Chapter 7
Hierarchical Bayes Models for Variability

This chapter discusses the Bayesian framework for expanding common likelihood functions introduced in earlier chapters to include additional variability. This variability can be over time, among sources, etc.

7.1 Variability in Trend Models

Chapter 5 introduced problems in which an aleatory model parameter such as p or λ is allowed to vary monotonically over time. Consider the example data in Table 7.1. If we assume a Poisson aleatory model for the number of events in each age period, with parameter $\lambda_i t_i$, and update the Jeffreys prior for λ with the event count for each age period, we get the caterpillar plot shown in Fig. 7.1, which suggests that λ may be decreasing with time.

If we model the apparent time trend in λ using the loglinear model introduced in Chap. 5, with flat prior distributions on the loglinear coefficients a and b, we find that the marginal posterior distribution for the slope parameter, b, lies almost entirely below zero, appearing to indicate a decreasing trend in λ, in agreement with Fig. 7.1. However, if we estimate the Bayesian p-value for such a model as in Chaps. 4 and 5, we find a value of about 0.009, a quite small value, suggesting that the loglinear model for λ is not very good at replicating the observed data. The plot of the predicted event count for each age bin under this model (Fig. 7.2) bears out the low Bayesian p-value. As can be seen in Fig. 7.2, a number of observed event counts are at the limit (especially the lower limit) of the 95% credible interval for the replicated event count, suggesting that the loglinear model has some difficulty replicating the observed data.

In an attempt to explore a richer model that could give rise to the data in Table 7.1, let us consider an extension to the loglinear model, in which there is still a monotonic trend with age, but with additional random variability in λ over time. Our expanded trend model becomes

D. Kelly and C. Smith, *Bayesian Inference for Probabilistic Risk Assessment*, Springer Series in Reliability Engineering, DOI: 10.1007/978-1-84996-187-5_7,

Table 7.1 Example data showing apparent decreasing trend over time in λ

x	Age	Exposure time
19	0.5	0.71
3	1.5	0.75
6	2.5	0.79
0	3.5	0.90
1	4.5	0.82
3	5.5	0.75
0	6.5	1.00
0	7.5	0.83
2	8.5	0.84

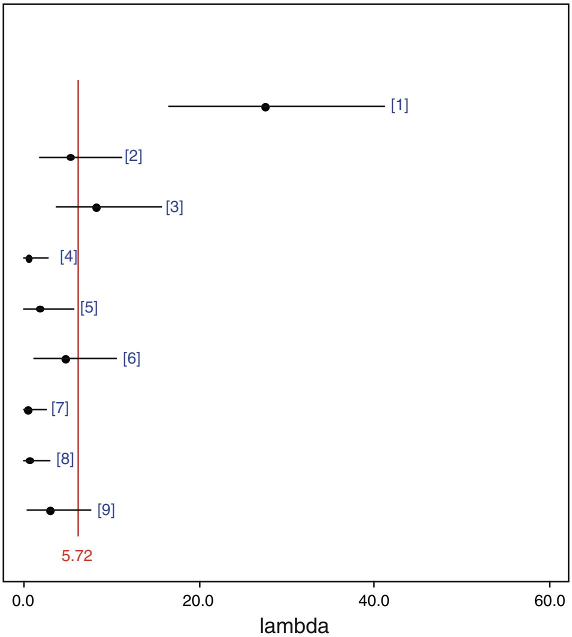


Fig. 7.1 Caterpillar plot showing apparent decreasing trend in λ over time

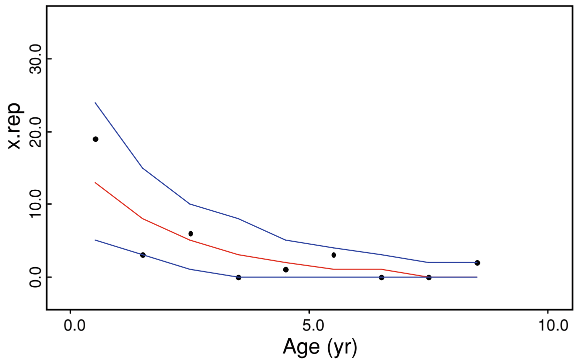


Fig. 7.2 Plot of 95% credible interval (*blue lines*) for predicted event count under loglinear model. Dots are observed event counts

Table 7.2 OpenBUGS script for loglinear model for λ with extra-Poisson variation in each time period

```
model
{
for (i in 1:N) {
        log(lambda[i]) <- a + b*age[i] + eps[i]
        eps[i] ~ dnorm(0, tau.eps)
#Poisson aleatory model and predictive distributions for model checking
        mu[i] <- lambda[i]*time[i]
        x[i] ~ dpois(mu[i])
        x.rep[i] ~ dpois(mu[i])
        diff.obs[i] <- pow(x[i] - mu[i], 2)/mu[i]
        diff.rep[i] <- pow(x.rep[i] - mu[i], 2)/mu[i]
        }
#Model checking
chisq.obs <- sum(diff.obs[])
chisq.rep <- sum(diff.rep[])
p.value <- step(chisq.rep - chisq.obs)
#Prior distributions
a ~ dflat() #Diffuse prior on a
b ~ dflat() #Diffuse prior on b
tau.eps <- pow(sigma.eps, -2)
sigma.eps ~ dunif(0, 10)
}
```

$$\log(\lambda_i) = a + bt_i + \varepsilon_i \tag{7.1}$$

In this equation, ε_i is a random error term in each age period, which we will take to be normally distributed with mean zero and constant (unknown) variance in each period. We will perform Bayesian inference on the unknown variance, along with the parameters of the loglinear trend model. The OpenBUGS script for this model is shown in Table 7.2. OpenBUGS parameterizes the normal distribution in terms of the mean and precision, which is the reciprocal of the variance. In this script we place the prior on the standard deviation in each time period, which is the square root of the variance. Note the use of a diffuse uniform distribution for the standard deviation. A uniform distribution is used to avoid problems one can encounter using the nearly improper gamma(0.0001, 0.0001) distribution, as discussed in [1].

If we run two MCMC chains, initialized at ($a = 5$, $b = -0.1$) and ($a = -5$, $b = 0.1$), we find that about 2,000 samples are needed to be reasonably sure of convergence (see Chap. 6 for details on checking for convergence). Running an additional 100,000 samples gives adequately low Monte Carlo error for parameter estimation. With variability introduced in each age period, the marginal posterior distribution for b now extends above 0, with a 95th percentile of 0.03. So while there is still evidence for a decreasing trend in λ, it is weaker than in the first case, which omitted extra-Poisson variability. The Bayesian p-value increases to 0.39, indicative of a model that is substantially better at replicating the observed data

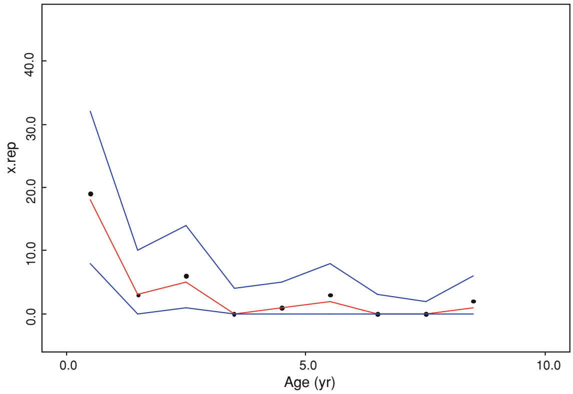


Fig. 7.3 Plot of 95% credible interval (*blue lines*) for predicted event count under loglinear model with additional random variability in λ. Dots are observed event counts

Table 7.3 Component performance data over time, from [2]

Time	Failures	Exp. time
0.5	1	31.64
1.5	1	44.135
3	3	156.407
6	1	608.38
8.5	4	189.545
9.5	9	214.63
10.5	5	216.355
11.5	6	204.091
12.5	12	185.926
13.5	2	157.536
14.5	1	127.608
15.5	3	104.105
16.5	5	77.86
20	3	121.167

than the loglinear model with its p-value of 0.009. The plot of credible intervals for the replicated event count in Fig. 7.3 graphically illustrates the improved predictive ability of this model in comparison with the loglinear model without additional variability in λ over time.

This problem illustrates the so-called *hierarchical Bayes* approach. Hierarchical Bayes is so-named because it utilizes hierarchical or multistage prior distributions. In each age bin in this example, the distribution of λ is conditional upon a value of ε, the error term describing the random year-to-year variation in log(λ) about a straight line. But ε is uncertain, and we model this uncertainty by introducing a prior distribution for the standard deviation of ε, which is assumed to be constant across time. Thus we have a hierarchy of prior distributions.

Recall from Chap. 5 the example using data from [2], reproduced in Table 7.3. In Chap. 5, we considered two aleatory models for these data: a simple Poisson model with constant λ, and a time-dependent Poisson model with a loglinear trend in λ. However, neither of these models performed well on either qualitative

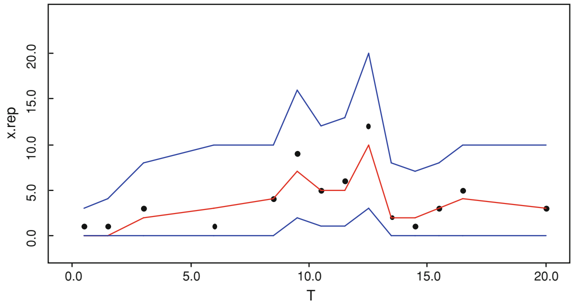


Fig. 7.4 Plot of 95% posterior predictive intervals for replicated data in each time bin. All observed data fall within these intervals, suggesting good predictive ability for a loglinear model with additional variability in λ over time

or quantitative posterior predictive checks. The Bayesian p-values were 0.004 for the constant-λ model, and 0.01 for the loglinear model.

Let us now examine a hierarchical time trend model, as above, where we introduce additional variability in λ over time. Using the script in Table 7.2, we find a less significant trend over time than with the simple loglinear trend model used in Chap. 5. The 90% credible interval for b, which was (0.034, 0.16) for a simple loglinear model, is now (−0.04, 0.14). The predictive validity has improved with the addition of variability over time, as shown by the posterior predictive plot in Fig. 7.4, and the Bayesian p-value has increased to 0.43.

7.2 Source-to-Source Variability

A more common past application of hierarchical Bayes analysis in PRA has been as a model of variability among data sources, for example variability in emergency diesel generator (EDGs) performance across plants, or across time. Such an example of source-to-source variability is discussed at length in [3] from the perspective of two-stage Bayes and empirical Bayes, which are both approximations to a hierarchical Bayes treatment, and have been commonly used in past PRAs, before the availability of tools such as OpenBUGS, which make a fully hierarchical Bayes analysis tractable. The hierarchical Bayes approach to the same problem is presented in [4]. Other examples related to PRA can be found in [4–6].

We will first examine the EDG example from [3]. The data for EDG failures are reproduced in Table 7.4. At each plant, the number of observed failures, X_i, is modeled with a binomial aleatory model with parameters n_i and p_i. To develop a hierarchical model for p, we specify a *first-stage prior*, which is often of a particular functional form, often a conjugate prior, although this is not necessary. For this example, we will take the first-stage prior to be a conjugate beta (α, β) distribution. We now need to specify a prior distribution on the first-stage parameters, α and β. This is called the *second-stage prior*, or *hyperprior*. Note that although nothing limits the analysis to two stages, the use of more than two stages has been rare in PRA applications. It is common, although not necessary, to employ independent, diffuse distributions at the second-stage.

Table 7.4 EDG failure data for 10 plants, taken from [3]

Plant	Failures	Demands
1	0	140
2	0	130
3	0	130
4	1	130
5	2	100
6	3	185
7	3	175
8	4	167
9	5	151
10	10	150

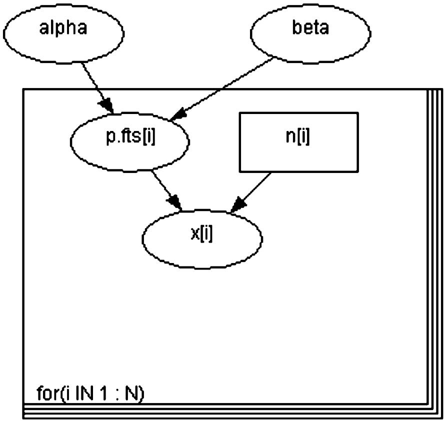


Fig. 7.5 DAG for hierarchical Bayes model of source-to-source variability

Hierarchical problems such as this one are most simply represented via a DAG. The DAG for this first example is shown in Fig. 7.5. Note in this figure that α and β are independent, prior to the observation of the data. This obviates the need to develop a joint prior distribution for α and β that includes dependence. Once the data are observed, these nodes become dependent, and the joint posterior distribution will reflect this dependence. Note that in some cases the high degree of correlation between α and β can lead to very slow convergence to the joint posterior distribution. In such cases, it may be helpful to reparameterize the problem in terms of the mean and a dispersion measure, which are approximately independent in the joint posterior distribution. We will examine this situation in more detail below.

Another point worth noting about the DAG in Fig. 7.5 is that the EDG failure probabilities (`p.fts[i]`) are conditionally independent, given values of α and β. The posterior predictive distribution of p, representing source-to-source variability, will be given by an average of the posterior distribution for p, conditional upon α and β (a beta distribution), weighted by the joint posterior distribution for α and β. We can take advantage of the fact that, as Fig. 7.5 illustrates, the components p_i are conditionally independent, given α and β, to write

Table 7.5 OpenBUGS script for modeling plant-to-plant variability in EDG performance

```
model {
for (i in 1 : N) {
        p[i] ~ dbeta(alpha, beta) #First-stage prior
        x[i] ~ dbin(p[i], n[i]) #Binomial dist. for failures at each plant
        }
p.pred ~ dbeta(alpha, beta)
alpha ~ dgamma(0.0001, 0.0001) #Vague hyperprior for alpha
beta ~ dgamma(0.0001, 0.0001) #Vague hyperprior for beta
inits
list(alpha = 1, beta = 25)
list(alpha = 0.5, beta = 75)
}
```

$$\pi\left(p_i|\tilde{x},\,\tilde{n}\right) = \int_0^1\int_0^1\cdots\int_0^1\left\{\int\int\left[\prod_{i=1}^{N}\pi_1\left(p_i|\alpha,\,\beta\right)\right]\pi_2\left(\alpha,\,\beta|\tilde{x},\,\tilde{n}\right)d\alpha d\beta\right\} dp_1dp_2\cdots dp_{i-1}dp_{i+1}\cdots dp_n$$

$$= \int\int\pi_1(p_i|\tilde{x},\tilde{n},\,\alpha,\,\beta)\pi_2(\alpha,\,\beta|\tilde{x},\,\tilde{n})d\alpha d\beta$$

The second line in this equation is obtained by interchanging the order of integration. Thus, the marginal posterior distribution for p at any particular plant is a continuous mixture of beta distributions, mixing over the joint posterior distribution of the hyperparameters, α and β. The distribution describing plant-to-plant variability in p is the posterior predictive distribution, sometimes referred to in PRA (especially older references) as the average population variability curve:

$$\pi(p^*|\tilde{p}) = \int\int\pi_1\left(p^*|\alpha,\,\beta,\,\tilde{\lambda},\,\tilde{x},\,\tilde{n}\right)\pi_2(\alpha,\,\beta|\tilde{x},\,\tilde{n})d\alpha d\beta$$

$$= \int\int\pi_1(p^*|\alpha,\,\beta,\,\tilde{x},\,\tilde{n})\pi_2(\alpha,\,\beta|\tilde{x},\,\tilde{n})d\alpha d\beta$$

It is thus a similar mixture of beta distributions. It is generated in OpenBUGS (node `p.pred` in the script) by sampling α and β from their joint posterior distribution, and then sampling p^* from the first-stage prior, a beta distribution in this example. The OpenBUGS script used to analyze this problem is shown in Table 7.5.

Two Markov chains, each starting from a separate point in the parameter space, were used with this script. More than two chains may be useful in some problems, although two are sufficient for this example. More than one chain aids in checking convergence, as discussed in Chap. 6. Each of the chains must be given an initial value of α and β. Reference [4] discuss the use of empirical Bayes as an aid in selecting starting values for the chains. The empirical Bayes estimates are 1.2 for α and 63 for β, and the initial values shown in the script in Table 7.5 were dispersed around these values. Note that the empirical Bayes estimates can be calculated using the online calculator at https://nrcoe.inel.gov/radscalc/Default.aspx.

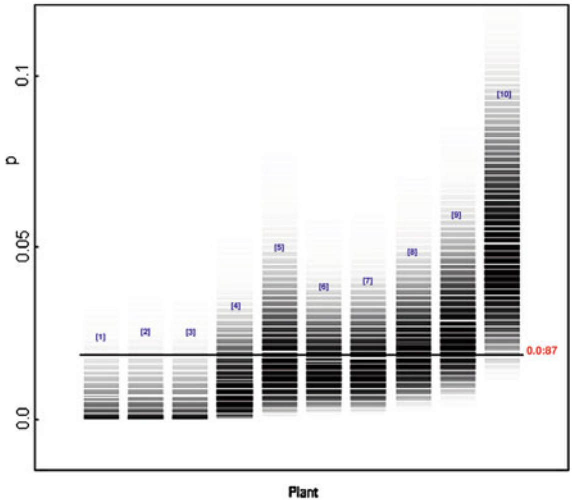


Fig. 7.6 Summary of marginal posterior distributions for p at each of the 10 plants

Table 7.6 Results for EDG #1, with comparison of other methods from [3]

	5th	50th	95th	Mean
Empirical Bayes	4.7E-04	4.4E-03	1.7E-02	5.9E-03
Two-stage Bayes	1.2E-04	3.3E-03	1.8E-02	5.2E-03
Hierarchical Bayes	5.9E-05	4.5E-03	1.9E-02	6.3E-03

The model converges quickly: 2,000 iterations are sufficient for burn-in. Running an additional 100,000 iterations with which to estimate parameter values gives a posterior predicted mean for `p.pred` of 0.02, with a 90% credible interval of (3.7E-4, 0.06). Figure 7.6 summarizes the marginal posterior distribution of p for each of the 10 plants. The marginal posterior distribution for p for EDG #1 is summarized in Table 7.6, along with the results of the other approaches described in [3]. The results from the hierarchical Bayes analysis are generally comparable to those from empirical and two-stage Bayes, but with somewhat wider uncertainty bounds. Also, as noted by [3], α and β are highly correlated in the posterior distribution: the rank correlation coefficient calculated by OpenBUGS is 0.98. Note that this correlation is automatically accounted for in the MCMC sampling process in OpenBUGS.

We consider next a similar example for a Poisson aleatory model. Table 7.7 presents component failure data for 11 sources. Figure 7.7 is a caterpillar plot of 95% credible intervals for λ for each of these sources, based on updating the Jeffreys prior for λ with the data from each source. The lack of overlap of the intervals in this figure suggests that there may be extra-Poisson variation among these sources, so that a model with a single λ may not be adequate.

Before we rush to use a hierarchical model in this case, we pause to do some posterior predictive checking of a simple Poisson aleatory model with a single λ, using the Jeffreys prior to focus attention on the aleatory model.

Table 7.7 Component failure rate data for hierarchical Bayes example

Source	Failures	Exposure time (year)
1	2	15.986
2	1	16.878
3	1	18.146
4	1	18.636
5	2	18.792
6	0	18.976
7	12	18.522
8	5	19.04
9	0	18.784
10	3	18.868
11	0	19.232

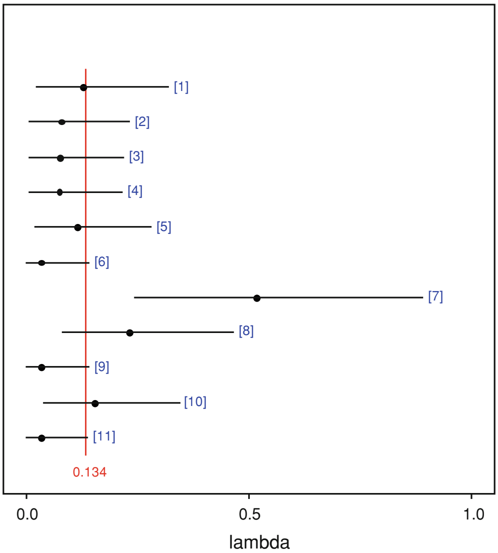


Fig. 7.7 Posterior 95% credible intervals for λ for each of the 11 sources shown in Table 7.7, illustrating apparent extra-Poisson variability among the sources

The plot of replicated vs. observed event counts in Fig. 7.8 suggests that a model with a single value of λ cannot reproduce the variability seen in the observed data. The Bayesian p-value for this simple model is about 0.0002, confirming its poor validity.

Because the simple Poisson model has poor predictive validity with respect to the observed data, we turn to a hierarchical Bayes model. This model is exactly analogous to that used for the EDGs in the first example. In this case, because the aleatory model is Poisson instead of binomial, we will use a gamma(α, β) distribution as the first-stage prior, although a conjugate first-stage prior is again not necessary. The OpenBUGS script for this model is shown in Table 7.8.

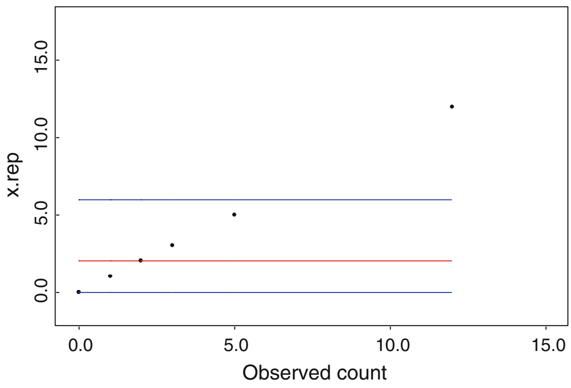


Fig. 7.8 Plot of 95% intervals for replicated event counts, illustrating inability of model with single value of λ to replicate observed variability in the data in Table 7.7

Table 7.8 OpenBUGS script for hierarchical Bayes analysis of variability in λ

```
model {
for (i in 1 : N) {
        lambda[i] ~ dgamma(alpha, beta) #Model variability in LOSP frequency
        mean[i] <- lambda[i] * time[i] #Poisson parameter for each plant
        x[i] ~ dpois(mean[i]) #Poisson dist. for events at each plant
        }
lambda.pred ~ dgamma(alpha, beta)
alpha ~ dgamma(0.0001, 0.0001) #Vague hyperprior for alpha
beta ~ dgamma(0.0001, 0.0001) #Vague hyperprior for beta
inits
list(alpha = 1, beta = 1)
list(alpha = 0.5, beta = 5)
}
```

Two Markov chains, each starting from a separate point in the parameter space, were used in the analysis, as above. We again used empirical Bayes as an aid in selecting starting values for the two chains. The empirical Bayes estimates are 0.85 for α and 6.4 year for β.

The model converges quickly: 2,000 iterations are sufficient. Running an additional 100,000 iterations with which to estimate parameter values gives a posterior predicted mean for `lambda.pred` of 0.16, with a 90% credible interval of (9.8E-4, 0.53). Figure 7.9 summarizes the marginal posterior distribution of λ for each of the 11 sources.

The posterior predictive plot shown in Fig. 7.10 illustrates the enhanced ability of the hierarchical model to replicate the variability in the observed data. This is reinforced by the Bayesian p-value of 0.45 for the hierarchical model.

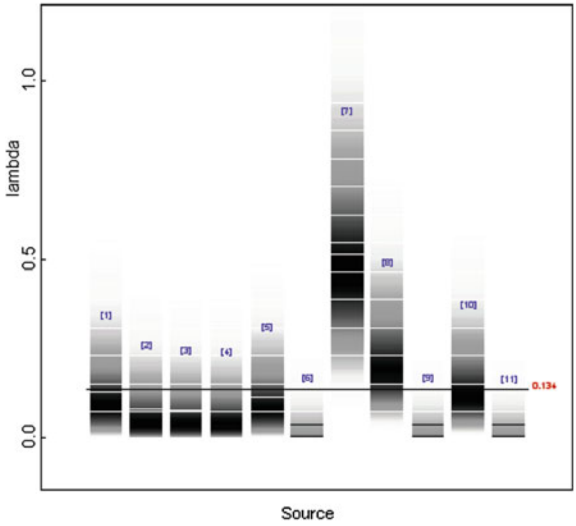


Fig. 7.9 Summary of marginal posterior distribution for λ for the 11 sources listed in Table 7.7

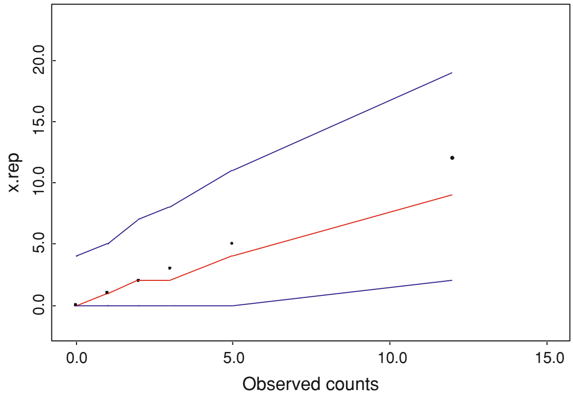


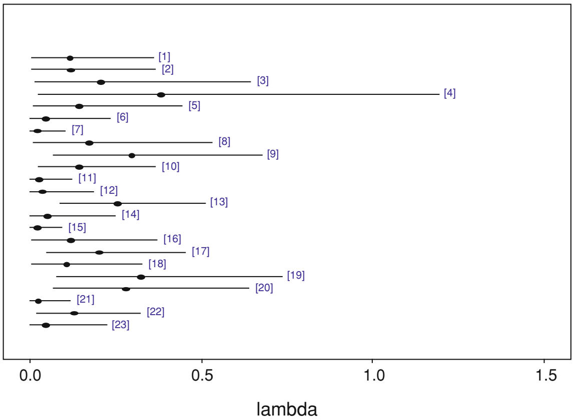
Fig. 7.10 Plot of 95% intervals for replicated event counts, illustrating ability of hierarchical Bayes model of source-to-source variability to replicate data in Table 7.7

7.3 Dealing with Convergence Problems in Hierarchical Bayes Models of Variability

We turn now to an example in which convergence to the posterior distribution is not so rapid.[1] In our experience, convergence issues with hierarchical Bayes models for population variability make them among the most challenging of common PRA Bayesian inference problems from a computational perspective.

[1] As noted earlier, the problem illustrated here is really a problem of poor mixing of the chains due to high correlation between the parameters of the second-stage prior; however, the effect is the same and must be ameliorated before the MCMC samples can be used for parameter estimation.

Fig. 7.11 Side-by-side plot of 95% credible intervals for 23 data sources in Table 7.9



Modern tools such as OpenBUGS have eliminated the computational challenges of the high-dimensional integrals encountered in these kinds of problems, but the burden is still upon the analyst to be on the alert for convergence problems. We use the following example to illustrate the problem and a practical solution.

Consider the 23 sources of data for loss of offsite power, taken from [7]. shown in Table 7.9.

Let us assume that the failure count for each source is Poisson-distributed with rate λ_i. Figure 7.11 shows a side-by-side plot of the 95% credible intervals for λ for each of these sources, based on updating the Jeffreys prior, and illustrates the variability in the sources. However, most of the intervals overlap, suggesting that the source-to-source variability is not large.

Let us first adopt a conjugate first-stage prior, which in this case is a gamma distribution. As above, empirical Bayes can be used as an aid in specifying initial values of α and β. The empirical Bayes estimates are 2.1 and 23 year, respectively. Using these point estimates of α and β in the first-stage gamma prior that describes source-to-source variability, the average rate is 0.09/year, with a 90% credible interval of (0.02, 0.2).

Turning now to hierarchical Bayesian analysis using OpenBUGS with the script given in Table 7.8 and the data in Table 7.9, we must specify hyperpriors for α and β. We will use independent, diffuse distributions, so that the data drive the results. A gamma distribution with both parameters very small gives a distribution that is essentially flat over the positive real axis. With both parameters of the gamma distribution equal to 10^{-4}, the 5th percentile is zero, effectively, and the 95th percentile is approximately 10^{-220}. Thus, there is a very sharp vertical asymptote at zero, and the density is essentially flat (and approximately equal to zero) for values greater than zero.

Following the guidance for checking convergence provided in Chap. 6, we run multiple chains in order to monitor convergence, both graphically via chain history plots, and quantitatively, using the BGR convergence diagnostic calculated by

Table 7.9 Example data from [7]

Events	Exposure time (year)
1	13.054
1	12.77
1	7.22
1	3.944
1	10.548
0	10.704
0	24
1	8.76
3	11.79
2	17.5
0	20.03
0	13.39
5	21.5
0	10.075
0	26.32
1	12.54
3	17.5
1	14.3
3	10.89
3	12.5
0	21.38
2	19.65
0	11.34

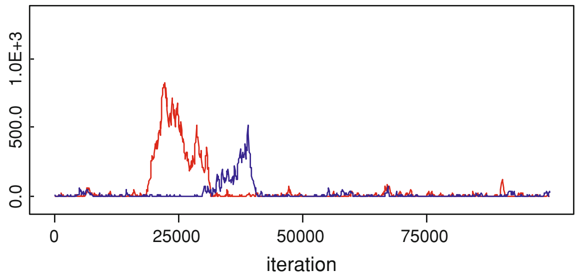


Fig. 7.12 History of 100,000 iterations for α, showing failure of chains to mix

OpenBUGS. OpenBUGS cannot generate initial values from the diffuse gamma hyperpriors used for α and β, so initial values must be supplied. We start with the point estimates from the empirical Bayes analysis above, and pick values around these, as these approximate the mode of the joint posterior distribution for α and β.

Figure 7.12 shows the history of 100,000 samples for α. The two chains are not well mixed, indicating potential convergence problems. The corresponding BGR plot is shown in Fig. 7.13. If the chains have converged to the posterior distribution, the normalized BGR ratio (red line) should be about 1.0, and all three lines

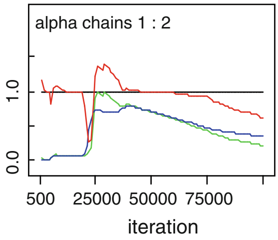


Fig. 7.13 BGR diagnostic for first 100,000 iterations for α, illustrating failure to converge

Table 7.10 OpenBUGS script for hierarchical Bayes model for λ, parameterized in terms of mean and coefficient of variation of gamma first-stage prior

```
model {
for (i in 1 : N) {
        lambda[i] ~ dgamma(alpha, beta) #Model variability in LOSP frequency
        mu[i] <- lambda[i] * time[i] #Poisson parameter for each plant
        x[i] ~ dpois(mean[i]) #Poisson dist. for events at each plant
        }
lambda.pred ~ dgamma(alpha, beta)
alpha <- pow(COV, -2)
beta <- alpha/mean
mean ~ dgamma(0.0001, 0.0001) #Vague hyperprior for alpha
COV ~ dgamma(0.0001, 0.0001) #Vague hyperprior for beta
inits
list(mean = 0.01, CV = 1.5)
list(mean = 0.1, CV = 0.5)
}
```

should be stable. This is clearly not the case, so we have not converged after 100,000 samples.

Running the chains for an additional 100,000 iterations gives results similar to those above. Even after more than 10^6 iterations, this behavior was still observed, and the parameter estimates provided by the two chains differed significantly.

Highly correlated parameters can be an obstacle to convergence, as noted by [8], among others. In this case, the correlation between α and β in the joint posterior distribution is estimated by OpenBUGS to be 0.98. We can ameliorate this problem by reparameterizing the gamma first-stage prior in terms of the mean and coefficient of variation, which we would expect to be less highly correlated. For a gamma distribution, the mean is given by α/β. The coefficient of variation is the ratio of the standard deviation to the mean, which for the gamma distribution will be $\alpha^{-0.5}$. We place independent, diffuse gamma hyperpriors on the mean and coefficient of variation. Initial values are again chosen using the results from the empirical Bayes analysis above. The OpenBUGS script for the reparameterized problem is shown in Table 7.10.

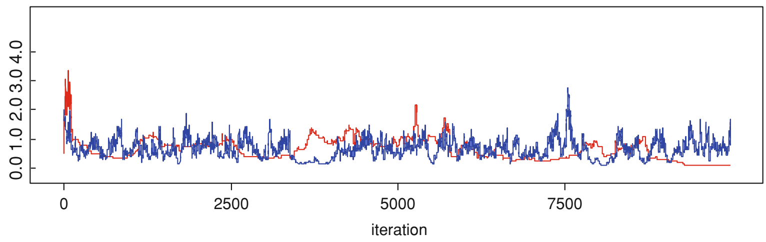
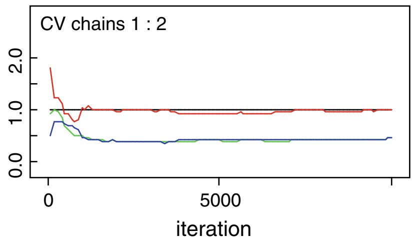


Fig. 7.14 History of 10,000 iterations for coefficient of variation, illustrating convergence

Fig. 7.15 BGR diagnostic for 10,000 iterations for coefficient of variation, illustrating convergence



As Figs. 7.14 and 7.15 illustrate, we now appear to have convergence within the first couple of thousand iterations. To be safe, we discard the first 10,000 iterations and do not use these samples in estimating parameter values.

An additional 100,000 iterations were used to estimate the desired parameters. The predicted distribution of λ has mean 0.09/year, with a 90% credibility interval of (0.02, 0.20). These results match those from the empirical Bayes analysis above; in general, the results are not necessarily so similar. In particular, credible intervals from the hierarchical Bayesian analysis tend to be wider, especially for the overall population variability distribution (i.e., the posterior predictive distribution of λ because of the inclusion of uncertainty about the hyperparameters α and β. This uncertainty is sometimes included in empirical Bayes analyses through asymptotic approximations, although to simplify the exposition we did not do so in this example. For this example, the uncertainty in the hyperparameters is relatively small, a result of the low degree of variability among the sources.

7.4 Choice of First-Stage Prior

We now examine the impact of choosing a different functional form for the first-stage prior describing the variability in λ from source to source. We illustrate this using the data in Table 7.9. As discussed by [6], use of conjugate first-stage priors can lead to unreasonably small lower percentiles, especially in cases where there are several orders of magnitude variability among the sources being modeled.

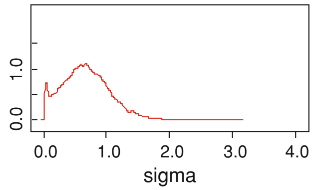


Fig. 7.16 Marginal posterior distribution for σ, with "spike" at 0, an indication of little source-to-source variability

With modern tools such as OpenBUGS, conjugate distributions are no longer a requirement of the analysis. A common nonconjugate alternative to the gamma first-stage prior is the lognormal distribution:

$$\pi_1(\lambda|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma\lambda}\exp\left[-\frac{(\log\lambda - \mu)^2}{2\sigma^2}\right], \quad -\infty < \mu < \infty,\ \sigma > 0$$

Care must sometimes be taken in the choice of hyperpriors in order to avoid numerical difficulties. In particular, the diffuse gamma distribution used as a second-stage prior above can be problematic as a hyperprior for σ (and for dispersion parameters generally). This issue is discussed at length in [1], where a uniform distribution with finite bounds is recommended as an alternative. A value of 1.4 for σ corresponds to a lognormal error factor of 10, and thus to a ratio of 100 between the 95 and 5th percentiles. Therefore, a uniform (0, 5) hyperprior for σ should be quite diffuse in practice. We will check the marginal posterior distribution for σ to ensure that the upper tail has not been truncated, a sign that a wider hyperprior distribution is needed.

With a uniform (0, 5) hyperprior on σ, and a flat hyperprior on μ (μ which is a logarithmic mean, can be negative), convergence was achieved in the first 1,000 iterations. Running another 10,000 samples to estimate parameter values, we find the mean of the posterior predictive distribution of λ to be 0.10. The 90% credible interval is (0.02, 0.24), slightly wider than with the gamma first-stage prior above. The marginal posterior density of σ is shown below. The right tail is not truncated, indicating that the uniform (0, 5) hyperprior was sufficiently diffuse. Note the "spike" at 0 in this distribution. This is an indication that there is little source-to-source variability in the data, as noted above (Fig. 7.16).

In [9] the authors present a hierarchical Bayesian analysis of failure rate data for digital instrumentation and control equipment. Figure 7.17 illustrates the extremely large variability in the sources the authors examined (many intervals do not overlap), and suggests that a highly skewed distribution will be required to adequately model the source-to-source variability.

With a gamma first-stage prior, and diffuse hyperpriors on α and β, OpenBUGS converges within 1,000 iterations, with no reparameterization needed. An additional 50,000 iterations gave a posterior predictive mean value of 0.09, a median of 0.011, and a 90% credible interval of (6.65×10^{-8}, 0.43). The posterior mean of the gamma shape parameter, α, was 0.24. Such a small value of α leads to a sharp vertical asymptote at zero in the first-stage prior, with an excessive amount of

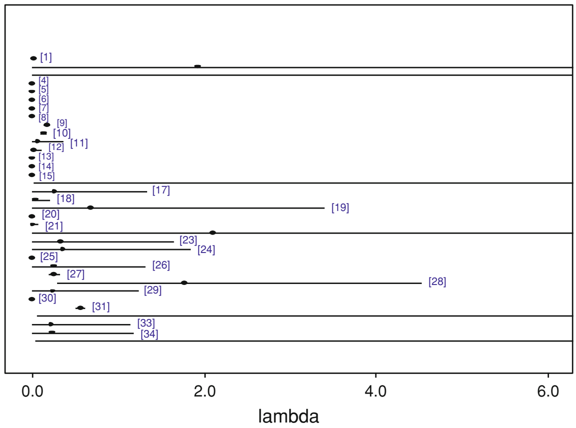


Fig. 7.17 Side-by-side plot of 95% intervals for data sources in [9]. Dots are posterior means from update of Jeffreys prior

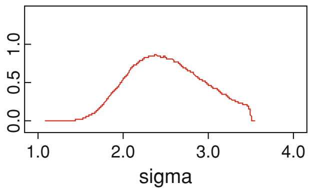


Fig. 7.18 Marginal posterior distribution for σ illustrating truncation of upper tail

probability mass near zero, which is the only way a gamma first-stage prior can accommodate such extremely large source-to-source variability.

The lognormal distribution is perhaps a better choice for a first-stage prior to model variability that ranges over several orders of magnitude. The authors of [9] recommend a lognormal distribution in place of the gamma distribution; however, they based their hyperpriors for μ and σ on the data, using a uniform(-7, -0.1) distribution for μ, and a uniform(1, 3.5) distribution for σ. With these choices, the posterior predictive distribution of λ has a mean of 0.29, a median of 0.007, and a 90% interval of (8.6×10^{-5}, 0.49). Figure 7.18 shows the marginal posterior distribution for σ, showing the truncation of the upper tail caused by the overly narrow data-based hyperprior.

The mean of the lognormal distribution is $\exp(\mu + \sigma^2/2)$. Thus, the mean is very sensitive to σ; the median, given by e^{μ}, is unaffected by σ. If we replace the data-based hyperpriors used by [9] with more diffuse distributions, we expect the mean to increase. We replaced the hyperprior for μ with a flat prior over the real axis, and the distribution for σ with a more diffuse uniform(0, 5) distribution. With these more diffuse hyperpriors, the mean of the predictive distribution for λ increased to 1.1, with a 90% credible interval of (6.3×10^{-5}, 0.55). Note that the mean, which lies well to the right of the 95th percentile, is no longer a representative value. The median, as expected, was robust against this change, remaining at 0.007. The marginal posterior distribution for σ is shown below. Note the lack of truncation in the upper tail (Fig. 7.19).

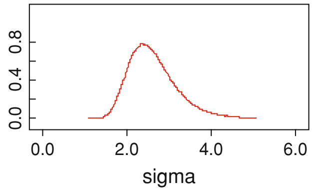


Fig. 7.19 Marginal posterior distribution for σ with more diffuse hyperprior illustrating no truncation in upper tail

Table 7.11 OpenBUGS script for modeling plant-to-plant variability in EDG performance, logistic-normal first-stage prior

```
model {
for (i in 1 : N) {
        x[i] ~ dbin(p[i], n[i]) #Binomial dist. for failures at each plant
        p[i] <- exp(log.p[i])/(1 + exp(log.p[i])) #First-stage prior for FTS probability at
each plant
        log.p[i] ~ dnorm(mu, tau)
        }
#Overall average distribution (posterior predictive distribution)
p.pred <- exp(log.p.pred)/(1 + exp(log.p.pred))
log.p.pred ~ dnorm(mu, tau) inits
#Hyperpriors (second-stage priors)
mu ~ dflat()
sigma ~ dunif(0, 10)
tau <- pow(sigma, -2)
}
list(mu = -5, sigma = 1)
list(mu = -7, sigma = 2)
```

To model source-to-source variability in p, the parameter of a binomial aleatory model, one could use a lognormal distribution. However, because the range of the lognormal distribution is unbounded, care must be exercised to ensure that values of $p > 1$ are not produced. To avoid this difficulty, the analyst may wish to use a logistic-normal distribution as a nonconjugate first-stage prior. The density function of the logistic-normal distribution is given by

$$\pi_1(p|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma p(1-p)} exp\left\{-\frac{\left[ln\left(p/1-p\right)-\mu\right]^2}{2\sigma^2}\right\}$$

In the case of the lognormal distribution, the logarithm of the lognormally distributed variable has a normal distribution with mean μ and variance σ^2. The relationship is similar for the logistic-normal distribution, only it is the logit of the variable that is normally distributed. The OpenBUGS script in Table 7.11 implements the logistic-normal hierarchical model for p, with diffuse hyperpriors on μ and σ (recall that OpenBUGS parameterizes the underlying normal distribution in terms of $\tau = \sigma^{-2}$.

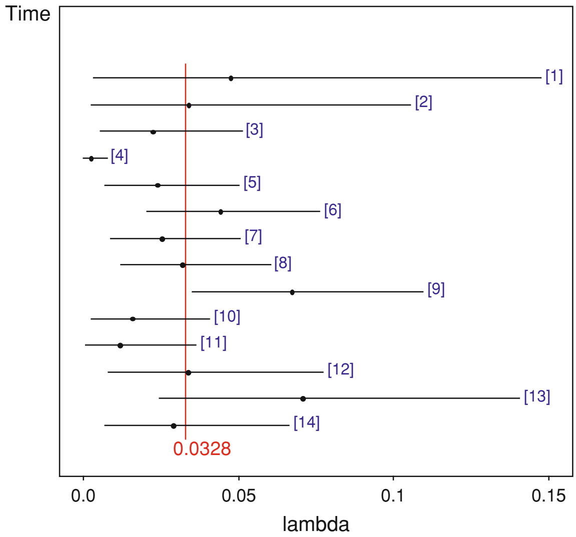


Fig. 7.20 Plot of 95% posterior credible intervals for λ, based on updating Jeffreys prior with data from Table 7.3

Using the data in Table 7.4, the model converges within about 1,000 iterations. Running an additional 100,000 iterations gives a posterior predicted mean for `p.pred` of 0.03, with a 90% credible interval of (6.4E-4, 012). Recall that with a beta first-stage prior, the posterior predicted mean and 90% interval were 0.02 and (3.7E-4, 0.06), so the choice of first-stage prior does have a discernible influence.

7.5 Trend Models Revisited

Let us return now to the example from [2], which we have explored both here and in Chap. 5. In Chap. 5, we saw that a simple loglinear trend model for λ did not have good predictive ability. Earlier in this chapter, we enhanced the loglinear model by introducing additional variability over time, and found that such a model had substantially better predictive ability than the simple trend model examined in Chap. 5. However, our initial graphical exploration in Chap. 5 was a caterpillar plot of λ, based on updating the Jeffreys prior for λ with the data in each time bin. This plot, reproduced in Fig. 7.20, was not suggestive of a monotonic time trend in λ, but appeared to be more indicative of random variability in λ across the time bins.

We can use the script in Table 7.8 to fit a hierarchical model representing extra-Poisson variability across the time bins, but without an underlying time trend. Doing this gives a model with good predictive ability, as shown by the posterior predictive plot in Fig. 7.21, and the Bayesian p-value of 0.51.

So we have two candidate models, both of which perform well on the posterior predictive checks: a loglinear trend model with additional variability over time and

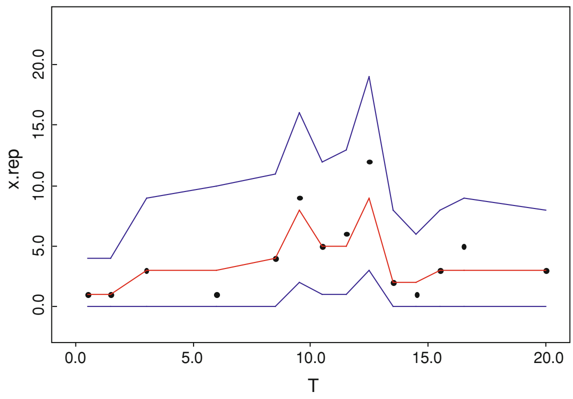


Fig. 7.21 Plot of 95% posterior predictive intervals for replicated data in each time bin. All observed data fall within these intervals, suggesting good predictive ability of hierarchical model for extra-Poisson variability across time

a model with no trend but with extra-Poisson variability over time. Which model do we choose?

The answer to this question depends on our assumptions about the data-generating process. Choosing the hierarchical model without a trend means that we are indifferent to re-ordering the data. In other words, we have no expectations that the event counts will be lower (or higher) in earlier time periods than in later ones. In technical terms, such an assumption is referred to as *exchangeability*. In contrast, if we are not willing to assume the data are exchangeable, perhaps because we have reasons to believe that a change in operating practice would lead to increasing event counts over time, then we should adopt a trend model.

7.6 Summary

In this chapter we have presented hierarchical Bayes models for extra variability that cannot be accounted for by simple models in which the parameter of interest is constant or is a simple monotonic function of time. The key feature of these models is that the prior distribution is specified in stages, with two stages being seen most commonly in PRA applications.

Such models have mostly been applied in PRA to account for source-to-source variability in data, and most of those applications have been approximations to full hierarchical Bayes models. The two common approximations are parametric empirical Bayes models, widely applied by the U.S. NRC, and the two-stage Bayes model of [10].

Convergence can be problematic for hierarchical Bayes models, and so care must be utilized to avoid errors caused by lack of convergence. The functional form of the first-stage prior distribution can have a significant influence on the results when modeling source-to-source variability, also. When variability among the source spans orders of magnitude, a nonconjugate lognormal or logistic-normal first-stage prior helps to avoid the extremely small lower percentiles that arise from a conjugate gamma or beta first-stage prior. However, the mean of the posterior predictive distribution may no longer be a representative value, and the analyst may wish to use the median instead.

7.7 Exercises

1. Using the data in Table 7.7, replace the gamma first-stage prior with a lognormal distribution. Use a *dflat*() hyperprior for the first parameter. For the second parameter, reparameterize in terms of σ and place a *dunif*(0, 10) hyperprior on σ.
   (a) How do the mean and median of `lambda.ind` compare to the results in the text? Explain.
   (b) How do the 90% intervals compare?
   (c) Any conclusions about choosing a first-stage prior?
2. Using the data in Table 7.4, with a logistic-normal first-stage prior for p, and the diffuse hyperpriors in the text, compare the resulting posterior mean and 90% credible interval for EDG #1 with the results for the beta first-stage prior given in Table 7.5.
3. The file "edg_data.txt" on the website for the text contains data for failure on demand for 195 EDGs. Recalling that the MLE of p is given by x/n, you should find that the MLE is > 0.05 for EDGs 183, 184, and 191–195. There is a desire to demonstrate that $\Pr(p > 0.05) < 0.05$. In English, we want to show that we are 95% sure that EDG reliability on demand is at least 95%. If we analyze each EDG separately, using the Jeffreys prior, we will find quite a few that do not meet the criterion (i.e., too many false positives). Pooling the data would also be inappropriate, giving a very narrow credible interval for p; all of the EDGs would meet the criterion by a wide margin. We would like to get a better answer than either of these simple approaches gives by developing a hierarchical Bayes model that describes the variation in p across the 195 EDGs in the dataset.
   (a) Use OpenBUGS to analyze a hierarchical Bayes model for this data. Treat the number of failures for each EDG as binomial, with $p_i \sim \text{beta}(\alpha, \beta)$. Use independent diffuse hyperpriors for α and β. We want to find which EDGs have $\Pr(p > 0.05) > 0.05$.
   (b) Re-analyze this model, using a *uniform*(0, 10) hyperprior for α, and an independent *uniform*(0, 100) hyperprior for β. Does this change in priors affect your conclusions about which EDGs have $\Pr(p > 0.05) > 0.05$?
4. Three plants report the following data on initiating events:

Plant	No. of events	Critical years
1	5	0.5
2	1	0.5
3	14	0.8

(a) Find a posterior mean and 90% interval for each plant, using the Jeffreys prior for a Poisson aleatory model, and investigate whether a model where the three plants have the same frequency of initiating events appears to be adequate.

(b) Perform a quantitative check of whether the frequency is the same at all three plants.

References

1. Gelman A (2006) Prior distributions for variance parameters in hierarchical models. Bayesian Anal 1(3):515–533
2. Rodionov A, Kelly D, Uwe-Klügel J (2009) Guidelines for analysis of data related to ageing of nuclear power plant components and systems. Joint Research Centre, Institute for Energy, Luxembourg: European Commission
3. Siu NO, Kelly DL (1998) Bayesian parameter estimation in probabilistic risk assessment. Reliab Eng Syst Saf 62:89–116
4. Kelly DL, Smith CL (2009) Bayesian inference in probabilistic risk assessment–the current state of the art. Reliab Eng Syst Saf 94:628–643
5. Singpurwalla ND (2006) Reliability and risk: a bayesian perspective. Wiley, Washington, DC
6. Atwood C, LaChance JL. Martz HF, Anderson DJ, Engelhardt ME, Whitehead D et al (2003) Handbook of parameter estimation for probabilistic risk assessment. U. S. Nuclear Regulatory Commission
7. Atwood CL et al (1997) Evaluation of Loss of Offsite Power Events at Nuclear Power Plants: 1980–1996, U. S. Nucl Regulatory Comm, NUREG/CR-5496
8. Gelman A et al (2004) Bayesian data analysis, 2nd edn. Chapman and Hall/CRC, London
9. Yue M, Chu T-L (2006) Estimation of failure rates of digital components using a hierarchical bayesian method. International conference on probabilistic safety assessment and management. New Orleans
10. Kaplan S (1983) On a two-stage bayesian procedure for determining failure rates. IEEE Trans Power Apparatus Syst 102(1):195–262