A Bayes Linear Bayes Method for Estimation of Correlated Event Rates

Risk Analysis, Vol. , No. , 0000 DOI: 000

Typically full Bayesian estimation of correlated event rates can be computationally challenging since estimators are intractable. When estimation of event rates represents one activity within a larger modelling process, there is an incentive to develop a more efficient inference than provided by a full Bayesian model. We develop a new subjective inference method for correlated event rates based on a Bayes linear Bayes model under the assumption that events are generated from a homogeneous Poisson process. To reduce the elicitation burden we introduce homogenisation factors to the model and, as an alternative to a subjective prior, an empirical method using the method of moments is developed. Inference under the new method is compared against estimates obtained under a full Bayesian model, which takes a multivariate gamma prior, where the predictive and posterior distributions are derived in terms of well-known functions. The mathematical properties of both models are presented. A simulation study shows that the Bayes linear Bayes inference method and the full Bayesian model provide equally reliable estimates. An illustrative example, motivated by a problem of estimating correlated event rates across different users in a simple supply chain, shows how ignoring the correlation leads to biased estimation of event rates.

KEY WORDS: Correlated event rates, Bayes linear kinematics, homogenisation factors, empirical Bayes, supply chain risk.

0272-4332//0100-0001$22.00/1

1. INTRODUCTION

Estimating multiple event rates when these unknown rates are not statistically independent is challenging. Practically there are many motivations for considering inference for multiple event rates in the presence of correlation. For example, consider risk and resilience analysis of supply networks where a model should capture the dependencies between events affecting suppliers, customers and the focal firm following some shock event [(1)] or a safety risk analysis of a national rail network where the rates of passenger slips, trips and fall events at multiple stations are to be estimated [(2)]. The consequences of these events may vary in severity and be classified as, say, minor injuries, major injuries and fatalities where correlation may exist between the event rates between these severity classes for a specific station given the common environment experienced by passengers. Finally, consider reliability databases, whether inter or intra-organizational, such as the German ZEDB database [(3)] and the Nordic t-book [(4)]. Typically these databases support estimation of event rates for relevant failure modes based on experience data for multiple components of identical design specification that are operated and maintained within industrial plants sited in different global locations. Such conditioning factors suggest that correlation may exist between event rates for multiple items of the same type.

Many approaches to the practical problems discussed, including the methods developed in [(5)], assume an underlying model where the event rates of each process are assumed to be statistically independent, conditional on some unknowns. Within the context of Bayesian inference, if these unknowns are hyperparameters which are specified exactly in the prior distribution, then the assumption of conditional statistical independence is equivalent to saying that making observations on one of the rates would not change our estimates, when updating, of any of the other rates. **This in general will not be the case.**

In the situation where correlation exists between event rates, we can define a full multivariate distribution for all the unknown parameters to incorporate the dependence, for example using a copula, and then apply Bayes Theorem in the light of observed data to update estimates [(6,7)]. This allows correlations to be incorporated into the modelling procedure and inform estimates of the rates based on information for many events. Doing so within a full Bayesian framework ensures that the resulting estimates are coherent and theoretically sound.

Two-stage Bayesian models have been used in risk analyses [(8,9,10)] and particularly in the estimation of failure rates in nuclear power plants [(11,12)]. Typically, such models assume that events are realised from Poisson distributions. Failure rates are then given a prior distribution and the rates are assumed independent conditional on the hyperparameters of the relevant distribution. The hyperparameters are themselves then given prior distributions. Such a model was utilised by [(8)] to estimate event rates as part of a probabilistic risk assessment. [(9)] took a two-stage model considering initiating event frequencies at nuclear power plants. [(10)] criticised the two-stage Bayesian model for switching the order of integrals when calculating posterior distributions, indicating that when such improper integrals do not converge uniformly over the parameter space this could affect the solution. However, Hofer's proposed solution to this appears to lead to conflicting assumptions in his model [(12)].

[(11)] used a two-stage Bayesian model to estimate failure rates in the Swedish nuclear sector. [(12)] considered the approach used in the German ZEDB database and developed a two-stage Bayesian model. As part of their model, [(12)] compared tree types of non-informative prior; Jeffrey's prior, a uniform prior and the one suggested by [(11)]. In such a model, when observed failure events are rare, the assumption of the form of the non-informative prior tends to dominate the inference.

Obtaining the posterior distribution generally involves numerical integration such as Markov Chain Monte Carlo (MCMC) or simulation, and hence can be extremely computationally intensive [(13)]. An alternative form of subjective analysis that allows updating to be performed analytically is Bayes linear methods [(14)]. A Bayes linear analysis is based on expectations rather than probability distributions. Updating takes the form of linear fitting rather than Bayes Theorem. Bayes linear kinematics [(15,16)], a form of Bayes linear analysis in which changes in the belief about some unknowns can be used to update beliefs about others in a Bayes linear structure, has potential computational advantages over full Bayesian updating for such problems. When estimation of event rates represents one activity within a larger modelling process, such as is often the case in risk analysis where event rates are estimated at component level but then feed into a larger system level model, then there is an incentive to develop a more efficient inference than provided by a full Bayesian model. While Bayes linear methods have previously been applied in reliability assessment by [(17,18)], these have been to different problems from that considered here.

Another challenge within a risk context is that we often require estimates of rare events. Thus specifying a subjective prior may be cognitively challenging for experts yet this prior will play a dominant role in estimates. An alternative to eliciting a prior is to use an empirical Bayes procedure [(19)] where the events generated by each process are pooled to estimate each event rate as a weighted average of the pooled event rate and the observed frequency of that particular event from data. In [(5)], homogenisation factors are introduced in order to increase the effectiveness of the pooling process and hence the accuracy of the estimates obtained. To date, empirical Bayes inference for event rates has only been developed under an assumption of statistical independence. See, for example, [(20,5)].

The major contribution in this article is the development of more efficient Bayesian inference for correlated event rates. Our new method is based on a Bayes linear Bayes model and we consider situations where we have both subjective and empirical priors. We bound our study to the case where events are assumed to be generated from a homogeneous Poisson process and the marginal prior distributions for the rates are assumed to be gamma. From the perspective of our motivating risk context this is not an unreasonable initial assumption in many situations, such as when components are in a so-called useful life period. While we motivate and illustrate our proposed methods by drawing upon examples from a risk context, the methods do have general applicability.

In order to support comparison between our new method and existing theory, we begin by considering a full Bayesian model. We express the predictive and posterior distributions in terms of well-known functions, and derive desirable properties about the posterior estimates. Our new method is then introduced by adjusting the prior estimates using a mixture of full Bayesian and Bayes linear kinematic updating to support faster and easier calculations than in the full Bayes model. So far, the prior is assumed to be specified subjectively, but we now provide a method for empirically obtaining priors and examine their statistical consistency. As well as examining the theoretical properties of our inference methods, we discuss the comparative performance of our proposed Bayes linear Bayes approach relative to full Bayesian inference based on a simulation study to examine the relative accuracy of estimates. An illustrative example, motivated by a real industrial problem across a supply chain but suitably de-sensitised, provides further insight into the use of the proposed inference and allows us to examine not only the results under the alternative inference methods but also the impact of failing to account for the correlation between events.

The remainder of the paper is structured as follows. Section 2 describes the full Bayesian model and provides some useful properties of the model. In Section 3 the new Bayes linear Bayes inference procedure for correlated event rates is explained. Section 4 then gives an alternative prior specification procedure based on empirical Bayes. In Section 5 we present the findings of our simulation study to evaluate the alternative Bayesian inference methods and in Section 6 we show the illustrative example. Finally, in Section 7, we suggest future work as well as making some concluding remarks.

2. A FULL BAYES MODEL

We consider the situation which can be represented by a pool of p Homogeneous Poisson Processes(HPP) where events for the i'th process are realised at a rate of λ_i and as such the conditional probability function for the number of events realised in an interval of length t is expressed in the following

$$P\left(N_i = n \middle| \lambda_i\right) = \frac{\left(\lambda_i t\right)^n e^{-\lambda_i t}}{n!}, \lambda_i > 0,\ i = 1, 2, ..., p,\ n = 0, 1, 2, ..\ t > 0.$$

That is, N_i, N_j are conditionally independent given λ_i, λ_j. The realisations from the HPP are assumed to be conditionally independent given the rates.

Many possible prior distributions can represent a situation in which the rates of different events, $\lambda_i, \lambda_j,\ i \neq j$, are not assumed independent. We use one in which posterior distributions can be expressed in terms of well known functions. This can then be used to model correlated event rates and then later as a measure of comparison for the Bayes linear Bayes approach. Specifically, a multivariate Gamma distribution will be taken as the form of the prior distribution. This is expressed in the following

$$\pi\left(\boldsymbol{\lambda}\right) = \left[\prod_{i=1}^{p} \frac{\left(\frac{\phi+\theta}{h_i}\right)^{r} \lambda_i^{r-1} e^{-\left(\frac{\phi+\theta}{h_i}\right)\lambda_i}}{\Gamma(r)}\right] \left(\frac{\phi}{\phi+\theta}\right)^{r} \times {}_0F_{p-1}\left(\left[\,\right], \left[r\right], \left(\phi+\theta\right)^{p-1} \theta \prod_{i=1}^{p} \frac{\lambda_i}{h_i}\right), \tag{1}$$

for $\boldsymbol{\lambda} = (\lambda_1 \ldots, \lambda_p)$, where r, ϕ, and θ are parameters associated with the marginal distributions and correlations between the rates and ${}_0F_{p-1}\left(.\right)$ is the hypergeometric function. That is,

$${}_0F_{p-1}\left(\left[\,\right], \left[r\right], \left(\phi+\theta\right)^{p-1} \theta \prod_{i=1}^{p} \frac{\lambda_i}{h_i}\right) = \sum_{m=0}^{\infty} \frac{\left[\Gamma(r)\right]^{p-1} \left[\left(\phi+\theta\right)^{p-1} \theta \prod_{i=1}^{p} \frac{\lambda_i}{h_i}\right]^{m}}{\left[\Gamma\left(r+m\right)\right]^{p-1} m!}.$$

The parameter $h_i > 0$ serves as a homogenisation factor, such that we construct a hierarchical model where we assume the parameters λ_i are sampled from a multivariate Gamma distribution with identical marginal distributions. We set $h_1 = 1$. We assume the remaining homogenisation factors are provided through expert judgement, see[(5)] for a further discussion on homogenisation factors within HPP's. That is, we treat r, ϕ as parameters which provide a "base rate" for the occurrence of any event and then elicit a quantity $h_i > 0$ which is used to provide a suitable order of magnitude for the rate of the event in question.

From[(21)] we know that the model described above has the following properties. Firstly, the marginal distributions are Gamma as in the following

$$\pi\left(\lambda_i\right) = \frac{\left(\frac{\phi}{h_i}\right)^{r} \lambda_i^{r-1}}{\Gamma\left(r\right)} e^{-\frac{\phi}{h_i}\lambda_i}, \lambda_i > 0, r > 0, \phi > 0.$$

The parameter r acts as the shape parameter and the ration h_i/ϕ is the scale parameter. That is, we assume that each of the rates has the same marginal prior distribution up to the homogenisation factor. Secondly, the correlation between two different rates is a constant[(21)]. In particular, it is given by

$$\mathrm{Corr}\left(\lambda_i, \lambda_j\right) = \frac{\theta}{\phi+\theta} = \rho, \text{ for } i \neq j.$$

The predictive distribution for the number of events $\boldsymbol{n} = (n_1, \ldots, n_p)$ for each process within the pool can be obtained by mixing the conditionally independent Poisson distributions over the multivariate gamma, to obtain

$$
\begin{aligned}
P\left(\boldsymbol{N}=\boldsymbol{n}\right) &= \left[\prod_{i=1}^{p}\frac{\Gamma(n_i+r)}{\Gamma(r)n_i!}\left(\frac{\phi+\theta}{\phi+\theta+h_it}\right)^r\left(\frac{h_it}{\phi+\theta+h_it}\right)^{n_i}\right]\left(\frac{\phi}{\phi+\theta}\right)^r \\
&\quad \times {}_pF_{p-1}\left([n_i+r],[r],\left[\prod_{i=1}^{p}\left(\frac{\phi+\theta}{\phi+\theta+h_it}\right)\right]\frac{\theta}{\phi+\theta}\right),
\end{aligned}
$$

where the hypergeometric function ${}_pF_{p-1}(\cdot)$ is

$$
{}_pF_{p-1}=\sum_{m=0}^{\infty}\frac{\Gamma(r)^{p-1}}{\Gamma(r+m)^{p-1}}\prod_{i=1}^{p}\frac{\Gamma(n_i+r+m)^p}{\Gamma(n_i+r)^p}\frac{1}{m!}\times\left[(\phi+\theta)^{p-1}\theta\prod_{i=1}^{p}\left(\frac{\phi+\theta}{\phi+\theta+h_it}\right)\frac{\theta}{\phi+\theta}\right]^m,
$$

This multivariate distribution has a number of properties which can be calculated analytically. Firstly, the marginal distribution for the number of events from a process is Negative Binomial:

$$
\begin{aligned}
P\left(N_i=n\right) &= \int_0^\infty P\left(N_i=n|\,\lambda_i\right)\pi\left(\lambda_i\right)d\lambda_i \\
&= \frac{\Gamma(r+n)}{\Gamma(r)n!}\left(\frac{\phi}{h_it+\phi}\right)^r\left(\frac{h_it}{h_it+\phi}\right)^n.
\end{aligned}
$$

The correlation coefficient can be expressed as follows

$$
\mathrm{Corr}\left(N_i\left(t\right),N_j\left(t\right)\right)=\left(\frac{\theta}{\phi+\theta}\right)\sqrt{\frac{h_ih_jt^2}{\left(\phi+h_it\right)\left(\phi+h_jt\right)}}. \tag{2}
$$

We can derive the posterior distribution for the rates. Doing so results in

$$
\begin{aligned}
\pi\left(\boldsymbol{\lambda}|\,\boldsymbol{n}\right) &= \left[\prod_{i=1}^{p}\frac{\left(\dfrac{\phi+\theta+h_it}{h_i}\right)^{n_i+r}\left(\lambda_i^{r+n_i-1}\right)e^{-\left(\frac{\phi+\theta+h_it}{h_i}\right)\lambda_i}}{\Gamma\left(n_i+r\right)}\right] \\
&\quad \times\frac{{}_0F_{p-1}\left([\,],[r],(\phi+\theta)^{p-1}\theta\prod_{i=1}^{p}\dfrac{\lambda_i}{h_i}\right)}{{}_pF_{p-1}\left([n_i+r],[r],(\phi+\theta)^{p-1}\theta\left[\prod_{i=1}^{p}\left(\dfrac{\phi+\theta}{\phi+\theta+h_it}\right)\right]\dfrac{\theta}{\phi+\theta}\right)}.
\end{aligned}
$$

Thus we can calculate the marginal posterior distributions and expectations of the rates. From the posterior marginal distribution we calculate the marginal posterior expectations. They are

$$
\mathrm{E}(\lambda_i\mid\boldsymbol{n})=\frac{1}{t}\left(1-\frac{\phi+\theta}{\phi+\theta+h_it}\right)\left[(n_i+r)\times\frac{{}_pF_{p-1}\left([n_i+r+1,n_j+r],[r],\dfrac{\theta}{\phi+\theta}\prod_{i=1}^{p}\dfrac{\phi+\theta}{\phi+\theta+h_it}\right)}{{}_pF_{p-1}\left([n_i+r],[r],\dfrac{\theta}{\phi+\theta}\prod_{i=1}^{p}\dfrac{\phi+\theta}{\phi+\theta+h_it}\right)}\right]. \tag{3}
$$

We express both the posterior densities and expectations in terms of standard functions. However, due to the nature of the generalised hypergeometric function, these quantities cannot be found analytically. Numerical procedures will be necessary to evaluate such quantities in practice.

It is also possible to observe multiple time periods for each event. Thus, instead of observing n_i occurrences of event i over time t we instead observe n_{ik} occurrences over time periods t_k for $k = 1, \ldots, q$. In terms of the full Bayes model above, this distinction is trivial as using Bayes Theorem to update beliefs is a fully coherent procedure. Thus making observations over multiple time periods does not alter the form of the posterior distribution. This may not be the case in other forms of subjective inference and so has to be considered more carefully.

2.1 **Properties of the Model**

Our full Bayes model has certain interesting properties, which we shall now derive allowing us to develop an intuitive understanding of how the correlations between the λ_i's affect the inference. We consider what happens to the posterior expectations of the event rates when $\theta \to 0$, which corresponds to the independent model, and when $\theta \to \infty$, which corresponds to perfect positive correlation between the rates. To do this we assume all $h_i = 1$.

Initially let $\theta \to 0$. In this case the ratio of the hypergeometric functions in (3) tends to 1 and so we have

$$\lim_{\theta \to 0} \mathrm{E}[\lambda_i \mid \boldsymbol{n}] = \frac{n_i + r}{\phi + t}.$$

Thus, using this model, as $\theta \to 0$, the posterior expectations of the rates tend to the posterior expectations we would obtain under the independent full Bayesian conjugate model.

We now examine what happens when $\theta \to \infty$. First we note that the expectation of λ_i can be re-expressed as

$$\frac{1}{\phi + t}\left(1 - \frac{\theta}{\phi + \theta + t}\right)\left[n_i + r + \left[\left(\prod_{j=1}^{p} \frac{\phi + \theta}{\phi + \theta + t}\right)\frac{\theta}{\phi + \theta}\right]\left[\frac{\prod_{j=1}^{p}(n_j + r)}{r^{p-1}}\right]\right.$$
$$\left. \times \frac{{}_pF_{p-1}\left([n_j + r + 1], [r + 1], \left[\prod_{j=1}^{p} \frac{\phi + \theta}{\phi + \theta + t}\right]\frac{\theta}{\phi + \theta}\right)}{{}_pF_{p-1}\left([n_j + r], [r], \left[\prod_{j=1}^{p} \frac{\phi + \theta}{\phi + \theta + t}\right]\frac{\theta}{\phi + \theta}\right)}\right]. \quad (4)$$

In order to develop an understanding of what happens when θ goes to infinity, we need to prove two properties of generalised hypergeometric functions. The first is in the form of a lemma.

Lemma 1: *If $n < p$ then*

$$\lim_{j \to \infty} (1 - x)\frac{{}_pF_n\left([a_i + 1 + j], [c_i + 1 + j], x\right)}{{}_pF_n\left([a_i + j], [c_i + j], x\right)} = 1.$$

Proof. The proof is given in Section 1 of the Supplementary Material.

We can then use this property in the proof of the following theorem.

THEOREM 1: *If $n = p-1$, then*

$$\lim_{x\to 1}(1-x)\frac{d\log\left({}_pF_n([a_i],[c_i],x^p)\right)}{dx} = \sum_{i=1}^{p} a_i - \sum_{i=1}^{p-1} c_i.$$

Proof. The proof of this theorem is also given in Section 1 of the Supplementary Material.

In our case $x = (\theta+\phi)/(\theta+\phi+t)$. We can also define $z = ((t+\phi)/t)x^p - (\phi/t)x^{p-1}$ allowing us to express the posterior expectation of λ_i as

$$\lim_{x\to 1}\left(\frac{1}{t}\right)\left[\left(\frac{t+\theta}{t}\right)x^p - \frac{\phi}{t}x^{p-1}\right]\frac{(1-x)}{(1-z)}(1-z)\times\frac{d\log({}_pF_{p-1}([n_j+r],[r],z))}{dz}.$$

By L'Hopital's rule we know that

$$\lim_{x\to 1}\frac{1-x}{1-z} = \lim_{x\to 1}\frac{t}{p(t+\phi)x^{p-1}-(p-1)\phi x^{p-2}} = \frac{t}{pt+\phi}.$$

By Proposition 1 we know that

$$\begin{aligned}\lim_{z\to 1}(1-z)\frac{d\log({}_pF_{p-1}([n_j+r],[r],z))}{dz} &= \sum_{j=1}^{p}(n_j+r)-(p-1)r \\ &= \lim_{x\to 1}(1-x)\frac{d\log({}_pF_{p-1}([n_j+r],[r],x))}{dx}.\end{aligned}$$

Therefore,

$$\begin{aligned}\lim_{x\to 1}\mathrm{E}(\lambda_i \mid \boldsymbol{n}) &= \lim_{x\to 1}\left(\frac{1}{t}\right)\left[\frac{(t+\phi)x^p-\phi x^{p-1}}{p(t+\phi)x^{p-1}-(p-1)\phi x^{p-2}}\right] \\ &\quad\times\left[(1-x)\frac{d\log({}_pF_{p-1}([n_j+r],[r],x))}{dx}\right] \\ &= \frac{\sum_{i=1}^{p} n_j + r}{pt+\phi},\end{aligned}$$

which corresponds to the posterior expectation of the rate of event i for the usual (independent) Bayesian model under the assumption that all of the observed number of events come from the same group. Thus in the extremes of weak and strong correlation the full Bayes model is well behaved.

We have established that the limit of the posterior mean for a rate, as the correlation between the rates approaches 0, i.e., $\rho = 0$, tends to the same estimate as obtained with an independent model. Further, as the correlation between the rates approaches 1, i.e., $\rho = 1$, the posterior estimate tends to the same as the one which we would obtain by treating the data as having been realised from the same underlying Poisson model.

The relationship between the posterior mean and the correlation coefficient however is not linear and can have local extrema in the interval for ρ between 0 and 1.

Figure 1 illustrates the relationship between the posterior mean and the correlation for a pool of two processes. Consider the posterior mean for one rate with the number of events realised ranging between 0 and 4 with a prior mean of 10 (i.e., $r = 10, \phi = 1, t = 1$), so that the posterior mean with the independent model ranges between 5 and 7. The second process in the pool is assumed to have realised 6 events and as such is also below the pool mean. The intercept on the vertical axis increases by 1/2, i.e. $1/(\phi + t)$, for each additional event realised by the process, but on the right side of the figure, such increments are half the magnitude as it increases the average of the pool by less.

The values for the illustration in Figure 1 were chosen to demonstrate the possibility of a local minimum for situations where the posterior mean ultimately decreases, remain the same and ultimately increases as correlation increases towards 1. That is, initially the estimator will head in the direction of the geometric mean of the pool but eventually converge to the arithmetic mean.

The initial effect on the posterior estimate of a rate having introduced correlation amongst the pool is to decrease (increase) the estimate if the geometric average of the posterior means of the other rates are below (above) the prior mean. As the correlation approaches 1 the posterior estimate will tend towards the arithmetic average of the pool as the data will be treated as exchangeable. This creates the prospect of local extrema for the posterior mean with respect to the correlation, which is described in Proposition 2. Such a situation creates the possibility that the estimates for all rates within a pool and as such the aggregate estimate of the rate of occurrence from the pool may initially decrease with the inclusion of correlation, although ultimately increase. These conditions highlight a need for careful investigation during a sensitivity analysis, for example, through calculation of credible intervals as the extremes are not at the limits of the parameter space.

The following theorem provides conditions for the existence of local minima.

THEOREM 2: *Assume a collection of m homogeneous Poisson processes whose rates of occurrence are realised from the multivariate gamma distribution specified in (1), where each pairwise comparison of rates as the same correlation $0 < \rho < 1$, then the posterior mean of the i'th process, i.e.* $\mathrm{E}[\lambda_i \mid \boldsymbol{n}]$, *has a local minimum with respect to ρ if the following conditions are met.*

(1) $\dfrac{r + n_i}{\phi + t} < \dfrac{r + \sum_{i=1}^{m} n_i}{\phi + mt}$,

(2) $\sqrt[m-1]{\prod_{j=1, j \neq i}^{m} \dfrac{r + n_j}{\phi + t}} < \dfrac{r}{\phi}$.

A local maximum exists if the inequalities in the two conditions are reversed.

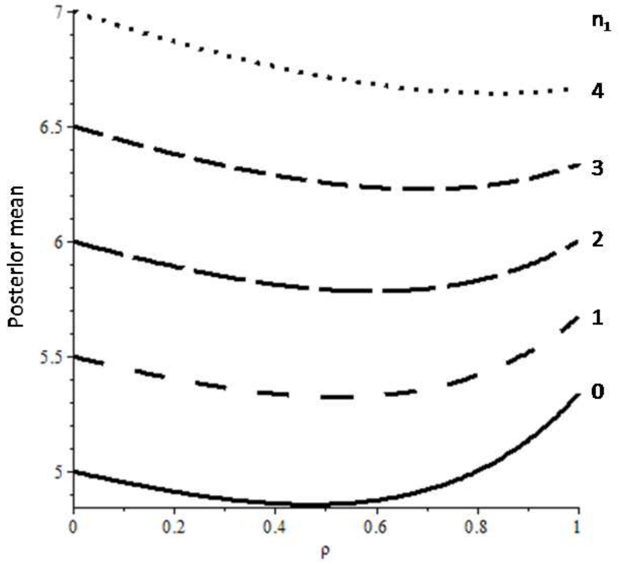

Fig. 1. The relationship between the posterior mean for a process as a function of the correlation in a pool of 2 illustrating the possibility of a local minima for situations where the posterior mean is ultimately increasing, remaining the same or decreasing.

Proof. The proof is given in Section 1 of the Supplementary Material.

3. THE BAYES LINEAR BAYES MODEL

In the previous section we considered a Bayesian model to estimate the unknown event rates. Updates were performed using a numerical procedure because the posterior distributions were not analytically tractable. In this section we consider a model which utilises Bayes linear methods. Bayes linear methods are, like Bayesian methodology, a form of statistical analysis which utilises subjective expert knowledge. Whereas a Bayesian analysis is based on subjective probability in the form of a probability distribution for the unknowns in the analysis, Bayes linear methods are instead based on expected values. Of course it is simple to move between the two concepts since a probability is just the expected value of an indicator variable.

In a Bayes linear analysis, for each unknown, X, an expectation $\mathrm{E}_0(X)$ and variance $\mathrm{Var}_0(X)$ is specified in the prior and, between every two unknowns, X_1 and X_2, a covariance $\mathrm{Cov}_0(X_1, X_2)$ is also given. This is known as a second order specification. Such a specification is adjusted on observation of data, not by Bayes Theorem as in a Bayesian analysis, but by minimising the expected squared loss between the unknowns and the observations. This gives the Bayes linear adjusted expectations and variances of the unknowns. More information on Bayes linear methods is given in Section 2 of the Supplementary Material.

Now, let N_{ik} denote the number of events of type i observed over time interval k, where the length of interval k is t_k. We can model this using the HPP

$$N_{ik} \mid \lambda_i \sim \mathrm{Po}(\lambda_i t_k),$$

as previously, for $i = 1, \dots, p$ and $k = 1, \dots, q$. The rates are assumed to have priors in the form of marginal gamma distributions which incorporate the use of homogenisation factors as in the previous section. That is,

$$\lambda_i \sim \mathrm{gamma}(r, \phi/h_i),$$

for parameters r, ϕ and homogenisation factor h_i. The prior expectation and variance of λ_i are

$$\mathrm{E}_0(\lambda_i) = \frac{rh_i}{\phi}, \quad \mathrm{Var}_0(\lambda_i) = \frac{rh_i^2}{\phi^2}.$$

Having specified the prior distributions in the form of specific values for r, ϕ, h_i and γ, we shall update the parameters using a mixture of full Bayesian and Bayes linear methods. This is known as a Bayes linear Bayes model[15].

However, in our case, we are not simply adjusting specifications directly as a result of observations as described for Bayes linear methods above. Rather, the expectation and variance of a single rate will change, using Bayes Theorem, as a result of of an observation on that rate. It is this change that we wish to propagate through to the other rates using Bayes linear methods. In the full probabilistic setting this situation is equivalent to that of probability kinematics, in which probabilities change over elements of a partition and these changes are then used to update the probabilities of some further events. The approach makes the assumption that the conditional probabilities of a future event B conditional on the elements of the partition $A_1, \dots, A_p$ do not change, i.e., $\mathrm{Pr}_1(B \mid A_i) = \mathrm{Pr}_0(B \mid A_i)$, for different specifications $\mathrm{Pr}_0(A_i)$ and $\mathrm{Pr}_1(A_i)$. The unconditional probability of B is then found in the usual way.

In a Bayes linear setting the equivalent assumption is that the adjusted expectation and variance of unknowns X do not change under different second order specifications $S_0(X)$ and $S_1(X)$. Imposing this condition leads to Bayes linear kinematics and provides a method for calculating the adjusted expectations of the unknowns. For more information on Bayes linear kinematics see Section 2 of the Supplementary Material. In our case, we update estimates within groups using Bayes Theorem as marginally the prior and likelihood are conjugate. Bayes linear kinematics is then used to update the estimates, given the full Bayes update in a certain group, in all of the other groups.

[16] discuss the application of Bayes linear kinematics to modelling failure rates and failure time distributions. They recommend performing updates on an unrestricted scale, i.e., $\mathbb{R}$ as opposed to $\mathbb{R}^+$, in order

that the linear fitting is being performed most appropriately. A discussion of the suitability and effectiveness of such transformations is given in Section 2.4.2 of[16]. Thus we shall also define

$$\eta_i = \log \lambda_i. \tag{5}$$

The prior mean and variance of η_i are

$$\mathrm{E}_0(\eta_i) = \psi(r) - \log \phi + \log h_i, \quad \mathrm{Var}_0(\eta_i) = \psi_1(r),$$

where $\psi(\cdot)$ and $\psi(\cdot)$ are the digamma and trigamma functions respectively. Derivations are found in[16].

Probability kinematics suffers from a lack of commutativity when successive updates of $\Pr(A_1), \ldots, \Pr(A_p)$ are made. That is, if updates are performed in different orders it is possible to get different answers for the final probability of B. Much work has been conducted to establish the conditions for when commutative solutions exist in probability kinematics - when updates performed in any order always give the same answer[22,23,24]. Multiple Bayes linear kinematic updates are also not necessarily commutative.[15] give necessary and sufficient conditions for a unique, commutative Bayes linear solution to exist. We utilise these conditions later when we update the rates on observation of the numbers of events. The transformation taken above will be helpful in providing a commutative solution.

First, however, consider the specification of prior correlations between the rates.

3.1 Specifying a Prior Covariance Structure

We require a prior covariance structure between all of the unrestricted parameters η_i, $i = 1, \ldots, p$. **The specification Bayes linear relationships between the transformed parameters means that the only restriction on the correlations which can be specified is the usual one that the correlation matrix must be positive semi-definite.** How to specify this will vary from problem to problem. However, below we give a brief indication of how this might be tackled in general.

In the model given in Section 2 all of the correlations between the rates are equal. This means that

$$\mathrm{Corr}_0(\lambda_i, \lambda_j) = \rho = \frac{\theta}{\phi + \theta}, \tag{6}$$

for correlation parameter $0 < \rho < 1$. We use this form in all examples in this paper. It will not always be reasonable to assume an exchangeable structure like this. In particular, we may believe that the correlations between rates depend on the degree of similarity of the events in question, with more similar events given higher correlations than very different events.

[25] discusses general methods for specifying covariance structures based on the idea of uncertainty factors, zero expectation quantities which can be common to many rates and serve to reduce the number of number of elicitations required to specify a full correlation structure in complex problems. Specific examples are given in the paper to show how uncertainty factors can be used in practice.

Of course it is important when specifying correlations to do so in such a way as to ensure that the resulting correlation matrix is positive definite. One way to do so utilises the structure of Markov Trees[26]. We introduce a latent rate λ_L, such that,

$$\lambda_L \sim \text{gamma}\left(r, \frac{\phi}{\bar{h}}\right),$$

where $\bar{h} = \frac{1}{p}\sum_i h_i$. That is, λ_L represents an average event rate. If we assume that λ_i, λ_j are independent conditional on λ_L for $i \neq j$, then we can construct correlations between the event rates by specifying $\text{Corr}_0(\lambda_L, \lambda_i) = \alpha_i$. If we specify each $\alpha_i = \alpha$ then this results in constant correlations as in (6) with $\text{Corr}(\lambda_i, \lambda_j) \approx \alpha^2$. Specifying non-constant α_i's results in non-constant correlations between the rates. For example, the correlations between rates could depend on h_L/h_i Defining the correlations using this structure ensures that the resulting correlation matrix is positive definite.

3.2 **Bayes Linear Kinematic Updating**

Suppose we have a full second order prior specification for $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)$ and so can update the estimates of the rates using Bayes Theorem and Bayes linear kinematics. Further, suppose that we observe each count for event i, that is $N_{ik} = n_{ik}$ over time periods t_k, $k = 1, \ldots, q$. Then, through conjugate Bayesian updating, the rate of event i becomes $\lambda_i \mid n_{i1}, \ldots, n_{iq} \sim \text{gamma}(r + \sum_k n_{ik}, (\phi/h_i) + \sum_k t_k)$. Thus the mean and variance of η_i become

$$\text{E}(\eta_i \mid \boldsymbol{n}_i) = \psi\left(r + \sum_k n_{ik}\right) - \log\left(\phi/h_i + \sum_k t_k\right),$$
$$\text{Var}(\eta_i \mid \boldsymbol{n}_i) = \psi_1\left(r + \sum_k n_{ik}\right),$$

where $\boldsymbol{n}_i = (n_{i1}, \ldots, n_{iq})$. We can use Bayes linear kinematics to update our beliefs about all of the other transformed rates given observation of $\boldsymbol{n}_i$. Using the formula from Appendix B, the adjusted expectation and variance of $\boldsymbol{\eta}$ are

$$\begin{aligned}
\mathrm{E}_1(\boldsymbol{\eta};\boldsymbol{n}_i) &= \mathrm{E}_0(\boldsymbol{\eta}) + \mathrm{Cov}_0(\boldsymbol{\eta},\eta_i)\mathrm{Var}_0^{-1}(\eta_i)\left[\mathrm{E}(\eta_i \mid \boldsymbol{n}_i) - \mathrm{E}_0(\eta_i)\right], \\
\mathrm{Var}_1(\boldsymbol{\eta};\boldsymbol{n}_i) &= \mathrm{Var}_0(\boldsymbol{\eta}) - \mathrm{Cov}_0(\boldsymbol{\eta},\eta_i)\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Cov}_0(\eta_i,\boldsymbol{\eta}) \\
&\quad + \mathrm{Cov}_0(\boldsymbol{\eta},\eta_i)\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Var}(\eta_i \mid \boldsymbol{n}_i)\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Cov}_0(\eta_i,\boldsymbol{\eta}).
\end{aligned}$$

One update of this kind is made for each i. A sufficient condition for a unique, commutative solution to exist for this to case, requires satisfying the general conditions of[15], $\psi_1(r + \sum_k n_{ik}) < \psi_1(r)$ for some i. Clearly this shall always be true because the trigamma function is monotonically decreasing on the positive real line.

Thus we have a unique, commutative solution and, from Section 2 of the Supplementary Material, the commutative adjusted expectation and variance are

$$\begin{aligned}
\mathrm{Var}_p^{-1}(\boldsymbol{\eta};\boldsymbol{n})\mathrm{E}_p^{-1}(\boldsymbol{\eta};\boldsymbol{n}) &= \sum_{i=1}^{p} \mathrm{Var}_1^{-1}(\boldsymbol{\eta};\boldsymbol{n}_i)\mathrm{E}_1(\boldsymbol{\eta};\boldsymbol{n}_i) - (p-1)\mathrm{Var}_0^{-1}(\boldsymbol{\eta})\mathrm{E}_0(\boldsymbol{\eta}), \\
\mathrm{Var}_p^{-1}(\boldsymbol{\eta};\boldsymbol{n}) &= \sum_{i=1}^{p} \mathrm{Var}_1^{-1}(\boldsymbol{\eta};\boldsymbol{n}_i) - (p-1)\mathrm{Var}_0^{-1}(\boldsymbol{\eta}),
\end{aligned}$$

for $\boldsymbol{n} = (\boldsymbol{n}_1,\ldots,\boldsymbol{n}_p)$. A simple way to recover estimates of the rates on the correct scale is to reverse the transformation given in (5). If we make the assumption that the posterior marginal distributions for the rates are still of the form of gamma distributions, that is, $\lambda_i;\boldsymbol{n} \sim \mathrm{gamma}(R,\Phi)$, then the posterior parameters R,Φ can be recovered from

$$\mathrm{E}(\eta_i;\boldsymbol{n}) = \psi(R) - \log\Phi, \quad \mathrm{Var}_p(\eta_i;\boldsymbol{n}) = \psi_1(R)$$

Thus using this procedure all updates are made using analytic formulae. Consequently, obtaining posterior quantities can be achieved more quickly than when using full Bayes models which require numerical or simulation methods. In particular, for the full Bayesian approach the number of computations required is dominated by the hypergeometric functions and hence exponential in p whereas in the Bayes linear Bayes model it is dominated by the matrix inversion and so $o(p^3)$. However it is important to investigate whether this procedure also gives estimates of event rates similar to those of the full Bayes model of Section 2. We shall explore this further in Sections 5 and 6.

4. PRIOR SPECIFICATION USING EMPIRICAL BAYES

So far we have considered the specification of the prior as a subjective exercise in which observable quantities are to be elicited from experts and then used to provide values for r, ϕ, ρ, and $h_1,\ldots,h_p$. In practice, however, there may be occasions where problem experts may be either unwilling or unable to fully specify all of these values. For example, when we wish to estimate the rates of many rare events, the prior will tend to dominate the inference. Therefore those with an interest in the analysis may be reluctant to provide expert judgements to specify the prior, especially in situations in which estimates will be in the public domain or used to inform public safety policy and so there is a need for seemingly defensibly objective methodologies.

Empirical Bayes offers a procedure to empirically estimate the parameters in the prior distribution or the Bayes linear parameters. For our model, to provide empirical Bayes estimates, the events generated by each of the Poisson processes are pooled and an estimate of the pooled event rate can be obtained. The accuracy of empirical Bayes estimates depends on the degree of homogeneity of the pool. The elicited homogenisation factors increase the effectiveness of the pooling process and hence the accuracy of the estimates obtained.

We can thus use empirical Bayes to estimate the prior values of r, ϕ, ρ. We could estimate the parameters r, ϕ using maximum likelihood, however this proved to be analytically intractable in the independent model of[(5)]. Instead they used the method of moments[(27)]. Whilst utilising a numerical procedure to evaluate the maximum likelihood estimates is possible, see[(20)], the method of moments produces consistent estimates quickly and without computational burden.

In order to use the method of moments we require the first two moments of N_{ik}, which are given by

$$\begin{aligned}
\mathrm{E}(N_{ik}) &= \mathrm{E}_{\lambda_i}\left[\mathrm{E}(N_{ik} \mid \lambda_i)\right] \\
&= \frac{rh_i}{\phi} t_k, && (7) \\
\mathrm{E}(N_{ik}^2) &= \mathrm{E}_{\lambda_i}\left[\mathrm{E}(N_{ik}^2 \mid \lambda_i)\right] \\
&= \frac{rh_i}{\phi} t_k + \left(\frac{rh_i^2}{\phi^2} + \frac{r^2 h_i^2}{\phi^2}\right) t_k^2. && (8)
\end{aligned}$$

We also wish to estimate ρ. For $i \neq j$, the cross moment between N_{ik} and N_{jk} is

$$\begin{aligned}
\mathrm{E}(N_{ik}N_{jk}) &= \mathrm{E}_{\lambda_i,\lambda_j}\left[\mathrm{E}(N_{ik}N_{jk} \mid \lambda_i, \lambda_j)\right] \\
&= \rho \frac{r}{\phi^2} h_i h_j t_k^2 + \frac{r^2}{\phi^2} h_i h_j t_k^2.
\end{aligned}$$

We can now use the method of moments to estimate r, ϕ and ρ. Doing so gives the empirical Bayes estimates

$$\hat{r} = \frac{U^2}{V - U^2}, \quad \hat{\phi} = \frac{U}{V - U^2} \quad \hat{\rho} = \frac{W - U^2}{V - U^2},$$

where

$$U = \frac{\sum_i \sum_k N_{ik}}{\sum_i \sum_k h_i t_k}, \quad V = \frac{\sum_i \sum_k N_{ik}^2 - \sum_i \sum_k N_{ik}}{\sum_i \sum_k (h_i t_k)^2},$$

and

$$W = \frac{\sum_i \sum_{j \neq i} \sum_k N_{ik} N_{jk}}{\sum_i \sum_{j \neq i} \sum_k h_i h_j t_k^2}.$$

The estimators $\hat{r}, \hat{\phi}$ and $\hat{\rho}$ are all unbiased and each is also a consistent estimator as $p \to \infty$. A proof of this is given in Section 3 of the Supplementary Material.

5. SIMULATION STUDY TO COMPARE METHODS

To examine the relative accuracy of estimates obtained using the Bayes linear Bayes method in Section 3 compared with the full Bayesian model described in Section 2, we report the results from a simulation experiment. We use both approaches in conjunction with the empirical Bayes prior estimates developed in Section 4. However, having made the prior specifications using the empirical approach, all of the comparisons between the two models are valid independently of the approach used to set the priors. That is, the results are equally valid for subjective priors specified using expert elicitation, **under the assumption that the quantities, such as quantiles, elicited from experts are consistent with the gamma distribution for some parameter values**.

We simulate data from the full Bayesian model with known parameter values. This will allow us to see how well we are recovering the true model in each case and how well the Bayes linear Bayes model is approximating the full Bayes model.

Let us consider just two events with rates λ_1 and λ_2. Further suppose that we have observed realisations of both events over 100 time periods each of length 1 time units. That is, we simulate 100 pairs of counts (N_{1k}, N_{2k}). The assumed homogenisation factors are $(h_1, h_2) = (1, 10)$ and the parameter values used in the simulation are

$$r = 1, \phi = 3, \rho = 0.2.$$

Using the procedure outlined in Section 4 we obtain the empirical Bayes estimates of the parameters from a single simulation. They are

$$\hat{r} = 1.022, \hat{\phi} = 2.982, \hat{\rho} = 0.288.$$

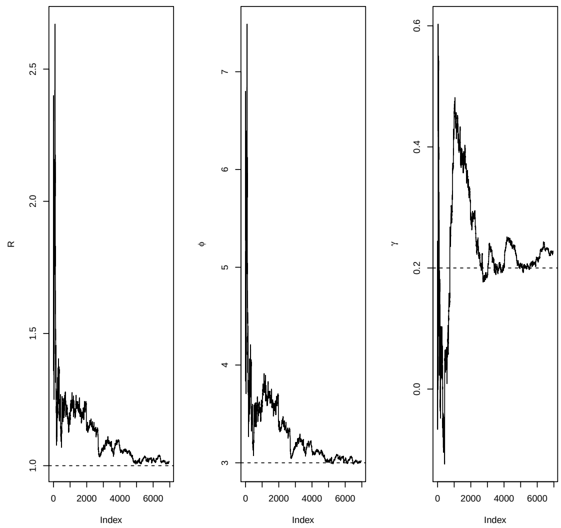

Fig. 2. The empirical Bayes parameter estimates for the simulation example resulting from different numbers of simulated observations.

Thus, for a reasonable sample size, the empirical Bayes estimation procedure appears to provide reasonable estimates of the true parameter values in the prior. Figure 2 shows the parameter values found using empirical Bayes under different numbers of observations from the same sample. We see that each of the parameters converges towards its correct value with large numbers of observations, but estimates are still relatively good for smaller numbers of observations in this sample.

Estimates are updated using the Bayes linear Bayes model to obtain the adjusted expectations for each of the rates as

$$\mathrm{E}_2(\lambda_1; \boldsymbol{n}) = 0.389, \quad \mathrm{E}_2(\lambda_2; \boldsymbol{n}) = 3.382.$$

The "approximate" posterior expectations can be compared to the "exact" posterior expectations found using the full Bayesian posterior distribution in (3). For our assumed data set, we find the "exact" posterior expectations to be

$$\mathrm{E}(\lambda_1 \mid \boldsymbol{n}) = 0.387, \quad \mathrm{E}(\lambda_2 \mid \boldsymbol{n}) = 3.360.$$

For this particular example, we obtain a close match between the "approximate" and "exact" values.

These are, of course, just results for a single simulation of 100 pairs of observations with a single set of parameter values. In order to obtain greater insight into how the models compare we have simulated 100 pairs of samples for each of three sets of parameters values $(r, \phi) =$

Table I . The differences between the posterior expectations from the full Bayes and Bayes linear Bayes methods, $\mathrm{E}(\lambda_i \mid \boldsymbol{n}) - \mathrm{E}_2(\lambda_i; \boldsymbol{n})$, given the same priors under a number of different parameter values.

		Group 1		Group 2	
(r, ϕ)	ρ	Stream 1	Stream 2	Stream 1	Stream 2
(1,3)	0.2	0.00722	0.00239	0.00724	0.00117
	0.4	-0.00129	-0.00221	-0.00590	-0.00249
	0.6	-0.00124	-0.00494	-0.00296	-0.00350
	0.8	-0.0231	-0.0144	-0.019	-0.0199
(2,2)	0.2	0.00706	-0.00887	0.00721	-0.00839
	0.4	0.00667	-0.0124	0.00690	-0.0128
	0.6	-0.0104	-0.00544	-0.0113	-0.00516
	0.8	-0.0276	-0.00137	-0.0298	-0.00223
(3,1)	0.2	-0.00427	-0.00338	-0.0529	-0.0384
	0.4	0.00056	0.00425	-0.0321	-0.228
	0.6	-0.0806	-0.0192	-0.0966	-0.0378
	0.8	-0.0418	-0.0318	-0.0576	-0.0965

$(1,3),(2,2),(3,1)$ **at four different correlations** $\rho = (0.2, 0.4, 0.6, 0.8)$ **using homogenisation factors as above. We repeat each simulation twice, and call the two runs of the simulation at the same parameter values two different "streams". The results given in Table I show the differences between the posterior expectations resulting from the two methods,** $\mathrm{E}(\lambda_i \mid \boldsymbol{n}) - \mathrm{E}_2(\lambda_i; \boldsymbol{n})$**.**

The results in the table suggest that the methods overall are performing very similarly over the range of different marginal parameter values and correlations. The relative differences between the estimates from two methods are almost never above 4% and are typically below 1%. The majority of the differences in the table are negative indicating that in the simulations, the posterior expectation from the Bayes linear Bayes model was higher than that of the full Bayes model. However, there are some occasions when the full Bayes model results in a higher estimate.

The posterior expectations from the two methods can also be compared for different numbers of observed intervals (sample sizes) for some chosen parameter values and correlations. Figure 3 shows the differences between the expectations of λ_1 and λ_2 under the two procedures for sample sizes between 1 and 200 for each of the two event rates using $(r, \phi) = (1, 3)$ and $\rho = 0.2$. Our results show that with increasing sample size, the difference between the estimates from the Bayes linear Bayes and full Bayes approaches becomes smaller. Even for small sample sizes the differences are small. We perform a further simulation, as before, to find the differences between the full Bayes and Bayes linear Bayes estimates of the posterior means of the rates but this time using $\rho = 0.6$. Plots for λ_1 and λ_2 are given in Figure 4. As previously, even for small sample sizes the Bayes linear Bayes and full Bayes methods are still producing results which are close. Both sets of plots suggest that the pattern of the differences between the means for the two methods are similar for λ_1 and λ_2. This is due to the fact that the same simulated data were used for (a) and (b) and similarly for (c) and (d).

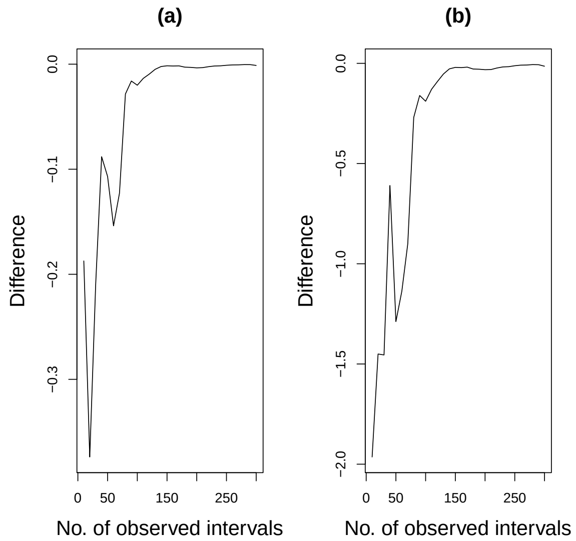

Fig. 3. The difference between the full Bayes and Bayes linear Bayes posterior expectations of λ_1 in (a) and λ_2 in (b), $\mathrm{E}(\lambda_i \mid \boldsymbol{n}) - \mathrm{E}_2(\lambda_i; \boldsymbol{n})$, for different sample sizes, with $\rho = 0.2$.

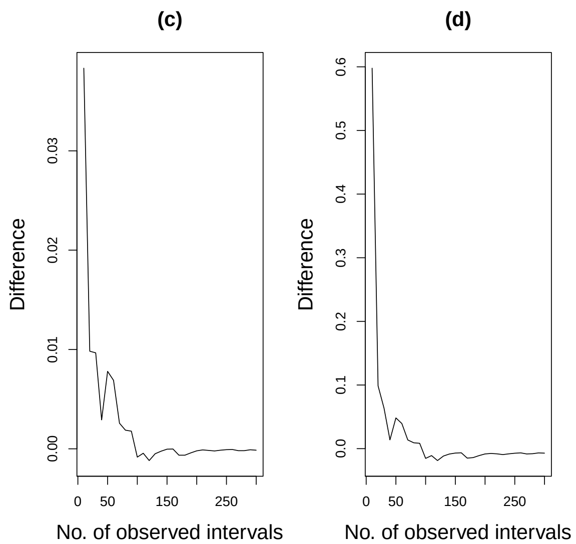

Fig. 4. The difference between the full Bayes and Bayes linear Bayes posterior expectations of λ_1 in (c) and λ_2 in (d), $\mathrm{E}(\lambda_i \mid \boldsymbol{n}) - \mathrm{E}_2(\lambda_i; \boldsymbol{n})$, for different sample sizes, with $\rho = 0.6$.

We can also compare the estimates resulting from the Bayes linear Bayes inference to the observed rates given sample sizes for a particular simulation. A plot of these quantities for λ_1 for sample sizes between 10 and 50 are given in Figure 5. The dashed line in the figure is the theoretical rate used for the simulation, $r/\phi = 0.33$. The plot shows that the posterior expectations under our inference procedure become closer to the observed rate with increasing sample size. **For smaller sample sizes the estimated rate from the Bayes**

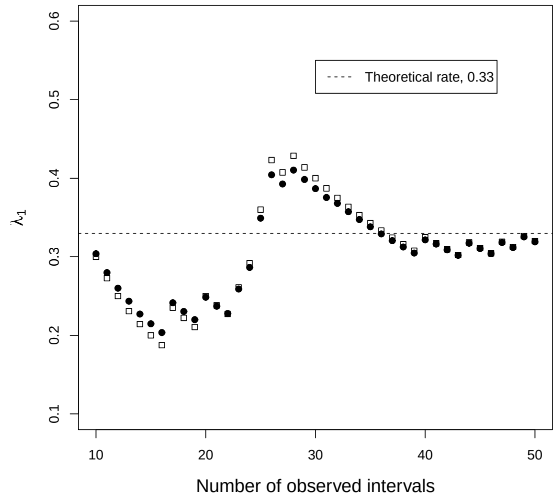

Fig. 5. The adjusted expectation of λ_1, $\mathrm{E}_2(\lambda_1;\boldsymbol{n})$ (circles ●) and observed rate of λ_1 (squares □) for different sample sizes.

linear Bayes model tends to be less extreme than the observed rate, both when the obseved rate is much higher and much lower than the theoretical rate. This is partly as a result of the correlation in the model. It is also apparent that both the data and the Bayes linear Bayes estimate are converging towards the correct rate.

6. ILLUSTRATIVE EXAMPLE

6.1 Background

A large engineering firm designs and manufactures electronic systems that are supplied to customers worldwide. For one high-valued product, the firm requires estimates of the failure rates experienced by each customer for different failure modes. There are over 2500 customers of the firm, each of whom have very different operational usage patterns. It is believed that these usage patterns are dominant over, for example, environmental conditions. Data are available for the counts of number of events experienced by each customer for a 5 year period together with the annual operating usage figures. Events related to the two major failure modes, labelled A and B, are identified and it is anticipated that their rates of occurrence will not be statistically independent due to the characteristics of the failure modes under different operating stresses. Further, it is considered reasonable to assume that the rates of events are constant given the stage of the product lifecycle.

While this example is motivated by a real problem, we have de-sensitised the data and so all results are indicative. However the example allows us to illustrate the proposed method and to understand the value of capturing the correlation between events.

6.2 The Model

For each customer, we observe the numbers of failures for both modes A and B yearly together with the annual usage figures, which represents our exposure time. That is, we have,

$$N_{ik} \mid \lambda_{ik} \sim \mathrm{Po}(\lambda_{ik} t_i),$$

where N_{ik} represents the number of failures of mode k for customer i, λ_{ik} is the rate of failures and t_i is the usage by the customer. Given that we believe correlations may exist between the two failure modes for each customer, then, λ_{ik} is not independent of λ_{jl} for $j \neq l$. However, it is unlikely that there will be the same number of events for each failure mode. Therefore the priors proposed for the rates are

$$\lambda_{ik} \sim \mathrm{gamma}(r, \phi/h_k),$$

for homogenisation factors h_k associated with each failure mode. It is further assumed that failure rates between different customers are statistically independent.

6.3 Failure Rate Estimation

We choose to specify the prior empirically using the approach described in Section 4. We obtain

$$r = 2.24,\ \phi = 6.13,\ \rho = 0.57,$$

and so establish that correlation between failure modes is evident. Through a structured process, we elicit subjective assessments of the homogenisation factors and hence set these to be $h = (1, 2)$ for failure modes A and B respectively. We can use these prior specifications to compare both the effect of including correlation in the modelling process and estimates obtained under the Bayes linear Bayes inference method with that under the full Bayes model.

Figure 6 gives the ratios of the posterior means of the failure rates for each customer under the full Bayes model to those found under the independent model using the same prior parameter values (but with $\rho = 0$). It is shown that the correlation between different failure modes for each customer is having an effect. Virtually all of the estimated rates under the correlated model are below those of the independent model for both failure

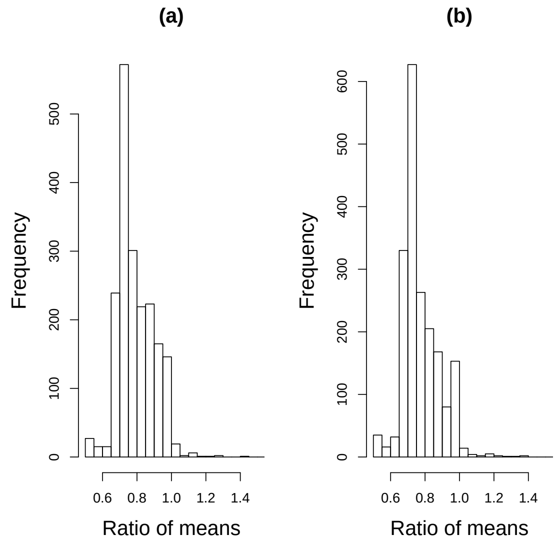

Fig. 6. Histogram of ratios of posterior means for each customer using the full Bayes correlated and independent models for (a) failure mode A and (b) failure mode B.

mode A in (a), and failure mode B, in (b). The mean of the correlated rates is 77.0% that of the independent model for failure mode A events and 75.1% for failure mode B events suggesting it is important to include correlations in the model.

We also compare the Bayes linear Bayes model posterior means to both the independent and full Bayes models. Figure 7 gives a plot of the ratios of the posterior expectations of the full Bayes model to the Bayes linear Bayes model. In this case, the ratio is more evenly spread around one. The modes for both failure modes A and B are larger than 1, suggesting that, on average, the full Bayes estimates are slightly lower.

A comparison of the Bayes linear and independent models provides a better insight into how the Bayes linear Bayes model is performing. A plot of the posterior ratios for all customers is given in Figure 8. A similar picture emerges to that found on comparing the full Bayes model with the independent model. The Bayes linear model typically provides lower estimates than the independent model.

A more comprehensive comparison of the Bayes linear Bayes model in relation to the fully Bayesian model is obtained by plotting the posterior means of the rates for the two methods against each other as in Figure 9. This figure indicates that the two methods are actually producing results which are very similar, in real terms, for all customers. Hence, the Bayes linear Bayes model provides a good approximation to the full Bayesian model for these data.

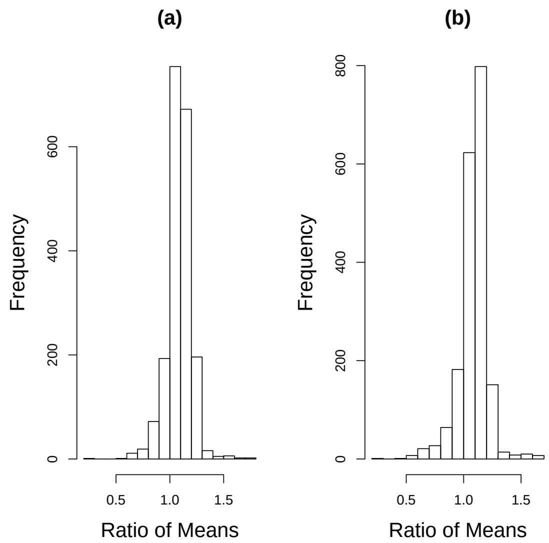

Fig. 7. Histogram of ratios of posterior means for each customer using the full Bayes correlated and Bayes linear Bayes models for (a) failure mode A and (b) failure mode B.

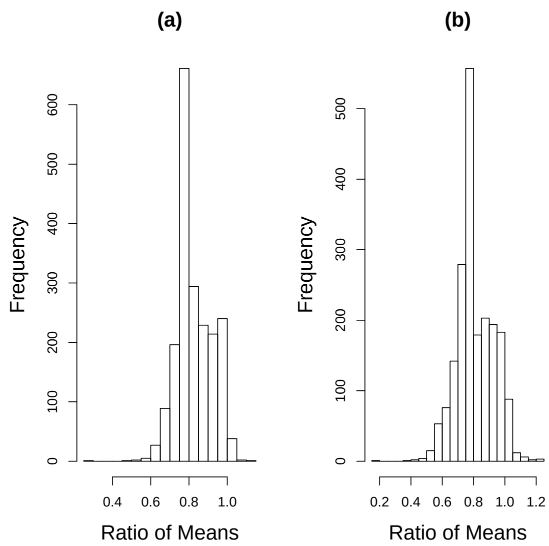

Fig. 8. Histogram of ratios of posterior means for each customer using the independent and Bayes linear Bayes models for (a) failure mode A and (b) failure mode B.

7. CONCLUSIONS AND FURTHER WORK

A new method of Bayesian inference is developed to estimate event rates which are considered correlated. The approach we have taken is subjective Bayesian, although guidance has also been given on an empirical Bayes method for specifying prior parameter values as we recognise that in many real life problems experts may be reluctant to include such explicit subjective information in their estimates.

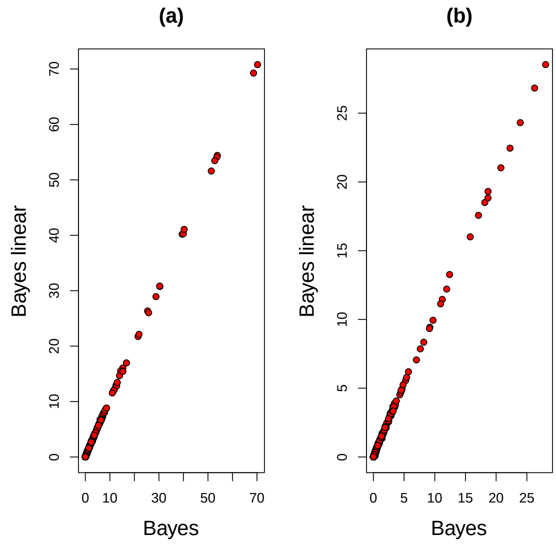

Fig. 9. The posterior means for the rates of (a) failure mode A and (b) failure mode B for each customer for full Bayes and Bayes linear Bayes models.

For our problem of correlated rates of events realised from a homogeneous Poisson process, we first considered a full Bayesian solution to the problem. We have derived the posterior and predictive distributions, as well as the posterior expectations, for the model in terms of hypergeometric functions. We have also showed that the model has desirable properties as correlation goes to 0 and 1. For this model it has been necessary to calculate posterior expectations numerically.

A new method based on a combined tractable full Bayes update for data within a group with Bayes linear kinematic updates between groups has been developed. We have found that our new method is relatively fast and efficient to implement, in contrast to the intensive calculations in the full Bayesian model. For our particular problem, using the full Bayesian approach the number of computations required is dominated by the hypergeometric functions and hence exponential in p whereas in the Bayes linear Bayes model it is dominated by the matrix inversion and so $o(p^3)$. Thus, when the estimation of the rates is part of a larger model, or the number of rates in the study and the number of observations is very large, the Bayes linear Bayes method could offer a solution when full Bayes methods do not.

We have evaluated the new Bayes linear Bayes model in comparison with the full Bayes model via a simulation study in which we simulate data from the full Bayes model with known parameter values and use an empirical prior. Our findings indicate that the empirical Bayes procedure provides accurate estimates of all three prior parameters for a sample size of 100 in each group. As sample size increases the inferences from both models were converging, with the models showing good agreement for even fairly small sample sizes.

Future work can progress in various directions. In our study we have used an empirical prior in our simulations and illustrative example even though we initially introduced the models with a subjective prior. A structured method for eliciting a subjective prior remains to be developed. We have assumed particular forms of the underlying point process and the nature of the correlation structure. We can relax the assumption that the rate of events is constant and hence consider, for example, events generated by non-homogeneous Poisson processes. In a reliability context such processes would allow us to capture growth or decay in the rate of events. Similarly, more complex correlations are inherent in some of the practical examples we discuss to motivate this work. Consider, for example, the temporal dependencies and more complex correlations we might expect in supply, transportation or power networks. While some of these types of problems could have an empirical Bayes solution based on current theory, there is future research to obtain a full Bayesian model.

REFERENCES

1. S. Ali Agha, R. van der Mear, and L. Walls. Modelling supply network reliability and reliance using dynamic Bayesian networks. *Proceedings of the Hamburg International Conference on Logistics*, 2012.
2. Annual Safety Performance Report. *Annual Safety Performance Report 2009/10: A reference guide to safety trends on GB railways*. Rail Safety and Standards Board, 2010.
3. G. Becker and B. Schubert. ZEDB - a practical application of a 2-stage Bayesian analysis code. In S. Lyderson, G.K. Hansen, and H.A. Sandtorv, editors, *ESREL 1998 conference: Safety and reliability*, volume 1, pages 487–494. 1998.
4. T-book. *Reliability Data of Components in Nordic Nuclear Power Plants*. Prepared by the TUD Office, Studsvik Eco and Safety AB and PSrn Consulting. Published by the ATV Office, Vattenfall AB, 1994.
5. J. Quigley, G. Hardman, T. Bedford, and L. Walls. Merging expert and empirical data for rare event frequency estimation: Pool homogenisation for empirical Bayes models. *Reliability Engineering and System Safety*, 96(6):687–695, 2011.
6. P. Congdon. *Bayesian Statistical Modelling*. John Wiley and Sons, Chichester, UK, 2001.
7. R. B. Nelson. *An introduction to copulas; second edition*. Springer, 2006.
8. S. Kaplan. Two-stage Poisson-type problem in probabilistic risk analysis. *Risk Analysis*, 5:227–230, 1985.
9. R. Iman and S. Hora. Bayesian methods of modeling recovery times with applications to the loss of off-site power at nuclear power plants. *Risk Analysis*, 9:25–36, 1989.
10. E. Hofer, S. Hora, R. Iman, and J. Peschke. On the solution approach for Bayesian modelling of initiating event frequencies and failure rates. *Risk Analysis*, 17:249–252, 1997.
11. K. Pörn. *On empirical Bayesian inference applied to Poisson probability models*. PhD thesis, Linkoping, 1990.
12. C. Bunea, C. Charitos, R.M. Cooke, and G. Becker. Two stage Bayesian models application to ZEDB project. *Reliability Engineering and System Safety*, 90:123–130, 2005.
13. P. Müller. Simulation-based optimal design. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 6*, pages 459–474. Oxford University Press, 1999.
14. M. Goldstein and D. Wooff. *Bayes Linear Statistics*. John Wiley and Sons, Chichester, UK, 2007.
15. M. Goldstein and S. Shaw. Bayes linear kinematics and Bayes linear Bayes graphical models. *Biometrika*, 91:425–446, 2004.

16. 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.
17. M. Goldstein and T. Bedford. The Bayes linear approach to inference and decision-making for a reliability programme. *Reliabiliy Engineering and System Safety*, 92:1344–1352, 2007.
18. M.J.T. Revie, T.Bedford, and L. Walls. Supporting reliability decisions during defense procurement using a Bayes linear methodology. *IEEE Transactions on Engineering Management*, 58:662 – 673, 2011.
19. B.P. Carlin and T.A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis (2nd ed.)*. Chapman and Hall, 2000.
20. J. Quigley, T. Bedford, and L. Walls. Estimating rate of occurrence of rare events with empirical Bayes: A railway application. *Reliability Engineering and System Safety*, 92(5):619–627, 2007.
21. A.T.M. Minhajuddin, I.R. Harris, and W.R. Schucancy. Simulating multivariate distributions with specific correlations. *Journal of Statistical Computing and Simulation*, 74, 2004.
22. H. Field. A note on Jeffrey conditionalization. *Philosophy of Science*, 45:361–367, 1978.
23. P. Diaconis and S. L. Zabell. Updating subjective probability. *Journal of the American Statistical Association*, 77:822–830, 1982.
24. F. Doring. Why bayesian psychology is incomplete. *Philosophy of Science*, 66:S379–S389, 1999.
25. M. Farrow. Practical building of subjective covariance structures for large complicatesd systems. *The Statistician*, 52:553–573, 2003.
26. T. Bedford and R. M. Cooke. Vines: a new graphical model for dependent random variables. *The Annals of Statistics*, 30:1031–1068, 2002.
27. S. F. Arnold. *Mathematical Statistics*. Prentice-Hall, New Jersey, 1990.