(\alpha, \lambda_i)$ if

$$T_i = \exp(-m_{T,i}^*) \left\{ -\log \left[1 - \Phi \left(\frac{Z_{T,i} - m_{T,i}^*}{\sigma_T} \right) \right] \right\}^{1/\alpha}, \quad (12)$$

where $\Phi()$ is the standard normal distribution function and $\lambda_i = \exp(\alpha m_{T,i}^*)$. We show this as follows. The distribution functions of $Z_{T,i}$ and T_i are $F_N(z; m_{T,i}^*, \sigma_T^2) = \Phi(\{z - m_{T,i}^*\}/\sigma_T)$ and $F_W(t; \lambda_i, \alpha) = 1 - \exp(-\lambda_i t^\alpha)$ respectively. Then setting $F_W(T_i; \lambda_i, \alpha) = F_N(Z_{T,i}; m_{T,i}^*, \sigma_T^2)$ and solving for T_i in terms of $Z_{T,i}$ gives (12).

The ratio of two quantiles in a Weibull(α, λ_i) distribution depends only on α . Similarly, the ratio of two quantiles in the lognormal($m_{T,i}^*, \sigma_T^2$) distribution of $\exp(Z_{T,i}) | m_{T,i}^*$ depends only on σ_T^2 . The quantiles at probability p for these two distributions are $[-\log(1-p)/\lambda_i]^{1/\alpha}$ and $\exp[m_{T,i}^* + \sigma_T \Phi^{-1}(p)]$ respectively. The logarithm of the ratio of two quantiles, at probabilities p_1 and p_2 , is $R_W(\alpha, p_1, p_2) = \alpha^{-1} \log[\log(1-p_1)/\log(1-p_2)]$ in the Weibull case and $R_{LN}(\sigma_T^2, p_1, p_2) = \sigma_T [\Phi^{-1}(p_1) - \Phi^{-1}(p_2)]$ in the lognormal case. We can fix the relationship between α and σ_T^2 by choosing p_1 and p_2 and setting $R_W(\alpha, p_1, p_2) = R_{LN}(\sigma_T^2, p_1, p_2)$. For example, we can use the upper and lower quartiles, so $p_1 = 0.75$ and $p_2 = 0.25$. This gives $\alpha = 1.166\sigma_T^{-1}$.

An alternative would be to set $\eta_i = Z_{T,i}$, rather than $\eta_i = m_{T,i}^*$, in a Weibull(α, λ_i) distribution, with $\eta_i = \log(\lambda_i)/\alpha$. This corresponds to introducing lognormal frailties. In this case τ_T^{-1} would be the variance of the logarithms of the frailties and we would need to learn about both α and τ_T in the offline learning. With univariate survival data, separating the effects of these two parameters in the likelihood depends on a change in the shape of the survival function and the degree of non-proportionality in the hazards.

Of course, distributions other than the Weibull distribution could be used. Wilson and Farrow (2017) used Bayes linear kinematics in survival analysis using a semi-parametric approach with a piecewise-constant hazard model.

6 Non-Hodgkin's Lymphoma Example

6.1 Background

As an example, we use survival prognosis for patients with non-Hodgkin lymphoma. To build our model we used historical data collected by the Scotland and Newcastle

Lymphoma Group from patients in Scotland and the North of England, UK, (Proctor and Taylor, 2000). The data set contained 1391 patients. A large proportion of these patients had at least some missing covariate values. The data set was used by Sieniawski et al. (2009) who used fourteen of the many covariates. Details of these are summarised in the Supplementary Materials (Al-Taie and Farrow, 2022). Three of the fourteen covariates are quantitative, seven are binary, two are ordinal and one is an interval-censored continuous variable, where the observation was recorded as “normal”, if the measurement was inside the normal range, or as an actual value if it was not. Only 636 (45.7%) of these patients had observations of all fourteen of these covariates recorded. The outcome variable is survival time, from diagnosis. This was right-censored in 653 cases, with 738 observed death times. For further details, see, for example, Zhao (2010); Abdul Jalal (2020). Abdul Jalal (2020) gives a survey of relevant medical literature which can be used in the construction of prior distributions.

Computer code used in the example is available for download at

<https://github.com/ZigZag1dr/AlTaieFarrow1>.

6.2 Example with Six Covariates

To illustrate our method, we initially use a subset of these fourteen covariates, consisting of **Age**, **Sex**, **Stage** (Ann Arbor Stage, Carbone et al., 1971), an ordinal variable, **Hb**, blood haemoglobin concentration, a continuous variable, **Wbc**, white blood cell count, also treated as a continuous variable, and **Albumin**, a binary variable indicating whether serum albumin concentration was within the normal range. We used the logarithm of the white blood cell count to make the distribution more symmetric.

We chose to treat **Age** and **Sex**, which are always observed, as permanent covariates and to condition the rest of the model on these. Thus the means of Z_1, \dots, Z_J, Z_{J+1} , but not the variances and covariances, depend via a linear model on age and sex. We adopted a general covariance structure for the Bayes linear network, with Z_1, \dots, Z_4 , for the remaining covariates, and Z_T related using the regression structure described in Section 5.2. Specifically, the covariates were numbered so that $X_{1,i}$, $X_{2,i}$, $X_{3,i}$ and $X_{4,i}$ are, respectively, **Hb**, the logarithm of **Wbc**, **Stage** and **Albumin** (0: normal, 1: abnormal) for patient i . Let $X_{5,i} = T_i$, the lifetime for patient i . Let $X_{a,i} + 60$ be the age in years and $X_{s,i}$ indicate the sex (-1 : female, 1 : male) of patient i . Then, for $j = 1, \dots, 5$, $m_{j,i} = \mu_j + \gamma_{j,a}X_{a,i} + \gamma_{j,s}X_{s,i}$. The regression structure is then given by (10).

The marginal models were as follows. For $j = 1, 2$, $X_{j,i} = Z_{j,i}$. For $j = 3$ (**Stage**), $X_{j,3} = s$ if and only if $c_{s-1} \leq Z_{3,i} < c_s$ for cut-points $c_0 \rightarrow -\infty, c_1 = 0, c_2, c_3 = 1, c_4 \rightarrow \infty$. For $j = 4$ (**Albumin**), $X_{4,i} = 0$ if $Z_{4,i} < 0$ and $X_{4,i} = 1$ if $Z_{4,i} \geq 0$. For $j = 5$ (lifetime), we used a Weibull distribution. Of course other forms of lifetime distribution could also be used, but, in our example, we used $T_i \sim \text{Weibull}(\alpha, \lambda_i)$, where

$$\log \lambda_i = \alpha \left\{ m_{5,i} + \sum_{k=1}^4 \phi_{5,k}(Z_{k,i} - m_{k,i}) \right\}.$$

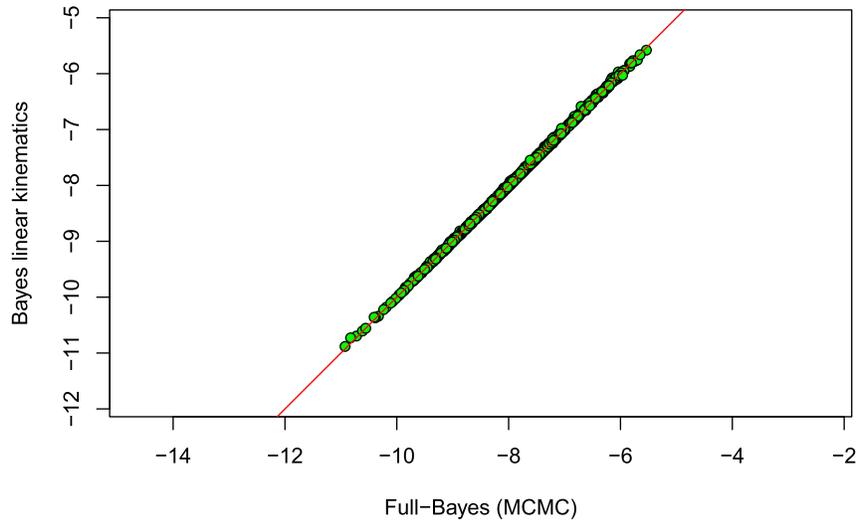


Figure 3: Bayes linear kinematic prognostic index values against the corresponding “full-Bayes” values in the 6-covariate example, with line of equality in red.

The computations in the offline-learning phase were done using **JAGS** through the **R** package `rjags` (Plummer, 2017; R Core Team, 2018). The **JAGS** model specification is given in the Supplementary Materials (Al-Taie and Farrow, 2022).

Having completed the offline learning, we calculated the Bayes linear kinematic (BLK) prognostic index values for all of the patients in the dataset with parameter values obtained from the offline learning. For comparison with the BLK prognostic index values, we also used MCMC to calculate “full Bayes” values. To make the results comparable and represent routine use in practice, we fixed parameters in the “full Bayes” calculation at the values obtained from the offline learning. The **R** code for the Bayes linear kinematic calculations and the **JAGS** model specification for the full-Bayes comparison are given in the Supplementary Materials (Al-Taie and Farrow, 2022).

Figure 3 shows the BLK index values plotted against the corresponding “full-Bayes” values. We can see that the values from the BLK adjustment are very close to the posterior means from the full-Bayes analysis. To be able to see the agreement more clearly we show a Bland and Altman agreement plot (Bland and Altman, 1986) in Figure 4. The difference between the BLK index value and the “full-Bayes” value is plotted against the mean of these values for each patient. We can see that the mean difference is very close to zero and the differences are small compared to the scale of the index values. Some points are shown coloured. These are cases where the value of at least one covariate is missing. Only **Hb**, **Wbc** and **Albumin** have missing values. There were no missing values for **Stage**. There is no obvious pattern to the locations of the missing-value cases.

Using non-conjugate marginal updates with normal marginal priors has brought the

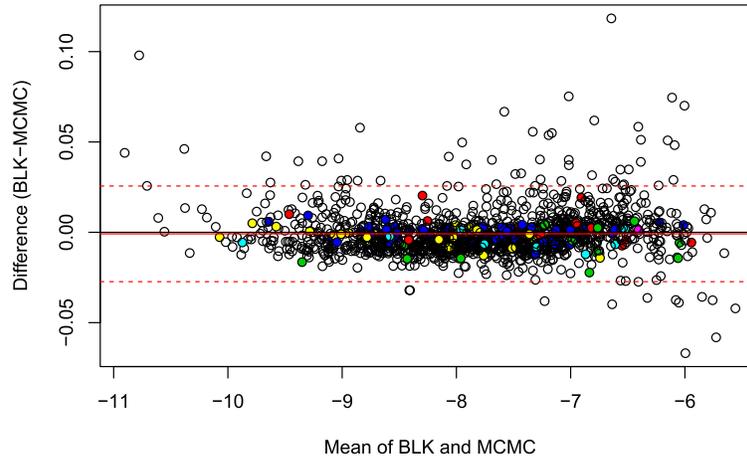


Figure 4: Bland-Altman plot for Bayes linear kinematic prognostic index values and “full-Bayes” values in the 6-covariate example. The red lines mark the mean difference and mean ± 2 std.dev. Coloured points are cases with missing covariate values as follows. Red: **Hb**, Green: **Wbc**, Dark blue: **Albumin**, Light blue: **Hb**, **Wbc**, Pink: **Hb**, **Albumin**, Yellow: **Wbc**, **Albumin**, Grey: **Hb**, **Wbc**, **Albumin**.

results of the Bayes linear kinematic analysis closer to those obtained in an analogous full-Bayes analysis when compared to earlier work. The difference between the Bayes linear kinematic analysis and the full-Bayes analogue is now largely in how changes in belief are propagated through \underline{Z} when data are observed, whether this is by Bayes linear kinematics using only first and second moments or a more complicated change in a joint probability distribution. This example suggests that, in practical terms, there is little difference between the results.

6.3 Example Using Ten Covariates

We also investigated the construction of a network using ten of the covariates used by Sieniawski et al. (2009). As in the 6-covariate example, we treated **Age** and **Sex** as permanent covariates. In the offline-learning phase, the additional four covariates were appended to the regression structure following the four in the six-covariate example, in the order as follows: **LDH**, serum lactate dehydrogenase level, units per litre, an interval-censored variable; **ECOG**, Eastern Cooperative Oncology Group performance status (Oken et al., 1982), an ordinal variable with five categories; **AP**, serum alkaline phosphatase concentration, a binary variable (normal or high); **Urea**, serum urea concentration, a binary variable (normal or high).

In the case of **LDH**, observations within the normal range were simply recorded as “normal” but values outside the normal range were recorded numerically. The normal range varies between centres of data collection. Zhao et al. (2014) converted LDH values into an ordinal variable by first dividing the LDH value by the upper limit of the normal

range and then categorising this ratio, R , into three categories: 1 for $R \leq 1$, 2 for $1 < R \leq 3$ and 3 for $R > 3$. Like Zhao et al. (2014), we transformed the **LDH** values using the limits of the normal range. However we did not convert observed values into an ordinal variable but rescaled so that, if Y is the original **LDH** value and y_0 and y_1 are the lower and upper limits of the normal range, then we used $X = (Y - y_0)(y_1 - y_0)$ as our covariate so that the lower and upper limits are always 0 and 1. Values between these limits are interval censored. If the observation is not censored, then it is treated as a continuous, normally distributed, observation. If the observation is censored, then it is treated as an observation on an ordinal variable with three categories, 0, 1 and 2, with the observed value being 1.

After completing the offline-learning, as in the six-covariate example, computation of the BLK index values for patients was very fast and no problems were encountered. The corresponding “full-Bayes” computation, used for comparison, involved sampling values for a number of unobserved elements of \underline{Z} . If no covariates were missing and **LDH** was not censored, this number was six. Any missing covariate increased this number by one, as did **LDH** being censored. Mixing and convergence of the MCMC chains need to be considered in the “full-Bayes” calculation. In contrast, the Bayes linear kinematic computation is deterministic so there were no concerns about mixing or convergence. As in the six-covariate example, there is a good agreement between the index values given by Bayes linear kinematics and the “full-Bayes” analogue. Further details of this larger example are given in the Supplementary Materials (Al-Taie and Farrow, 2022).

7 Conclusions

In this paper, we have extended the applicability of Bayes linear kinematics and Bayes linear Bayes graphical models by developing the use of non-conjugate marginal prior distributions for observable variables. We have then described the application of these ideas to the routine calculation of prognostic index values in medical survival.

The purpose of our Bayes linear Bayes prognostic network is the quick and easy routine calculation of prognostic index values for new patients, potentially using a large number of covariates but able to work when only some of these are observed. The missing-data ability can be achieved in a full-Bayes model by modelling the joint distribution of all of the variables, rather than just the conditional distribution of the lifetime given the covariates. However, in a full-Bayes model, we would need to integrate over the joint distribution of the missing covariates, conditional on the observed values, which may be computationally demanding. In our Bayes linear Bayes network, even with non-conjugate marginal updates, we typically need no more than a series of one-dimensional integrations which can usually be done very quickly.

Furthermore, a full-Bayes analysis typically requires a lot of decisions to be made about the forms of relationships between variables and these choices may have little basis either in expert judgement or the analysis of historical data. In contrast, the Bayes linear Bayes approach requires a more limited specification of relationships in terms of first and second moments and focussing on these more limited judgements might lead to sounder choices.

Our method might be regarded as an approximation to a full-Bayes analysis. Wilson and Farrow (2017) compared the behaviour of Bayes linear kinematic belief adjustments with full-Bayes posterior inferences in the case of a piecewise constant hazard survival model and found that the results were generally close. Our use of non-conjugate updates brings our model closer to the corresponding full-Bayes model and gives Bayes linear kinematic adjusted expectations which are even closer to the full-Bayes posterior means.

Our prototype prognostic network produces prognostic index values using all, or only some, of the possible covariates almost instantly and has the potential to be used, for example, as a Web-based calculator.

An alternative method, not involving BLK or MCMC in the computation of an index value for a new patient, might be developed by using the offline-learning phase to develop a separate predictive formula for each possible configuration of missing and censored covariates. However the number of such configurations might be large and some might not be present in the historical data, or only represented by a small number of cases. It would therefore be necessary to use a model for the joint distribution of the covariates and, when different types of covariates are involved, the resulting calculations for the conditional expectation of the index might not be straightforward.

This paper is not intended to deal with questions such as variable selection. It is assumed that the covariates are all thought to contain information which is at least potentially useful in prognosis. Similarly, the paper is not concerned specifically with answering questions of whether particular covariates are or are not useful predictors of survival. We might, however, learn something about this latter question in the offline-learning phase. For example, a particular latent variable Z_j might turn out to be almost conditionally uncorrelated with Z_T given some other elements of \underline{Z} . This latter possibility is interesting since there may be little to be gained by measuring X_j if certain other covariates have already been measured. This leads to the interesting decision problem of whether to observe additional covariates when making the observations might entail costs or risks.

Supplementary Material

Bayes linear Bayes networks with an application to prognostic indices Supplementary Materials (DOI: [10.1214/22-BA1314SUPP](https://doi.org/10.1214/22-BA1314SUPP); .pdf). The Supplementary Material contains the following items. Numbers in parentheses refer to sections of this paper. Sections 1 and 2 give further details of the square-root-free Cholesky decomposition of the precision matrix (5.2) and the adjusted expectation of Z_T (5.4) respectively. Section 3 gives details of the data set in the non-Hodgkin's lymphoma example (6) and the analyses with six and ten covariates are described in Sections 4 and 5. JAGS model specifications for the offline learning and the "full-Bayes" comparison are given in Sections 7 and 9 while Section 8 gives R functions used in the Bayes linear kinematic calculations. The code is all available in a repository specified in Section 6.

References

- Abdul Jalal, M. I. (2020). “Bayesian Survival Analysis with Missing Data using Integrated Nested Laplace Approximation.” Ph.D. thesis, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK. 451, 455
- Al-Taie, W. A. J. and Farrow, M. (2022). “Supplementary Material for “Bayes linear Bayes networks with an application to prognostic indices”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/22-BA1314SUPP>. 452, 453, 455, 456, 458
- Bland, J. A. and Altman, D. G. (1986). “Statistical methods for assessing agreement between two methods of clinical measurement.” *The Lancet*, 327(8476): 307–310. 456
- Carbone, P., Kaplan, H., Mushoff, K., Smithers, D., and Tubiana, M. (1971). “Report of the Committee on Hodgkin’s Disease Staging Classification.” *Cancer Research*, 31: 1860–1861. 455
- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2009). “Handling sparsity via the horseshoe.” In van Dyk, D. and Welling, M. (eds.), *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research*, volume 5, 73–80. Clearwater Beach, Florida, USA. 452
- Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (1999). *Probabilistic Networks and Expert Systems*. New York: Springer-Verlag. MR1697175. 443
- Daniels, M. J. and Pourahmadi, M. (2002). “Bayesian analysis of covariance matrices and dynamic models for longitudinal data.” *Biometrika*, 89: 553–566. MR1929162. doi: <https://doi.org/10.1093/biomet/89.3.553>. 452
- Diaconis, P. and Zabell, S. L. (1982). “Updating subjective probability.” *Journal of the American Statistical Association*, 77: 822–830. MR0686405. 441
- Farrow, M. (2003). “Practical building of subjective covariance structures for large complicated systems.” *The Statistician*, 52: 553–573. MR2011147. doi: <https://doi.org/10.1046/j.0039-0526.2003.00418.x>. 450
- Field, H. (1978). “A note on Jeffrey conditionalization.” *Philosophy of Science*, 45(3): 361–367. MR0505922. doi: <https://doi.org/10.1086/288812>. 441
- Gamerman, D. (1991). “Dynamic Bayesian models for survival data.” *Applied Statistics*, 40: 63–79. 445
- Goldstein, M. and Shaw, S. (2004). “Bayes linear kinematics and Bayes linear Bayes graphical models.” *Biometrika*, 91: 425–446. MR2081311. doi: <https://doi.org/10.1093/biomet/91.2.425>. 437, 440, 441, 443
- Goldstein, M. and Wilkinson, D. J. (2000). “Bayes linear analysis for graphical models: The geometric approach to local computation and interpretive graphics.” *Statistics and Computing*, 10(4): 311–324. 443
- Goldstein, M. and Wooff, D. (2007). *Bayes Linear Statistics*. Wiley. MR2335584. doi: <https://doi.org/10.1002/9780470065662>. 437, 440

- Gosling, J. P., Hart, A., Owen, H., Davies, M., Li, J., and MacKay, C. (2013). “A Bayes linear approach to weight-of-evidence risk assessment for skin allergy.” *Bayesian Analysis*, 8: 169–186. MR3036258. doi: <https://doi.org/10.1214/13-BA807>. 443
- Guerra, B., Gaveikaite, V., Bianchi, C., and Puhan, M. A. (2017). “Prediction models for exacerbations in patients with COPD.” *European Respiratory Review*, 26(143): 160061. 444
- Heckerman, D., Geiger, D., and Chickering, D. M. (1995). “Learning Bayesian networks: The combination of knowledge and statistical data.” *Machine learning*, 20(3): 197–243. MR1615024. 450
- Henderson, R., Jones, M., and Stare, J. (2001). “Accuracy of point predictions in survival analysis.” *Statistics in Medicine*, 20(20): 3083–3096. 438
- Ibrahim, J. G., Chen, M., and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer-Verlag. MR1876598. doi: <https://doi.org/10.1007/978-1-4757-3447-8>. 452
- Jeffrey, R. C. (1965). *The Logic of Decision*. New York: McGraw-Hill. MR0233448. 440
- Jensen, F. V. and Nielsen, T. D. (2007). *Bayesian Networks and Decision Graphs*. New York: Springer, 2nd edition. MR2344166. doi: <https://doi.org/10.1007/978-0-387-68282-2>. 438
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford: Clarendon Press. MR1419991. 443
- Lauritzen, S. L. and Spiegelhalter, D. J. (1988). “Local computations with probabilities on graphical structures and their application to expert systems.” *Journal of the Royal Statistical Society, B*, 50: 157–224. MR0964177. 438
- Margaritis, D. (2003). “Learning Bayesian Network Model Structure from Data.” Ph.D. thesis, School of Computer Science, Carnegie Mellon University. 450
- Martino, S., Akerkar, R., and Rue, H. (2011). “Approximate Bayesian inference for survival models.” *Scandinavian Journal of Statistics*, 38: 514–528. MR2833844. doi: <https://doi.org/10.1111/j.1467-9469.2010.00715.x>. 451
- Naylor, J. C. and Smith, A. F. M. (1982). “Applications of a method for the efficient computation of posterior distributions.” *Applied Statistics*, 31: 214–225. MR0694917. doi: <https://doi.org/10.2307/2347995>. 448
- Neapolitan, R. E. (2003). *Learning Bayesian Networks*. Morgan Kaufmann. 450
- Oken, M., Creech, R., Tormey, D., Horton, J., Davis, T. E., McFadden, E. T., and Carbone, P. P. (1982). “Toxicity and response criteria of the Eastern Cooperative Oncology Group.” *American Journal of Clinical Oncology*, 5: 649–655. 457
- Plummer, M. (2017). *JAGS: Just Another Gibbs Sampler, Version 4.3.0*. URL <http://mcmc-jags.sourceforge.net/> 456
- Pourahmadi, M. (1999). “Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation.” *Biometrika*, 86: 677–690. MR1723786. doi: <https://doi.org/10.1093/biomet/86.3.677>. 452

- Proctor, S. and Taylor, P. (2000). “A practical guide to continuous population-based data collection (PACE), a process facilitating uniformity of care and research into practice.” *QJM*, 93(2): 67–73. [MR0507062](#). 455
- Quigley, J., Wilson, K. J., Walls, L., and Bedford, T. (2013). “A Bayes linear Bayes method for estimation of correlated event rates.” *Risk Analysis*, 33: 2209–2224. 443
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/> 456
- Rue, H., Martino, S., and Chopin, N. (2009). “Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations.” *Journal of the Royal Statistical Society, B*, 71: 319–392. [MR2649602](#). doi: <https://doi.org/10.1111/j.1467-9868.2008.00700.x>. 451
- Sieniawski, M., Farrow, M., Zhao, X., Wilkinson, J., Mainou-Fowler, T., White, J., MacIntyre, J., and Proctor, S. J. (2009). “A novel Bayesian prognostic index for diffuse large B-cell lymphoma: A new powerful tool for prediction of outcome.” 51st American Society of Hematology Annual Meeting and Exposition, New Orleans. URL <https://doi.org/10.1182.V114.22.1666.1666> 455, 457
- Tierney, L. and Kadane, J. B. (1986). “Accurate approximations for posterior moments and marginal densities.” *Journal of the American Statistical Association*, 81: 82–86. [MR0830567](#). 448
- Wang, H. (2015). “Scaling it up: Stochastic search structure learning in graphical models.” *Bayesian Analysis*, 10: 351–377. [MR3420886](#). doi: <https://doi.org/10.1214/14-BA916>. 450, 452
- West, M., Harrison, P. J., and Migon, H. S. (1985). “Dynamic generalized linear models and Bayesian forecasting.” *Journal of the American Statistical Association*, 80: 73–83. [MR0786598](#). 445
- Wilson, K. J. (2011). “Belief representation for counts in Bayesian inference and experimental design.” Ph.D. thesis, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK. 441
- Wilson, K. J. and Farrow, M. (2010). “Bayes linear kinematics in the analysis of failure rates and failure time distributions.” *Journal of Risk and Reliability*, 224: 309–321. 443, 445, 446, 447
- Wilson, K. J. and Farrow, M. (2017). “Bayes linear kinematics in a dynamic survival model.” *International Journal of Approximate Reasoning*, 80: 239–256. [MR3581559](#). doi: <https://doi.org/10.1016/j.ijar.2016.09.010>. 443, 445, 446, 447, 454, 459
- Wilson, K. J., Quigley, J., Bedford, T., and Walls, L. (2013). “Bayes linear Bayes graphical models in the design of optimal test strategies.” *International Journal of Performativity Engineering*, 9: 715–728. 443
- Zhao, X. (2010). “Bayesian survival analysis for prognostic index development with

many covariates and missing data.” Ph.D. thesis, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK. [452](#), [455](#)

Zhao, Z., Sehn, L. H., Rademaker, A. W., Gordon, L. I., Lacase, A. S., Crosby-Thompson, A., and Vanderplas, A. (2014). “An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era.” *Blood*, 123(26): 837–842. [457](#), [458](#)

Acknowledgments

The authors thank the three reviewers for their constructive comments and suggestions which have helped to improve the paper.