

## 8. PARAMETER ESTIMATION USING EXISTING DATA SOURCES

### 8.1 Selection of Parameter Estimates from Existing Data Bases

The need to select parameter estimates from existing generic data bases may arise when performing a probabilistic risk assessment (PRA). This can occur when a PRA is being performed on a new plant that has no operating history or it may occur when no plant-specific information exists for a specific component. Whatever the reason, when it becomes necessary to select parameter estimates from generic data bases, certain cautions should be observed. These are

- (1) The generic data base should contain failure probability estimates for components that are identical or comparable to the ones in the PRA model in terms of size, component boundary definition, intended operational history (e.g., normally operating versus standby), and expected or postulated operating environment.
- (2) The generic data base should contain a recommended point estimate and an uncertainty distribution for each identified failure.
- (3) If possible, the primary sources of information used to develop the generic data base's failure probabilities and distributions should be information from other nuclear power plants. Supplemental information from non-nuclear sources should be used only when necessary to provide failure probabilities and distributions for components that cannot be obtained from nuclear power plant generic data sources.
- (4) Where possible, the generic data base's failure probabilities and distributions should be derived from actual failure events. If such information is not available, then failure probabilities and distributions generated by other techniques (e.g., expert elicitation) are acceptable.

- (5) Generic data base failure probabilities and distributions should reflect current trends. If significant trends exist within the failure data indicating either an increase or decrease in the failure probabilities or distributions, the underlying event failure information used to generate the failure probabilities and distributions should represent these recent events. However, if no significant trends exist, then data from all years can be used to estimate the failure probabilities.
- (6) The failure probability estimates contained within the generic data base should not be based on incestuous sources, i.e., the estimates should not be derived from two different sources that employed similar or different analysis techniques to the same ultimate set of failure information.

### 8.2 Combining Data From Different Sources

In this section, data come from a number of similar, but not identical, sources. For simplicity, the discussion is in terms of data from a number of nuclear power plants. However, the ideas can be applied much more widely.

#### 8.2.1 The Hierarchical Model

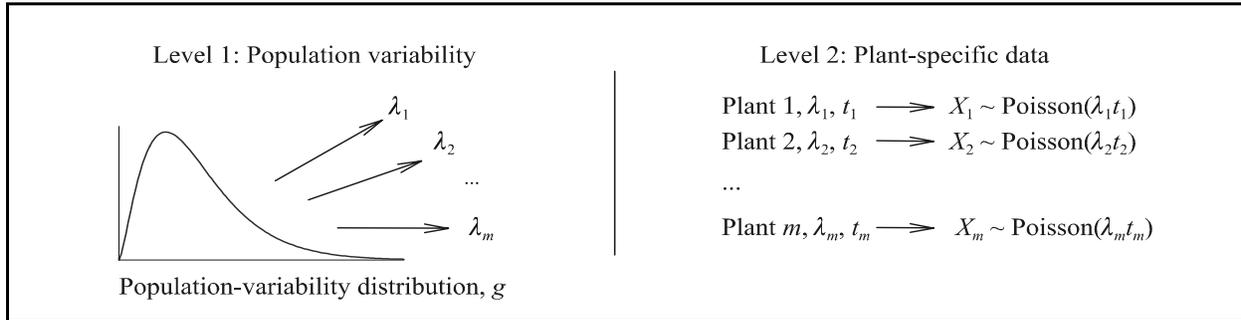
The situation is described by a **hierarchical model**, with two levels. The first level models the plants as a family, with the members resembling each other. The second level models the data that are generated at each plant.

To be more specific, suppose that initiating events are to be modeled, so the parameter of interest is  $\lambda$ . Level 1 of the model says that  $\lambda$  varies among the  $m$  plants, but only to a limited degree. Thus, the plants are not identical, but they resemble each other. This is modeled by a distribution  $g$  that describes the population variability. Before any data are generated, the distribution  $g$  is invoked  $m$  times, producing values  $\lambda_1$  through  $\lambda_m$ . These values of  $\lambda_i$  are independently generated, but they all come from the same distribution,

g. For each  $i$ ,  $\lambda_i$  is assigned to plant  $i$ . That is Level 1 of the hierarchical model. It is shown on the left side of Figure 8.1.

Level 2 of the model says that, conditional on the  $\lambda_i$  values, the plants independently produce data. Thus,

for each  $i$ , plant  $i$  is observed for time  $t_i$ , and it experiences a random number of initiating events,  $X_i$ , with  $X_i$  having a  $\text{Poisson}(\lambda_i t_i)$  distribution. This is shown on the right side of Figure 8.1.



**Figure 8.1** Hierarchical model for Poisson data.

The population-variability distribution  $g$  could be a gamma( $\alpha, \beta$ ) distribution, which has the computational advantage of being conjugate to the Poisson distribution. But that is not the only allowed distribution. It could also be a lognormal( $\mu, \sigma^2$ ) distribution, or some other distribution.

The data consist of the observation times,  $t_1$  through  $t_m$ , which are known and treated as fixed, and the event counts,  $x_1$  through  $x_m$ , which are treated as randomly generated. The unknown parameters consist of  $\lambda_1$  through  $\lambda_m$ , as well as any unknown parameters of  $g$ . These parameters of  $g$  could be  $\alpha$  and  $\beta$ , or  $\mu$  and  $\sigma^2$ , or some other parameters, depending on the assumed form of  $g$ . To emphasize the contrast between the two levels, the parameters of  $g$ , such as  $\alpha$  and  $\beta$  or  $\mu$  and  $\sigma^2$ , are sometimes called **hyperparameters**.

When the data instead are failures on demand, the situation is very similar. The population-variability distribution generates parameters  $p$ , one for each plant. The distribution  $g$  might be a beta distribution, or it might be some non-conjugate distribution, such as (truncated) lognormal or logistic-normal. The unknown parameters consist of the parameters of  $g$ , and the parameters  $p_1$  through  $p_m$ . The data consist of the counts of failures and demands at each plant,  $(x_1, n_1)$  through  $(x_m, n_m)$ . Examples 8.1 and 8.2 illustrate the two types of data.

In Example 8.1, most of the plants experience at least

one initiating event, and the total number of events is 361. Thus, the data set is large, and the methods given below perform well. Example 8.2, on the other hand, is a small data set. That is, most of the plants experienced no failures, and the total number of failures is only 6. This example was deliberately chosen for this handbook to illustrate problems that can occur with sparse data.

Two methods are given below for analyzing data by means of a hierarchical model. The results of each analysis include both an estimate of the population-variability distribution,  $g$ , and estimates of all the plant-specific parameters,  $\lambda_i$  or  $p_i$ .

## 8.2.2 The Parametric Empirical Bayes Method

### 8.2.2.1 General Approach

In spite of the name, this is not a truly Bayesian method. Instead, it is a kind of hybrid, involving a non-Bayesian step followed by a Bayesian step.

Step 1. Look at the variability in the data from the plants, and estimate  $g$ . That is, based on the data from all the plants, estimate the parameters of  $g$  by maximum likelihood, and obtain the resulting estimate of the distribution. Call the estimate  $\hat{g}$ .

Step 2. Now treat  $\hat{g}$  as a prior distribution. Perform the usual Bayesian update, with the prior distribution  $\hat{g}$  and data from a single plant,

to get the posterior distribution for the plant-specific parameter,  $\lambda_i$  or  $p_i$ .

**Example 8.1 Initiating events from many plants.**

The number of unplanned scrams at power,  $x$ , and the number of 1000 critical hours,  $t$ , are listed below for one calendar year (1984) at 66 plants. The data are from Martz et al. (1999).

Plant	$x$	$t$	Plant	$x$	$t$	Plant	$x$	$t$	Plant	$x$	$t$
Arkansas 1	3	6.2500	Farley 2	6	8.3333	Monticello	0	0.8106	San Onofre 2	5	5.2632
Arkansas 2	12	7.6433	Fort Calhoun	1	5.2632	North Anna 1	8	4.7619	San Onofre 3	7	5.0725
Beaver Val. 1	4	6.4516	GINNA	1	6.6667	North Anna 2	4	6.1538	St. Lucie 1	6	5.5556
Big Rock Point	2	6.8966	Grand Gulf	7	2.0896	Oconee 1	3	7.5000	St. Lucie 2	9	7.3770
Brunswick 2	3	2.6549	Haddam Neck	3	6.5217	Oconee 2	0	8.7840	Summer	11	5.5556
Callaway	12	1.5038	Hatch 1	7	5.6452	Oconee 3	4	6.5574	Surry 1	8	5.2980
Calvert Cliffs 1	5	7.5758	Hatch 2	7	3.1111	Oyster Creek	2	1.6949	Surry 2	14	7.4468
Cook 1	3	8.1081	Indian Pt. 2	4	4.7059	Palisades	1	1.5625	Susquehan. 1	7	6.5421
Cook 2	7	5.3030	Indian Pt. 3	7	6.9307	Pt. Beach 1	0	6.4201	Susquehan. 2	7	2.1472
Cooper Station	3	6.0000	Kewaunee	4	7.5472	Pt. Beach 2	0	7.5442	Turkey Point 3	8	7.3394
Crystal River	2	8.3333	LaSalle 1	9	6.2937	Prairie Island 1	4	8.3333	Turkey Point 4	9	5.0847
Crystal River 3	4	5.5556	LaSalle 2	11	5.4726	Prairie Island 2	0	7.8440	Vermont	2	7.1429
Davis-Besse	5	1.0846	Maine Yank.	7	6.6667	Quad Cities 1	3	4.7619	Yank.	23	4.3643
Davis-Besse	3	6.5217	McGuire 1	4	6.0606	Quad Cities 2	2	6.8966	Wash. Nucl. 2	6	6.3158
Diablo Cany. 1	8	3.8835	McGuire 2	16	6.9869	Robinson 2	0	0.6161	Zion 1	7	6.3063
Dresden 2	6	6.5934	Millstone 1	0	6.9902	Salem 1	10	2.6738	Zion 2		
Dresden 3	2	6.8966	Millstone 2	3	8.5714	Salem 2	10	3.3898			
Duane Arnold											
Farley 1											

**Example 8.2 Failure to start of AFW motor-driven segments at many plants.**

The number of failures to start on unplanned demands for motor-driven segments of the AFW system are tabulated for 68 plants, for 1987-1995. Here,  $x$  is the number of failures and  $n$  is the number of demands. Common-cause failures are excluded. The data are from Poloski et al. (1998, Table E-4).

Plant	$x$	$n$									
-------	-----	-----	-------	-----	-----	-------	-----	-----	-------	-----	-----

## 8. Parameter Estimation Using Existing Data Sources DRAFT

Arkansas 1	0	14	Crystal River 3	1	16	North Anna 2	0	18	Seabrook	0	17
Arkansas 2	0	9	Diablo Canyon 1	0	46	Oconee 1	0	18	Sequoyah 1	0	30
Beaver Valley 1	0	24	Diablo Canyon 2	0	30	Oconee 2	0	18	Sequoyah 2	0	41
Beaver Valley 2	0	43	Farley 1	0	34	Oconee 3	0	12	South Texas 1	0	69
Braidwood 1	0	13	Farley 2	0	54	Palisades	0	13	South Texas 2	0	87
Braidwood 2	0	24	Fort Calhoun	0	5	Palo Verde 1	0	7	St. Lucie 1	0	35
Byron 1	0	11	Ginna	0	28	Palo Verde 2	0	12	St. Lucie 2	0	21
Byron 2	0	26	Harris	0	98	Palo Verde 3	0	9	Summer	0	24
Callaway	0	57	Indian Point 2	1	24	Point Beach 1	0	8	Surry 1	0	26
Calvert Cliffs 1	0	12	Indian Point 3	2	32	Point Beach 2	0	16	Surry 2	0	32
Calvert Cliffs 2	0	15	Kewaunee	0	26	Prairie Island 1	0	3	Three Mile Isl 1	0	6
Catawba 1	0	41	Maine Yankee	0	23	Prairie Island 2	0	7	Vogtle 1	0	103
Catawba 2	0	89	McGuire 1	0	45	Robinson 2	1	28	Vogtle 2	0	45
Comanche Pk 1	0	66	McGuire 2	0	44	Salem 1	0	24	Waterford 3	0	38
Comanche Pk 2	0	14	Millstone 2	1	11	Salem 2	0	32	Wolf Creek	0	51
Cook 1	0	18	Millstone 3	0	54	San Onofre 2	0	13	Zion 1	0	13
Cook 2	0	36	North Anna 1	0	20	San Onofre 3	0	17	Zion 2	0	8

Thus, the method yields both an estimate of the population variability and plant-specific estimates at each plant.

The method as just explained underestimates the uncertainty in the answers, because it treats  $\hat{g}$  as if it were equal to the true distribution  $g$ . Therefore, the best implementations of the empirical Bayes method add a final adjustment to Step 2, which makes the plant-specific posterior distributions somewhat more diffuse. This largely accounts for the inaccuracy in equating  $\hat{g}$  to  $g$ .

The name "empirical Bayes" follows from the two steps. The plant-specific estimates are found in a Bayesian way, by updating a prior distribution with plant-specific data. But the prior is not based on prior data or on prior belief, but instead is based on the existing data — the prior is determined empirically.

Step 1 can be carried out in a simple way only if the distributions have convenient forms. Thus, parametric empirical Bayes estimation assumes that  $g$  is conjugate to the distribution of the data at any plant. That is,  $g$  is a gamma distribution when the data are initiating events or other data from a Poisson process, and  $g$  is a beta distribution when the data are failures on demand. This is a limitation to the method. One reason for introducing the hierarchical Bayes method, in Section 8.2.3, is to overcome this limitation.

Some people might object that the method double counts the data. It uses the data to decide on the prior

distribution, and then it uses the same data again to update the prior to obtain the plant-specific estimates. There are two responses to this. (1) The objection is not important in practice, unless the number of plants in the study is very small, or if a small number of plants dominate the data. If no single plant contributes much to the estimate of  $g$ , then there is very little double counting that influences the final estimate for that plant. (2) The hierarchical Bayes method, given in Section 8.2.3, will avoid this difficulty entirely.

For failures on demand, Martz et al. (1996) give a tutorial on the empirical Bayes method, illustrated with NPP data. Siu and Kelly (1998) also explain the method as part of their tutorial article. Carlin and Louis (2000) give a full treatment, including worked-out examples.

### 8.2.2.2 MLE Equations for the Gamma-Poisson Model

The gamma-Poisson model is used for initiating events. The data at plant  $i$  consist of a count of events,  $x_i$ , in time  $t_i$ . Conditional on the plant-specific parameter  $\lambda_i$ , it is assumed that  $x_i$  was generated from a Poisson( $\lambda_i t_i$ ) distribution. However  $\lambda_i$  was generated from the distribution  $g$ , which is assumed to be gamma( $\alpha$ ,  $\beta$ ). The equations for the MLEs of  $\alpha$  and  $\beta$  are now given.

The *conditional distribution* of  $X$ , conditional on  $\lambda$ , is Poisson. However, the *unconditional distribution* of  $X$ , when  $\lambda$  might be any value generated by the population-variability distribution  $g$ , is more complicated. It can be shown that the unconditional distribu-

tion equals the distribution conditional on  $\lambda$ , averaged over the possible values of  $\lambda$ . In equation form, this is

$$\Pr(X = x | \alpha, \beta) = \int \Pr(X = x | \lambda) g(\lambda | \alpha, \beta) d\lambda.$$

Substituting the formulas for the Poisson and gamma distributions, it can be shown that this equals

$$\Pr(X = x | \alpha, \beta) = \frac{\Gamma(\alpha + x)}{n! \Gamma(\alpha)} (t / \beta)^x (1 + t / \beta)^{-(\alpha + x)}. \quad (8.1)$$

As mentioned in Section 6.2.3.5, this is the gamma-Poisson distribution, also often called the negative binomial distribution. This distribution is not conditional on  $\lambda$ . Therefore,  $\lambda$  does not appear in this expression. Instead, it is the probability of seeing  $x$  events in time  $t$  at a plant with a randomly assigned  $\lambda$ .

To find the MLEs of  $\alpha$  and  $\beta$ , write the product of terms of the form of Expression 8.1, using values  $(x_i, t_i)$  for  $i$  from 1 to  $m$ . That product is the joint unconditional likelihood of the data. Take the logarithm of that expression and maximize it. There are several ways to accomplish this.

One approach is to maximize it numerically as a function of two variables, using some version of Newton's method. This is the approach of Siu and Kelly (1998). The derivatives of the log-likelihood, used in performing the maximization, are given below, as stated by Engelhardt (1994). Here  $\ln L$  denotes the logarithm of the likelihood.

$$\frac{\partial}{\partial \alpha} \ln L = \sum_{i=1}^m [\psi(\alpha + x_i) - \psi(\alpha) - \ln(1 + t_i / \beta)]$$

$$\frac{\partial}{\partial \beta} \ln L = -\frac{1}{\beta} \sum_{i=1}^m \left[ x_i - \frac{\alpha + x_i}{\beta + t_i} t_i \right].$$

The function  $\psi$  is the **digamma function**,  $\psi(u) = (d/du) \ln \Gamma(u)$ . It is built into many computer packages. Because  $x_i$  is an integer, the expression involving  $\psi$  can be rewritten as

$$\psi(\alpha + x_i) - \psi(\alpha) = \sum_{j=1}^{x_i} \frac{1}{\alpha + j - 1}. \quad (8.2)$$

A second approach reduces the problem to solving one equation, as follows. At the maximum of the log-likelihood, the two derivatives are equal to zero. Therefore, do the following.

Set the two derivatives equal to zero. The solutions, to be found, are  $\hat{\alpha}$  and  $\hat{\beta}$ .

Solve the second equation for  $\hat{\alpha}$ , as a function of  $\hat{\beta}$ . Substitute this expression into the first equation.

Solve the resulting equation numerically for  $\hat{\beta}$ .

Calculate  $\hat{\alpha}$  from the numerical value of  $\hat{\beta}$ .

The necessary equations to carry out these steps are the following. The equation for  $\hat{\alpha}$ , as a function of  $\hat{\beta}$ , is

$$\hat{\alpha} = \left( \sum_{i=1}^m \frac{x_i \hat{\beta}}{\hat{\beta} + t_i} \right) / \left( \sum_{i=1}^m \frac{t_i}{\hat{\beta} + t_i} \right). \quad (8.3)$$

Substitute Equation 8.3 into

$$\sum_{i=1}^m [\psi(\hat{\alpha} + x_i) - \psi(\hat{\alpha}) - \ln(1 + t_i / \hat{\beta})] = 0$$

and solve that equation numerically for  $\hat{\beta}$ . Having obtained the numerical value of  $\hat{\beta}$ , find  $\hat{\alpha}$  from Equation 8.3.

Sometimes the equations do not have a solution. If the plants do not appear to differ much — for example, the naive plant-specific estimates  $x_i/t_i$  are all similar — the maximum likelihood estimate of  $g$  may be degenerate, concentrated at a single point. That says that the plants appear to have a single common  $\lambda$ . Engelhardt (1994) recommends aborting the estimation process, not trying to fit a model, if the estimate of  $\beta$  becomes greater than  $\sum t_i$  during the iterations. The population-variability distribution  $g$  would be gamma( $\alpha, \beta$ ), with the second parameter greater than  $\sum t_i$ . But simply pooling the data (and using a Jeffreys prior) would result in a gamma( $\sum x_i + 1/2, \sum t_i$ ) posterior distribution. Thus the empirical Bayes distribution would produce a between-plant distribution that is more concentrated (larger

second parameter) than the distribution when the plant data are pooled. This is not the intent of the hierarchical model.

### 8.2.2.3 MLE Equations for the Beta-Binomial Model

The beta-binomial model is used for failures on demand. The data at plant  $i$  consist of a count of failures,  $x_i$ , and demands,  $n_i$ . Conditional on the plant-specific parameter  $p_i$ , it is assumed that  $x_i$  was generated from a binomial( $n_i, p_i$ ) distribution. However  $p_i$  was generated from the distribution  $g$ , which is assumed to be beta( $\alpha, \beta$ ). The equations for the MLEs of  $\alpha$  and  $\beta$  are now given.

The *conditional distribution* of  $X$ , conditional on  $p$ , is binomial. However, the *unconditional distribution* of  $X$ , when  $p$  might be any value generated by the population-variability distribution  $g$ , equals the distribution conditional on  $p$ , averaged over the possible values of  $p$ . That is,

$$\Pr(X = x|\alpha, \beta) = \int \Pr(X = x|p)g(p|\alpha, \beta)dp.$$

Substituting the formulas for the binomial and beta distributions, and using some standard identities relating the beta function and the gamma function, it can be shown that this equals

$$\Pr(X = x) = \frac{n!}{x!(n-x)!} \frac{\Gamma(\alpha+x)}{\Gamma(\alpha)} \frac{\Gamma(\beta+n-x)}{\Gamma(\beta)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)}.$$

This is called the beta-binomial distribution. If both  $x$  and  $n$  are integers, this can be rewritten without the gamma function, as follows.

$$\Pr(X = x) = \frac{n!}{x!(n-x)!} \prod_{j=0}^{x-1} (\alpha + j) \prod_{j=0}^{n-x-1} (\beta + j) / \prod_{j=0}^{n-1} (\alpha + \beta + j).$$

As just stated, this is valid if  $x$  and  $n$  are integers. Are they not always integers? No, they are not, if the data set only gives an *estimate* of the number of demands, which is not necessarily an integer. In that case, the

expression with the gamma function is the only one that can be used.

The likelihood is the product of terms of one of these forms, containing values  $(x_i, n_i)$  for  $i = 1$  to  $m$ . To find the MLE, take the logarithm of the likelihood and maximize it.

The maximization can be done in a variety of ways. One approach, following Atwood (1994), does not deal with  $\alpha$  and  $\beta$  directly. Instead, it reparameterizes, working with

$$\delta = \alpha + \beta \quad \text{and} \\ \mu = \alpha / \delta.$$

The intuitive reason for this reparameterization is that  $\mu$  is the mean of the binomial distribution, and in most models the mean is one of the easiest things to estimate. The letter  $\delta$  was chosen as a mnemonic for "dispersion," because the variance of the binomial distribution is  $\mu(1 - \mu)/(\delta+1)$ . Thus,  $\delta$  is related directly to the variance. Recall that in Section 6.3.2.2.2, the prior and posterior values of  $\alpha + \beta$ , which we are calling  $\delta$  here, were interpreted as the prior and posterior number of demands.

After working with  $\mu$  and  $\delta$ , and finding the MLEs of these parameters, we will translate back to find the MLEs of  $\alpha$  and  $\beta$ , using the equations

$$\alpha = \mu\delta \\ \beta = (1 - \mu)\delta.$$

The MLE is found by setting the derivatives with respect to  $\mu$  and  $\delta$  to zero. After some manipulation, the equations can be expressed as

$$\sum_{j=1}^m \{\psi(\mu\delta + x_i) - \psi(\mu\delta)\} \tag{8.4a}$$

$$= \sum_{i=1}^m \{\psi((1-\mu)\delta + n_i - x_i) - \psi((1-\mu)\delta)\}$$

$$\sum_{j=1}^m \{\psi((1-\mu)\delta + n_i - x_i) - \psi((1-\mu)\delta)\} \tag{8.4b}$$

$$= \sum_{j=1}^m \{\psi(\delta + n_i) - \psi(\delta)\}.$$

Here  $\psi$  is the digamma function, the derivative of  $\ln\Gamma$ , just as in Section 8.2.2.2. If  $x_i$  and  $n_i$  are integers for all  $i$ , Equation 8.2 can be used to rewrite Equation 8.4 as

$$\begin{aligned} & \sum_{i=1}^m \left[ \sum_{j=0}^{x_i-1} \frac{1}{\mu\delta + j} \right] \\ &= \sum_{i=1}^m \left[ \sum_{j=0}^{n_i-x_i-1} \frac{1}{(1-\mu)\delta + j} \right] \end{aligned} \quad (8.5a)$$

and

$$\begin{aligned} & \sum_{i=1}^m \left[ \sum_{j=0}^{n_i-x_i-1} \frac{1}{(1-\mu)\delta + j} \right] \\ &= \sum_{i=1}^m \left[ \sum_{j=0}^{n_i-1} \frac{1}{\delta + j} \right]. \end{aligned} \quad (8.5b)$$

The Equations 8.4 or 8.5 must be solved for  $\mu$  and  $\delta$ . One method, suggested by Atwood (1994) is to begin with a trial value of  $\delta$ . Solve Equation 8.4a or 8.5a numerically for  $\mu$ . This typically needs only a few iterations. Substitute this value into Equation 8.4b or 8.5b, and solve the resulting equation for  $\delta$ . Continue alternating between the two equations until the estimates converge.

The estimates do not always converge. If the plants have very similar data, the maximum likelihood estimate of  $g$  may be concentrated at a single point, degenerate. This would say that the plants all have the same  $p$ . Atwood (1994) recommends aborting the iterations if the value of  $\delta$  becomes greater than  $\Sigma n_i$ . Allowing  $\delta$  to be greater than  $\Sigma n_i$  would produce a population-variability distribution that is more concentrated than the distribution corresponding to simply pooling the data.

#### 8.2.2.4 Adjustment for Uncertainty in the Estimate of $g$

As mentioned above, the method as presented so far underestimates the uncertainty in the final answers, because it does not account for the uncertainty in  $\hat{g}$ . Kass and Steffey (1989) present a correction to the

final estimates, to approximately account for this uncertainty. The plant-specific posterior means are unchanged, but the posterior variances are increased somewhat. Kass and Steffey state that the adjustment is very important if there are few data subsets (plants, in the present discussion) and many observations (initiating events or demands). Conversely, the adjustment is unimportant when there are many data subsets and few observations. No harm is done by automatically applying the adjustment in every case. The formulas are given here.

##### 8.2.2.4.1 Equations for the Gamma-Poisson Case

It is computationally advantageous to reparameterize in terms of  $\mu = \alpha/\beta$  and  $\alpha$ . Denote the maximum likelihood estimators for the hyperparameters  $\mu$  and  $\alpha$  by  $\hat{\mu}$  and  $\hat{\alpha}$ . It turns out that these estimators are asymptotically uncorrelated, causing certain terms in the formulas to be zero.

The method as given in Section 8.2.2.2, finds the estimates  $\hat{\alpha}$  and  $\hat{\beta}$ , which can be reparameterized as  $\hat{\mu} = \hat{\alpha} / \hat{\beta}$  and  $\hat{\alpha}$ . These are the estimated parameters of the gamma prior distribution  $g$ . The method then updates the estimated prior by plant-specific data. The posterior distribution of  $\lambda_i$  is also a gamma distribution, with posterior mean

$$E_{\text{post}}(\lambda_i) = \frac{\hat{\alpha} + x_i}{\hat{\alpha} / \hat{\mu} + t_i} \quad (8.6)$$

and posterior variance

$$\text{var}_{\text{post}}(\lambda_i) = \frac{\hat{\alpha} + x_i}{(\hat{\alpha} / \hat{\mu} + t_i)^2}.$$

The Kass-Steffey adjustment increases the variance to

$$\begin{aligned} & \text{var}_{\text{post}}(\lambda_i) \\ &+ \left[ \frac{\partial E_{\text{post}}(\lambda_i)}{\partial \hat{\mu}} \right]^2 \text{var}(\hat{\mu}) + \left[ \frac{\partial E_{\text{post}}(\lambda_i)}{\partial \hat{\alpha}} \right]^2 \text{var}(\hat{\alpha}). \end{aligned} \quad (8.7)$$

A covariance term would normally also be present, but this term is zero when the parameterization is in terms of  $\mu$  and  $\alpha$ .

We now develop the formulas that must be substituted into Equation 8.7. From Equation 8.6, the derivatives of  $E_{\text{post}}(\lambda_i)$  are

$$\frac{\partial E_{\text{post}}(\lambda_i)}{\partial \hat{\mu}} = -\frac{\hat{\alpha}(\hat{\alpha} + x_i)}{(\hat{\alpha} + \hat{\mu}t_i)^2} \quad \text{and}$$

$$\frac{\partial E_{\text{post}}(\lambda_i)}{\partial \hat{\alpha}} = -\frac{\hat{\mu}(x_i - \hat{\mu}t_i)}{(\hat{\alpha} + x_i\hat{\mu})^2}.$$

From the asymptotic theory of maximum likelihood estimation, the variances are found as follows. The information matrix,  $J$ , is the negative of the matrix of second derivatives of the log-likelihood:

$$J = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} = -E \begin{bmatrix} \frac{\partial^2 L}{\partial \mu^2} & \frac{\partial^2 L}{\partial \mu \partial \alpha} \\ \frac{\partial^2 L}{\partial \mu \partial \alpha} & \frac{\partial^2 L}{\partial \alpha^2} \end{bmatrix} \quad (8.8)$$

evaluated at  $\hat{\mu}$  and  $\hat{\alpha}$ . The inverse of this matrix is asymptotically equal to the variance-covariance matrix:

$$\begin{bmatrix} \text{var}(\hat{\mu}) & \text{cov}(\hat{\mu}, \hat{\alpha}) \\ \text{cov}(\hat{\mu}, \hat{\alpha}) & \text{var}(\hat{\alpha}) \end{bmatrix}.$$

When this is carried out, we have

$$J_{11} = \frac{\hat{\alpha}}{\hat{\mu}} \sum_{i=1}^m \frac{t_i}{\hat{\alpha} + \hat{\mu}t_i}$$

$$J_{22} = \sum_{i=1}^m [\psi'(\hat{\alpha}) - \psi'(\hat{\alpha} + x_i)] - \frac{\hat{\mu}}{\hat{\alpha}} \sum_{i=1}^m \frac{t_i}{\hat{\alpha} + \hat{\mu}t_i}$$

$$J_{12} = J_{21} = 0.$$

If all the  $x_i$  values are integers, the difference of  $\psi$  terms can be rewritten using Equation 8.2, and the

difference of derivatives  $\psi'$  can be written algebraically, avoiding use of a special function.

Because the off-diagonal elements are zero, the inverse consists of the inverses of the diagonal terms, and it follows that

$$\text{var}(\hat{\mu}) = 1 / J_{11} \quad \text{and}$$

$$\text{var}(\hat{\alpha}) = 1 / J_{22}.$$

The final step of the empirical Bayes method is to substitute the expressions just found into the Kass-Steffey adjustment for the posterior variance, Equation 8.7. Then approximate the posterior distribution by a gamma distribution having the original posterior mean and the adjusted posterior variance. An example will be given below.

#### 8.2.2.4.2 Equations for the Beta-Binomial Case

As in Section 8.2.2.3, we parameterize in terms of  $\mu = \alpha/(\alpha + \beta)$  and  $\delta = \alpha + \beta$ . Denote the maximum likelihood estimators by  $\hat{\mu}$  and  $\hat{\delta}$ . Although these estimators are asymptotically not exactly uncorrelated, as was the case for the gamma-Poisson model, they are nearly uncorrelated. The equations are given by Atwood (1995).

The method as given in Section 8.2.2.3 finds the estimates  $\hat{\mu}$  and  $\hat{\delta}$ , the estimated parameters of the beta prior distribution  $g$ . The method then updates the estimated prior by plant-specific data. The posterior distribution of  $p_i$  is also a beta distribution, with posterior mean

$$E_{\text{post}}(p_i) = \frac{\hat{\mu}\hat{\delta} + x_i}{\hat{\delta} + n_i} \quad (8.9)$$

and posterior variance

$$\text{var}_{\text{post}}(p_i) = E_{\text{post}}(p_i)[1 - E_{\text{post}}(p_i)] / (\hat{\delta} + n_i + 1).$$

The Kass-Steffey adjustment increases the variance to

$$\begin{aligned} & \text{var}_{\text{post}}(p_i) \\ & + \left[ \frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\mu}} \right]^2 \text{var}(\hat{\mu}) + \left[ \frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\delta}} \right]^2 \text{var}(\hat{\delta}) \quad (8.10) \\ & + \left[ \frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\mu}} \right] \left[ \frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\delta}} \right] \text{cov}(\hat{\mu}, \hat{\delta}) . \end{aligned}$$

From Equation 8.9, the two derivatives of  $E_{\text{post}}(p_i)$  are

$$\begin{aligned} \frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\mu}} &= \frac{\hat{\delta}}{\hat{\delta} + n_i} \\ \frac{\partial E_{\text{post}}(p_i)}{\partial \hat{\delta}} &= -\frac{x_i - n_i \hat{\mu}}{(\hat{\delta} + n_i)^2} . \end{aligned}$$

The variances and covariance are found from inverting the matrix in Equation 8.8, with  $\delta$  used now instead of  $\alpha$ . The terms can be found as follows. Define

$$\begin{aligned} S_1 &= \sum_{i=1}^m \left[ \psi'(\hat{\mu}\hat{\delta}) - \psi'(\hat{\mu}\hat{\delta} + x_i) \right] \\ S_2 &= \sum_{i=1}^m \left[ \psi'((1-\hat{\mu})\hat{\delta}) - \psi'((1-\hat{\mu})\hat{\delta} + n_i - x_i) \right] \\ S_3 &= \sum_{i=1}^m \left[ \psi'(\hat{\delta}) - \psi'(\hat{\delta} + n_i) \right] . \end{aligned}$$

Then the information matrix is given by

$$\begin{aligned} J_{11} &= \hat{\delta}^2(S_1 + S_2) \\ J_{22} &= \hat{\mu}^2 S_1 + (1-\hat{\mu})^2 S_2 - S_3 \\ J_{12} &= J_{21} = \hat{\mu} \hat{\delta} S_1 - (1-\hat{\mu})\hat{\delta} S_2 . \end{aligned}$$

The variances and covariance follow from standard formulas for inverting a 2x2 matrix. Define the determinant

$$D = J_{11}J_{22} - (J_{12})^2 .$$

Then we have

$$\begin{aligned} \text{var}(\hat{\mu}) &= J_{22} / D \\ \text{var}(\hat{\delta}) &= J_{11} / D \\ \text{cov}(\hat{\mu}, \hat{\delta}) &= -J_{12} / D . \end{aligned}$$

To complete the Kass-Steffey adjustment, substitute the above expressions into the equation for the adjusted variance, Equation (8.10). Then find the beta distribution having the posterior mean and the adjusted posterior variance. Use this as the approximate posterior distribution for  $p_i$ .

### 8.2.2.5 Application to Examples

The parametric empirical Bayes method is now illustrated with Examples 8.1 and 8.2. First, a chi-squared test will be performed, to test whether the plants can be pooled. In each example, the difference between plants will be found to be either statistically significant or very close to statistically significant. Then plant-specific confidence intervals will be found, each based only on the data for a single plant. Then the empirical Bayes method will be used, and the resulting 90% credible intervals will be shown, based on the plant-specific posterior distributions, using the Kass-Steffey adjustment. The plant-specific intervals resulting from the empirical Bayes analysis will be compared to the (less sophisticated) plant-specific confidence intervals.

#### 8.2.2.5.1 Example 8.1, Initiating Events

To test poolability of the plants in Example 8.1, the Pearson chi-squared test was performed, as presented in Section 6.2.3.1.2. The test statistic  $X^2$  was equal to 378.5. Because there were 66 plants, the value of  $X^2$  should be compared to a chi-squared distribution with 65 degrees of freedom. The value of 378.5 is very far out in the tail of the chi-squared distribution, off the table. Thus, the evidence is extremely strong, beyond question, that the plants do not all have the same  $\lambda$ .

To show this graphically, 90% confidence intervals for  $\lambda$  were plotted, with each confidence interval based on the data from a single plant. These are shown in Figure 8.2. Because  $t$  has been written in terms of 1000 critical hours, the units of  $\lambda$  are events per thousand critical hours. The order of the plants is not alphabetical, but instead is by decreasing estimate of  $\lambda$ . Because the example has so many plants, only the plants with the 10 highest and 10 lowest estimates are individually identified in the figure.

8. Parameter Estimation Using Existing Data Sources DRAFT

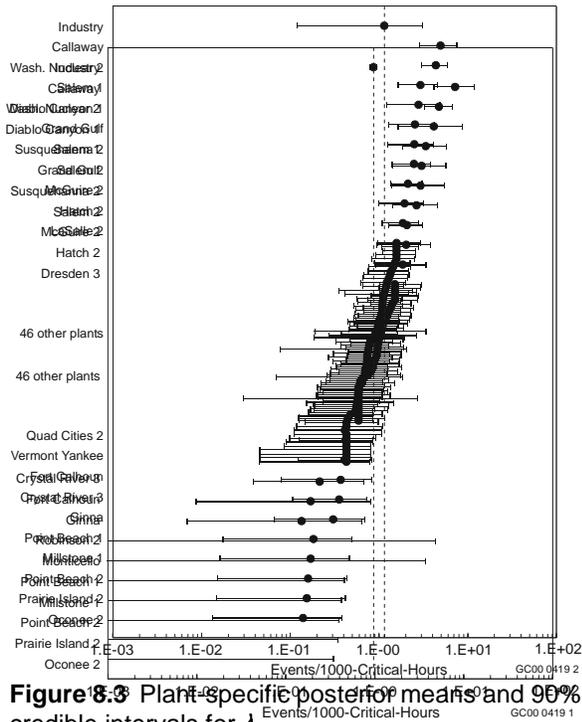


Figure 8.2 Plant-specific MLEs and 90% confidence intervals for  $\lambda$ .

A 90% confidence interval is plotted at the top of the plot for the pooled industry data. Of course the interval is too short to be realistic, because pooling of the data is completely unjustified in this example. In fact, the interval is too short even to be visible. Nevertheless, the overall pooled mean is a useful reference value for comparison with the individual plant results. For this reason, a vertical dashed line is drawn through the pooled mean.

Because the plant-specific estimates differ so greatly, the figure uses a logarithmic scale. This means that some of the point estimates, those with zero values, cannot be plotted.

Figure 8.3 is based on the empirical Bayes method. For each plant, the mean and 90% credible interval are shown, based on the posterior distribution and the Kass-Steffey adjustment. An interval for the industry is also plotted, a 90% interval based on the mean of this distribution is also shown, and a vertical dashed line is drawn through the mean.

Those who wish to make some detailed comparisons can find a few numerical values listed in Tables 8.1 through 8.3.

Table 8.1 Portion of frequentist analysis results for Example 8.1.

Plant	$x, t$	MLE and 90% conf. int. <sup>a</sup>
Industry	361, 374.229	(0.883, 0.965, 1.05)
Callaway	12, 1.5038	(4.60, 7.98, 12.9)
Wash. Nuc. 2	23, 4.3643	(3.60, 5.27, 7.47)
Diablo Can. 1	5, 1.0846	(1.82, 4.61, 9.69)
Salem 1	10, 2.6738	(2.03, 3.74, 6.34)
Grand Gulf	7, 2.0896	(1.57, 3.35, 6.293)
Pt. Beach 2	0, 7.5442	(0.0, 0.0, 0.397)
Prairie Isl. 2	0, 7.8440	(0.0, 0.0, 0.382)
Oconee 2	0, 8.7840	(0.0, 0.0, 0.341)

a. Format is (lower bound, MLE, upper bound).

The most striking feature seen by comparing the two figures is that the empirical Bayes estimates vary less from each other than do the MLEs. Of course, if a plant has no events, the lower confidence limit is zero, and any Bayesian method will give a non-zero lower limit. Such a difference appears enormous when plotted on a logarithmic scale. However, the effect is seen not only at the bottom of Figures 8.2 and 8.3 but also at the top: the largest plant-specific posterior means are closer to the industry average than are the corresponding MLEs. Indeed, just as was seen for Bayes methods in general, the empirical Bayes method gives posterior means that are between the MLEs and the industry (i.e., the prior) mean.

**Table 8.2 Portion of empirical Bayes analysis, without Kass-Steffey adjustment.**

Plant	Gamma parameters, $\alpha$ , $\beta$	Posterior mean and 90% credible interval <sup>a</sup>
Industry	1.39, 1.211	(0.118, 1.15, 3.07)
Callaway	13.39, 2.715	( 2.94, 4.93, 7.34)
Wash. Nuc. 2	24.39, 5.575	( 3.03, 4.37, 5.93)
Salem 1	11.39, 3.885	( 1.66, 2.93, 4.49)
Diablo Can. 1	6.39, 2.296	( 1.25, 2.78, 4.81)
Grand Gulf	8.39, 3.301	( 1.29, 2.54, 4.14)
Pt. Beach 2	1.39, 8.755	(0.0164, 0.159, 0.424)
Prairie Isl. 2	1.39, 9.055	(0.0158, 0.154, 0.410)
Oconee 2	1.39, 9.995	(0.0143, 0.139, 0.372)

a. Format is (lower bound, mean, upper bound).

**Table 8.3 Portion of empirical Bayes analysis, with Kass-Steffey adjustment.**

Plant	Gamma parameters, $\alpha$ , $\beta$	Posterior mean and 90% credible interval <sup>a</sup>
Industry	1.39, 1.211	(0.118, 1.15, 3.07)
Callaway	12.13, 2.460	( 2.86, 4.93, 7.47)
Wash. Nuc. 2	23.40, 5.348	( 3.00, 4.37, 5.96)
Salem 1	11.03, 3.762	( 1.65, 2.93, 4.52)
Diablo Can. 1	6.08, 2.185	( 1.22, 2.78, 4.86)
Grand Gulf	8.15, 3.204	( 1.27, 2.54, 4.16)
Pt. Beach 2	1.33, 8.382	(0.0151, 0.159, 0.431)
Prairie Isl. 2	1.33, 8.665	(0.0146, 0.154, 0.417)
Oconee 2	1.33, 9.554	(0.0132, 0.139, 0.378)

a. Format is (lower bound, mean, upper bound).

The order of the plants is not exactly the same in Figures 8.2 and 8.3. The reason is that estimates for different plants are pulled toward the industry mean by different amounts. This can cause some rearrangement of the ranking of the plants. For example Salem 1 and Diablo Canyon 1 appear in reverse order in the two figures (and in Tables 8.1 and 8.3). The reason is that Diablo Canyon 1 has about half as much data (5 events in 1085 hours) as Salem 1 (10 events in 2674 hours). Therefore Diablo Canyon 1 is pulled more toward the industry mean.

We notice also, by comparing Tables 8.2 and 8.3, that the Kass-Steffey adjustment is very small in this example. The data set is so large that  $g$  can be estimated quite well. Any error in equating the

estimate to the true distribution is minor, as reflected in the small effect of the Kass-Steffey adjustment.

An empirical Bayes estimator is sometimes called a **shrinkage estimator**, or a **shrinker**, because the method pulls all the MLEs in towards the industry mean. The intuitive justification for such shrinkage is the recognition that extreme data are produced by a combination of extreme parameters and luck. Thus, the plant with the highest observed frequency appears so extreme because of a combination of large  $\lambda$  and some bad luck. Likewise, the plant with the best performance, Oconee 2, which ran for 366 days straight without a single scram, can attribute its perfect performance to a combination of low  $\lambda$  and good luck. The empirical Bayes method tries to remove the effect of luck when estimating the  $\lambda$  values.

As always when performing a statistical analysis, one should try to combine statistical calculations with engineering understanding. It is known that newly licensed plants sometimes experience more initiating events than they do after acquiring more experience. This was mentioned in the discussion of Example 2.1, and it is seen again here.

Of the 66 plants, 9 did not have their commercial starts until 1984 or later. These 9 young plants are all among the 19 with the highest event frequencies. For example, consider the two plants with the highest estimated frequencies, based on the 1984 data. Both of these plants had their commercial starts in December 1984.

The hierarchical model is intended for plants that are nominally identical. The variability among the plants is unexplained, and modeled as random. An important assumption is that each plant is assigned a  $\lambda$  from the same distribution,  $g$ . As a result, each plant is as likely as any other to have a large  $\lambda$  or a small  $\lambda$ . The parameters  $\lambda_i$  are called **exchangeable** if any  $\lambda_i$  is as likely as any other to correspond to a particular plant. As discussed by Gelman et al. (1995, Section 5.2), when we know nothing about the plants, exchangeability is a reasonable assumption. When we know the ages of the plants, however, exchangeability is no longer reasonable. The most immature plants are expected to have larger values of  $\lambda$ .

Thus, the analysis of Example 8.1 really should be modified. One way would be to separate the plants into two groups, mature and immature, and perform an empirical Bayes analysis on each group. A more sophisticated way would be to try to model the age of the plant as a continuous explanatory variable. Then the otherwise random  $\lambda_i$  would be multiplied by some function of the age of plant  $i$ , a large factor for immature plants and a smaller factor for mature plants. Such models are beyond the scope of this handbook, however.

**8.2.2.5.2 Example 8.2, AFW Segment Failures to Start**

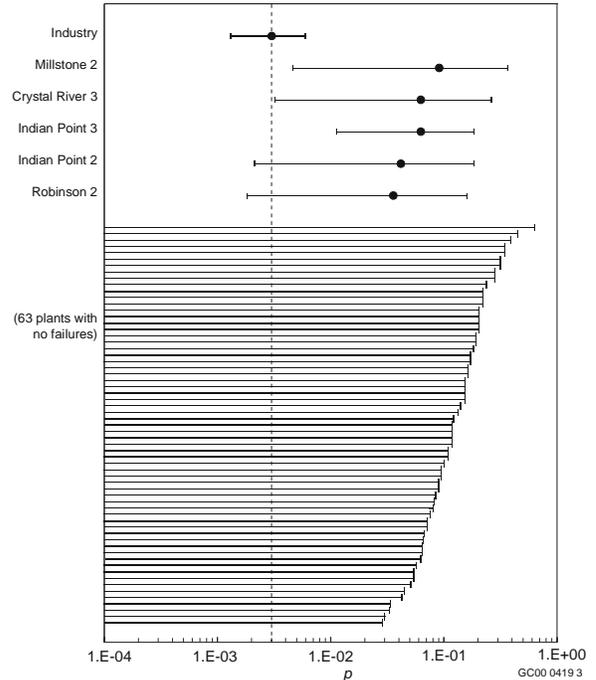
This example has 68 plants, with 6 failures in 1993 demands.

Poloski et al. (1998) perform a chi-squared test to see if  $p$  is the same at all plants. This test is explained in Section 6.3.3.1.2. The test statistic  $X^2$  equals 113.1. Because there are 68 plants, the degrees of freedom is 67. The reported p-value is 0.0004, meaning that 113.1 is the 99.96th percentile of the chi-squared distribution with 67 degrees of freedom. However, the data set has so few failures that the chi-squared distribution is not a good approximation for the distribution of  $X^2$ . The expected number of failures at a plant with 30 demands (a typical number of demands) is  $6 \times 30 / 1993 = 0.09$ . This is much less than the recommended minimum of 0.5. Therefore, the calculated p-value is quite suspect.

Poloski et al. choose to model between-plant differences with a hierarchical model, partly because of the above calculated p-value, and partly on the grounds that modeling possible differences between plants is more conservative (reflecting more variability) than simply pooling the data.

The empirical Bayes estimate of the population-variability distribution,  $g$ , is a beta(0.137, 36.34) distribution. The mean of this distribution is  $3.77E-3$ . The 5th and 95th percentiles are  $5.99E-12$  and  $2.12E-2$ . Note, the first parameter of the distribution is very small, well below 0.5. As a result, the 5th percentile is unrealistically small, implying less than one failure in one hundred billion demands. This unrealistic lower bound carries over to the posterior distribution of all plants that have zero failures. Figures 8.4 and 8.5 are the analogues of Figures 8.2 and 8.3.

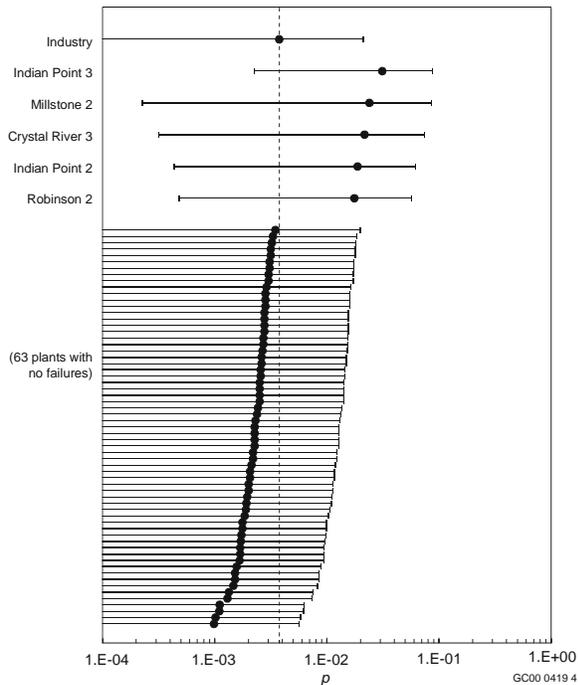
The first figure shows plant-specific MLEs and 90% confidence intervals, while the second shows the results of the empirical Bayes analysis, posterior means and 90% credible intervals. Only the five plants that had failures are individually identified in the figures.



**Figure 8.4** Plant-specific MLEs and 90% confidence intervals for  $p$ .

Some numerical details are given in Table 8.4 through 8.6.

Just as with Example 8.1, the empirical Bayes method pulls the plant-specific MLEs toward the industry mean. This is seen in both the figures and the tables. Also, the Kass-Steffey adjustment increases the width of the plant-specific intervals by a noticeable amount, for example, by about 30% for Indian Point 3. This is best seen by comparing Tables 8.5 and 8.6. This comparison shows that the estimates of the parameters have noticeable uncertainty, even if the assumption of a beta distribution is accepted.



**Figure 8.5** Plant-specific posterior means and 90% credible intervals for  $p$ .

**Table 8.4** Portion of frequentist analysis results for Example 8.2.

Plant	$x, d$	MLE and 90% conf. int. <sup>a</sup>
Industry	6, 1993	(1.31E-3, 3.01E-3, 5.93E-3)
Millstone 2	1, 11	(4.65E-3, 9.09E-2, 3.64E-1)
Crystal River 3	1, 16	(3.20E-3, 6.25E-2, 2.64E-1)
Indian Point 3	2, 32	(1.12E-2, 6.25E-2, 1.84E-1)
Indian Point 2	1, 24	(2.13E-3, 4.17E-2, 1.83E-1)
Robinson 2	1, 28	(1.83E-3, 3.57E-2, 1.59E-1)
Prairie Island 1	0, 3	(0.0, 0.0, 3.32E-1)
Vogtle 1	0, 103	(0.0, 0.0, 2.87E-2)

a. Format is (lower bound, MLE, upper bound).

Poloski et al. (1998) carry out the above empirical Bayes analysis, but they do not report the calculated lower bounds for plants that experience no failures. They recognize that those lower bounds are unrealistically small, and that such calculated values are an artifact of using a beta distribution. Therefore, they simply report that the lower bound is  $<1E-8$ . The next section gives a way to avoid entirely the assumption of a beta distribution.

**Table 8.5** Portion of empirical Bayes analysis, without Kass-Steffey adjustment.

Plant	Beta parameters, $\alpha, \beta$	Posterior mean and 90% credible interval <sup>a</sup>
Industry	0.137, 36.34	(6.0E-11, 3.77E-3, 2.12E-2)
Indian Point 3	2.137, 68.34	(6.14E-3, 3.12E-2, 7.15E-2)
Millstone 2	1.137, 47.34	(1.70E-3, 2.40E-2, 6.78E-2)
Crystal River 3	1.137, 52.34	(1.53E-3, 2.17E-2, 6.14E-2)
Indian Point 2	1.137, 60.34	(1.33E-3, 1.88E-2, 5.33E-2)
Robinson 2	1.137, 64.34	(1.24E-3, 1.76E-2, 5.01E-2)
Prairie Isl. 1	0.137, 39.34	(5.5E-12, 3.48E-3, 1.96E-2)
Vogtle 1	0.137, 139.3	(1.6E-12, 9.86E-4, 5.52E-3)

a. Format is (lower bound, mean, upper bound).

**Table 8.6** Portion of empirical Bayes analysis, with Kass-Steffey adjustment.

Plant	Beta parameters, $\alpha, \beta$	Posterior mean and 90% credible interval <sup>a</sup>
Industry	0.137, 36.34	(6.0E-11, 3.77E-3, 2.12E-2)
Indian Point 3	1.149, 35.65	(2.27E-3, 3.12E-2, 8.77E-2)
Millstone 2	0.596, 24.29	(2.26E-4, 2.40E-2, 8.54E-2)
Crystal R. 3	0.663, 29.94	(3.16E-4, 2.17E-2, 7.44E-2)
Indian Point 2	0.754, 39.34	(4.34E-4, 1.88E-2, 6.17E-2)
Robinson 2	0.793, 44.14	(4.78E-4, 1.76E-2, 5.69E-2)
Prairie Isl. 1	0.133, 37.98	(2.6E-12, 3.48E-3, 1.97E-2)
Vogtle 1	0.127, 128.9	(2.8E-13, 9.86E-4, 5.59E-3)

a. Format is (lower bound, mean, upper bound).

### 8.2.3 The Hierarchical Bayes Method

### 8.2.3.1 General Approach

In the preceding discussion of parametric empirical Bayes, the unknown hyperparameters of the population-variability (or prior) distribution were estimated by maximum likelihood. The empirical Bayes estimate of the population-variability distribution is the corresponding distribution in which these maximum likelihood estimates have been inserted.

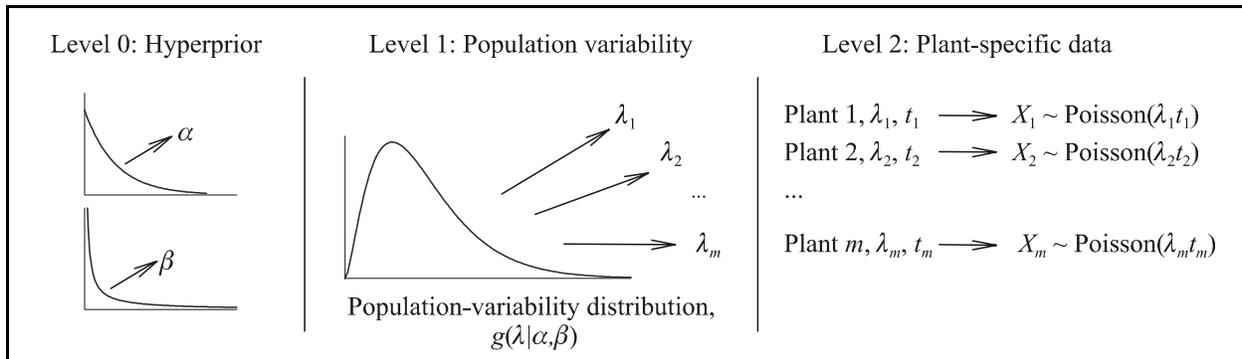
The hierarchical Bayes method is entirely different. It embodies a complete (or full) Bayesian approach to the problem of estimating the unknown population-variability distribution based on the available data. The hierarchical Bayes approach expresses the initial uncertainty (that is, uncertainty before the data are considered) about the unknown hyperparameters using yet another prior, a so-called second-order or **hyper-prior** distribution. For example, in Example 8.1, the population-variability distribution can be a gamma distribution, with parameters (called hyperparameters in this context)  $\alpha$  and  $\beta$ . The distribution  $g$  could also be lognormal with parameters  $\mu$  and  $\sigma^2$ . Any desired distribution can be used, with any parameterization. Figure 8.6 denotes the parameters of  $g$  generically as  $\alpha$  and  $\beta$ . The uncertainty in the state-of-knowledge about the values of  $\alpha$  and  $\beta$  is expressed by a suitably specified joint hyperprior distribution on  $\alpha$  and  $\beta$ . This expands Figure 8.1 to be Figure 8.6. We almost always desire such hyperprior distributions to be diffuse because we almost never have very precise (or informative) information at the hyperprior level of such a model.

Figure 8.6 is drawn showing  $\alpha$  and  $\beta$  with separate distributions. In general, the hyperparameters together have a joint distribution, which does not have to be the product of independent distributions.

In the full Bayesian model all the unknown parameters, including prior-distribution hyperparameters, are assigned prior distributions that express the analyst's initial uncertainty about these parameters. This is known as a **hierarchical Bayes** model. Berger (1985) and Gelman et al. (1995) discuss the basic notions of hierarchical Bayes modeling. In Example 8.1, the parameters of interest to be estimated at the overall population-variability level of the analysis are  $\alpha$  and  $\beta$ , while the plant-specific parameters to be estimated are

the 66  $\lambda_i$  values. Each of these 68 parameters is assigned an appropriate prior distribution in a hierarchical Bayes analysis.

The solution to the hierarchical Bayes method requires conditioning on the data and obtaining the required posterior distributions of all the parameters of interest. This is done using Markov chain Monte Carlo (MCMC) simulation (see Section 8.2.3.3). The desired point and interval estimates of the parameters are then directly



**Figure 8.6** Hierarchical Bayes model for Poisson data.

(and easily) obtained from these posterior distributions. This process will be illustrated for Examples 8.1 and 8.2 in Sections 8.2.3.4 and 8.2.3.5, respectively.

It is well known (Berger 1985 and Gelman et al. 1995) that parametric empirical Bayes can be viewed as an approximation to a full hierarchical Bayes analysis. However, there are several important advantages of hierarchical Bayes over parametric empirical Bayes.

First, parametric empirical Bayes essentially requires the use of a conjugate population-variability distribution in order to obtain the required unconditional distribution of  $X$  in closed form. Because hierarchical Bayes analysis is implemented in practice using Monte Carlo simulation, non-conjugate population-variability distributions can be used as easily as conjugate distributions. For the Poisson example, a lognormal( $\mu, \sigma^2$ ) distribution on  $\lambda$  is as easy as a gamma( $\alpha, \beta$ ) distribution.

Second, when using hierarchical Bayes, there is no need to worry about double counting of the data. The hierarchical model and associated Bayesian analysis ensures that this cannot occur.

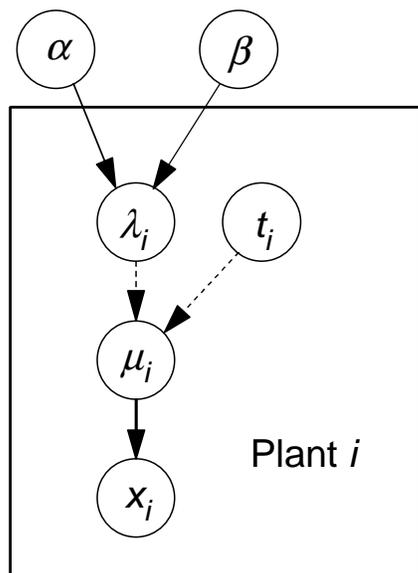
Finally, as mentioned above, the hierarchical Bayes method is conveniently and easily implemented in practice by means of Markov chain Monte Carlo simulation using existing software, which is presently available for free download from the Web (see Section 8.2.3.3.3).

### 8.2.3.2 Directed Graphs

The first conceptual step in a hierarchical Bayes analysis should be to construct a **directed graph** representing the hierarchical Bayes model. Briefly, such a graph represents all quantities as nodes in a directed graph, in which arrows between nodes represent directed influences. A directed graph for Example 8.1 is shown in Figure 8.7, where we have defined  $\mu_i = \lambda_i t_i$ .

Note that a solid arrow indicates a stochastic dependency, while a dashed arrow indicates a logical function. The hierarchical Bayes approach for the gamma-Poisson case proceeds as follows. First specify hyperprior distributions for the two hyperparameters outside the "plant  $i$ " box in Figure 8.7. Inference on the hyperparameters  $\alpha, \beta$  and the scam rate vector  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{66})$  requires that we obtain Monte Carlo samples from the joint posterior  $g(\alpha, \beta, \lambda | \mathbf{x})$ , where the data vector  $\mathbf{x}$  is defined as  $\mathbf{x} = (x_1, x_2, \dots, x_{66})$ . The letter  $g$  is used here to denote both prior and posterior densities. Generate these samples, and then use summary statistics from these samples to obtain the desired estimates, such as point and interval estimates of these parameters.

In order to calculate samples from this joint posterior we must successively sample from the full conditional distributions. That is, we must successively sample from the conditional distribution of each stochastic



**Figure 8.7** Directed graph for the hierarchical Bayes analysis of Example 8.1.

node given all the other stochastic nodes in the graph. However, conditional independence is expeditiously exploited in directed graphs in order to simplify these full conditional distributions. For example, given  $\beta$  and  $\lambda_j$ ,  $\alpha$  in Figure 8.7 is conditionally independent of  $x_j$ .

### 8.2.3.3 Markov Chain Monte Carlo (MCMC) Simulation

**Markov chain Monte Carlo (MCMC)** sampling techniques give the required samples from the joint posterior distribution of all the unknown parameters. The desired hierarchical Bayes point and interval estimates can thus be directly computed from the corresponding simulated sample observations without the need for tedious analytical or numerical calculations. MCMC is a Monte Carlo integration technique that is implemented using Markov chains. MCMC draws these samples by running a cleverly constructed Markov chain for a long period of time. Good introductions to MCMC are provided by Gilks et al. (1996) and Gelman et al. (1995).

In the Poisson example, the required hierarchical Bayes estimates can be obtained by means of Gibbs sampling, a basic MCMC technique that is described next. The

equations are sketched here first. Then a publicly available software package, BUGS, is described. The package implements the equations without requiring the users to understand the details.

#### 8.2.3.3.1 Gibbs Sampling

**Gibbs sampling** is a technique that can be used to generate a random sample from the joint posterior distribution indirectly, provided that we can directly sample each of the full conditional distributions (which are described below).

The Gibbs sampling method, sometimes also called simply the **Gibbs sampler**, is briefly described here. Suppose that we have a set of  $p$  parameters  $\theta_1, \theta_2, \dots, \theta_p$  whose joint posterior distribution  $g(\theta_1, \theta_2, \dots, \theta_p | \mathbf{x})$  is unknown but is of interest to be estimated. This is the usual case when using the hierarchical Bayes method. In Example 8.1, the  $\theta_j$ s consist of the two hyperparameters plus the 66  $\lambda_j$ s, and the number of parameters,  $p$ , is 68.

However, suppose that the full conditional distributions  $g(\theta_i | \theta_j, \mathbf{x}, j \neq i) i = 1, 2, \dots, p$ , are known in the sense that sample values of  $\theta_i$ , conditional on values of  $\theta_j, j \neq i$ , may be generated from these by some appropriate method. Under mild conditions, these conditional distributions uniquely determine the required joint posterior distribution  $g(\theta_1, \theta_2, \dots, \theta_p | \mathbf{x})$ ; hence, they determine all the unconditional marginal distributions  $g(\theta_i | \mathbf{x}), i = 1, 2, \dots, p$ , as well.

The Gibbs sampler generates samples from the required joint distribution as follows:

- (1) Select an arbitrary starting set of values  $\theta_1^0, \dots, \theta_p^0$ . Set  $j = 0$ .
- (2) Draw  $\theta_1^{j+1}$  from  $g(\theta_1 | \theta_2^j, \dots, \theta_p^j, \mathbf{x})$ , then  $\theta_2^{j+1}$  from  $g(\theta_2 | \theta_1^j, \theta_3^j, \dots, \theta_p^j, \mathbf{x})$ , and so on up to  $\theta_p^{j+1}$  from  $g(\theta_p | \theta_1^j, \dots, \theta_{p-1}^j, \mathbf{x})$  to complete one iteration of the sampler.
- (3) Increment  $j$  and repeat Step (2) until  $j+1 = n$ . After  $n$  such iterations of Step (2), we have obtained the sample  $(\theta_1^n, \dots, \theta_p^n)$ .

Under mild conditions, as  $n \rightarrow \infty$  this  $p$ -tuple converges in distribution to the unknown joint posterior distribution  $g(\theta_1, \theta_2, \dots, \theta_p | \mathbf{x})$ .

Typical implementation of the Gibbs sampling algorithm generates an initial "large" number of iterations (called the **burn-in**) until the samples have converged. The burn-in samples are discarded, and the samples generated thereafter are used as sample observations from the joint posterior distribution  $g(\theta_1, \theta_2, \dots, \theta_p | \mathbf{x})$ . Nonparametric density estimators, such as those given in Section 6.7.3, can then be used to approximate the posterior distribution using the post burn-in samples. Examples 8.1 and 8.2 are analyzed in this way in Sections 8.2.3.4 and 8.2.3.5.

In Example 8.1, sixty-eight full conditional distributions are required in order to use the Gibbs sampler:

$$\begin{aligned}
 (1) \quad & g(\alpha | \beta, \lambda, x) \equiv g(\alpha | \beta, \lambda) \\
 & \propto \left[ \frac{\beta^\alpha}{\Gamma(\alpha)} \right]^{66} \left( \prod_{i=1}^{66} \lambda_i \right)^\alpha e^{-\alpha} \\
 (2) \quad & g(\beta | \alpha, \lambda, x) \equiv g(\beta | \alpha, \lambda) \\
 & = \text{gamma} \left( 66\alpha + 0.0625, \sum_{i=1}^{66} \lambda_i + 0.0625 \right) \\
 (3) \quad & g(\lambda_i | \alpha, \beta, x) = \text{gamma}(\alpha + x_i, \beta + t_i), \quad i = 1, \dots, 66.
 \end{aligned}$$

It is easy to sample directly from the gamma distributions. The first distribution, however, the distribution for  $\alpha$ , is not of a familiar form, and in fact is not known fully. It is only known to be proportional to the stated expression, with the normalizing constant unknown. The Metropolis-Hastings method is described next, as a way to sample from a distribution such as the distribution for  $\alpha$ .

### 8.2.3.3.2 Metropolis-Hastings Step

It is sometimes the case that one or more of the full conditional distributions  $g(\theta_i | \theta_j, \mathbf{x}, j \neq i) \quad i = 1, 2, \dots, p$ , required in Step (2) of the Gibbs sampler may not be available in closed form. This may happen as follows. These full conditional distributions are usually obtained using Bayes' Theorem and, while the two terms (the

likelihood and the prior distribution) in the numerator of Bayes' Theorem are usually known, the integration required to obtain the normalizing factor (the denominator) in Bayes' Theorem cannot be performed in closed form. That is, the required full conditional distribution is known only up to a multiplicative constant, the normalizing factor. The corresponding full conditional distribution is thus unavailable in closed form, and sample values from this distribution cannot be directly obtained as required in Step (2) of the Gibbs sampler.

Denote the full conditional distribution, which is known only up to a normalizing constant, as  $g(\theta_i | \theta_j, \mathbf{x}, j \neq i)$ . For convenience, we suppress the conditioning terms in the notation below. In this situation, sample observations may be obtained in Step (2) of the Gibbs sampler by using a so-called **Metropolis-Hastings step** (Hastings 1970) as follows:

- (4) Initialize  $\theta_i^0$  and set  $j = 0$ .
- (5) Generate an observation  $\theta_i^*$  from a candidate distribution  $q(\theta_i^* | \theta_i^j)$ , where  $q(y|x)$  is a probability density in  $y$  with parameter (for example mean)  $x$ .
- (6) Generate a uniform(0, 1) observation  $u$ .
- (7) Let
 
$$\theta_i^{j+1} = \begin{cases} \theta_i^*, & \text{if } u \leq \alpha(\theta_i^j, \theta_i^*) \\ \theta_i^j, & \text{otherwise} \end{cases}$$

where

$$\alpha(x, y) = \frac{g(y)q(x|y)}{g(x)q(y|x)}.$$

- (8) Increment  $j$  and go to (2)

Because  $\alpha$  uses a *ratio*,  $g(y)/g(x)$ , it can be calculated even though the normalizing constant for  $g$  is unknown. The candidate distribution in Step (2) can be almost any distribution (Gilks et al. 1996), although a symmetric distribution such as a normal distribution results in a simplification of the algorithm, and is called simply a **Metropolis step** (as opposed to a Metropolis-Hastings step). A common choice for  $q(y|x)$  is a normal distribution with mean  $x$  and a variance that allows the random deviates to be a representative sample from the entire complete conditional distribu-

tion. A preliminary rule of thumb suggested Gelman et al. (1995, Sec. 11.5) is that the variance be such that the new value,  $\theta_i^*$ , is picked in Step (4) about 30% of the time, and the old value,  $\theta_i^j$ , is picked about 70% of the time. The new value should be picked more often in problems with few parameters and less often in problems with many parameters.

Actually, BUGS favors a method called **adaptive rejection sampling** (Gilks and Wild 1992) instead of the Metropolis-Hastings algorithm. This method uses more storage space but fewer iterations. It requires that the conditional distributions in the Gibbs sampler be log-concave (George et al. 1993). This requirement is satisfied for the commonly used prior distributions. If the user happens to select a prior that leads to a problem, BUGS will give a diagnostic message.

### 8.2.3.3.3 BUGS (Bayesian Inference Using Gibbs Sampling)

Fortunately, for a wide range of common problems, there is little need to actually program the Gibbs sampler in practice. Gibbs sampling has been conveniently implemented through the **BUGS** software project (BUGS 1995, Spiegelhalter et al. 1995, and Gilks et al. 1994). It is currently available for free download at WWW URL

<http://www.mrc-bsu.cam.ac.uk/bugs/> .

The classic BUGS program uses text-based model description and a command-line interface, and versions are available for major computing platforms.

A Windows version, WinBUGS, has an optional graphical user interface (called DoodleBUGS) as well as on-line monitoring and convergence diagnostics. BUGS is reasonably easy to use and, along with a user manual, includes two volumes of examples. Section 8.2.3.4 illustrates how WinBUGS was used in Example 8.1, to obtain the hierarchical Bayes estimates of the 66 plant-specific scram rates and of the corresponding population-variability distribution. Section 8.2.3.5 uses WinBUGS in Example 8.2, to obtain the hierarchical Bayes estimates of  $p$  at each of the 68 plants and of the corresponding population-variability distribution.

### 8.2.3.4 Application to Example 8.1, Initiating Events

Let us now illustrate the hierarchical Bayes method in Example 8.1. We use the Gibbs sampler in BUGS to calculate all the required population and plant-specific scram rate estimates. To begin, we assume that the population-variability distribution is gamma( $\alpha, \beta$ ) just as in Section 8.2.2. The hyperparameter  $\alpha$  is given an exponential hyperprior distribution with mean and variance of 1, while the hyperparameter  $\beta$  is given an independent gamma(0.0625, 0.0625) hyperprior distribution. Thus,  $\beta$  is assumed to have a hyperprior mean and standard deviation of 1 and 4, respectively. The hyperpriors are diffuse (large variances), and have plausible means, so they will probably not bias the final answers much.

Figure 8.8 contains the WinBUGS model used here for this Poisson example. The initial values considered are:  $\alpha = 1$ ,  $\beta = 1$ , and  $\lambda[i] = 1$ ,  $i = 1, \dots, 66$ .

```

model
{
  for (i in 1:M) {
    lambda[i] ~ dgamma(alpha,beta)
    mu[i] <- lambda[i]*t[i]
    x[i] ~ dpois(mu[i])
  }
  alpha ~ dexp(1.0)
  beta ~ dgamma(0.0625, 0.0625)
}

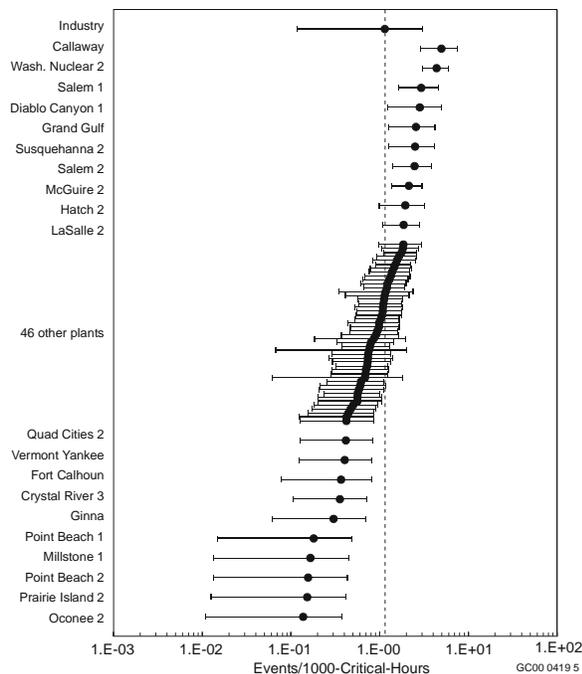
```

**Figure 8.8** WinBUGS model specification for Example 8.1.

After 1,000 burn-in iterations (to remove the effect of the initial starting values and to achieve convergence of the Markov chain), 10,000 additional simulated posterior sample values of  $\alpha$ ,  $\beta$ , and  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{66})$  were recorded. These 10,000 sample values were then used to calculate the required posterior point and credible interval estimates of  $\alpha$ ,  $\beta$ , and each  $\lambda_i$ . For example, the hierarchical Bayes estimated posterior mean of the Callaway scram rate is calculated to be 4.97 per 1,000 critical hours. The corresponding 90% credible interval on  $\lambda_1$  is [2.87, 7.51].

In addition, the marginal posterior mean and standard deviation of  $\alpha$  are calculated to be 1.38 and 0.30, respectively, whereas those for  $\beta$  are computed to be 1.21 and 0.32. A hierarchical Bayes 90% credible interval for  $\alpha$  is [0.94, 1.93], while the corresponding interval for  $\beta$  is [0.76, 1.78]. The marginal posterior correlation between  $\alpha$  and  $\beta$  is also easily calculated from the 10,000 pairs of corresponding posterior  $(\alpha, \beta)$  values to be 0.89. From Table 8.3 we see that the empirical Bayes point estimates of  $\alpha$  and  $\beta$  are 1.39 and 1.21, respectively, which are in near perfect agreement with the hierarchical Bayes estimates.

Figure 8.9 illustrates the hierarchical Bayes results. For each plant, the posterior mean and 90% credible interval are shown. The mean and 90% credible interval for the population-variability distribution are also shown, and a vertical dashed line is drawn through the mean. Actually, this population-variability distribution is the gamma distribution evaluated when  $\alpha$  and  $\beta$  are set equal to their posterior means. It does not reflect the uncertainty in these two hyperparameters. Figure 8.9 agrees very closely with Figure 8.3.



**Figure 8.9** Plant-specific posterior means and 90% credible intervals for  $\lambda$ , from hierarchical Bayes analysis.

Table 8.7 contains the same portion of the numerical hierarchical Bayes analysis results for Example 8.1 as are displayed in Table 8.3.

The point and interval estimates in Table 8.7 are all in good agreement with the empirical Bayes estimates in Table 8.3.

**Table 8.7** Portion of hierarchical Bayes analysis results for Example 8.1.

Plant	Posterior mean and 90% credible interval <sup>a</sup>
Industry	(0.116, 1.14, 3.06)
Callaway	(2.87, 4.97, 7.51)
Wash. Nuc. 2	(3.01, 4.39, 5.99)
Salem 1	(1.65, 2.94, 4.57)
Diablo Can. 1	(1.24, 2.83, 4.99)
Grand Gulf	(1.27, 2.57, 4.21)
Pt. Beach 2	(0.013, 0.156, 0.429)
Prairie Isl. 2	(0.012, 0.152, 0.410)
Oconee 2	(0.011, 0.137, 0.374)

a. Format is (lower bound, mean, upper bound). The least significant digit may be inaccurate by 2 or more, because of Monte Carlo sampling error.

### 8.2.3.5 Application to Example 8.2, AFW Segment Failures to Start

Recall that this example has 68 plants with sparse failure data consisting of only 6 failures in 1993 demands. Because the data are so sparse, the form of the prior, the population-variability distribution, can strongly influence the answers. Therefore, the example is analyzed using two population-variability distributions, first a beta distribution, as in Section 8.2.2, and then a logistic-normal distribution. In each case, diffuse hyperpriors with plausible means are used. Therefore, the exact choices made for the hyperpriors have little influence on the answer.

#### 8.2.3.5.1 Analysis with Beta Prior

We assume that the population-variability distribution is a beta( $\alpha, \beta$ ) distribution. To begin, the hyperparameter  $\alpha$  is assigned an exponential(1) hyperprior

distribution with a hyperprior mean and standard deviation of 1, while the hyperparameter  $\beta$  is assigned an independent gamma(1.0, 0.035) hyperprior distribution. Thus, we assume that  $\beta$  has a hyperprior mean and standard deviation both equal to 28.6. The forms of these hyperprior distributions ensure that the joint posterior distribution will be log-concave, and the diffuseness of the hyperpriors ensures that they will not influence the final answers greatly.

Fig. 8.10 contains the WinBUGS model used here for this binomial example. The initial values were:  $\alpha = 1$ ,  $\beta = 1$ , and  $p[i] = 0.01$ ,  $i = 1, \dots, 68$ .

```

model
{
  for (i in 1:M) {
    x[i] ~ dbin(p[i],n[i])
    p[i] ~ dbeta(alpha, beta)
  }
  alpha ~ dexp(1.0)
  beta ~ dgamma(1.0,0.035)
}
    
```

**Figure 8.10** WinBUGS model specification for a beta prior in Example 8.2.

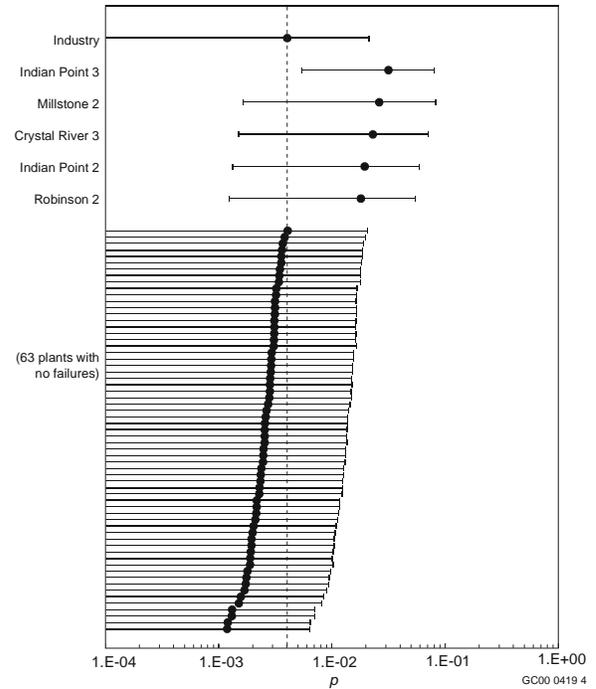
After 10,000 burn-in iterations (to remove the effect of the initial starting values and to achieve convergence of the Markov chain), 90,000 additional simulated posterior sample values of  $\alpha$ ,  $\beta$ , and  $\mathbf{p} = (p_1, p_2, \dots, p_{68})$  were recorded. These 90,000 sample values were then used to calculate the required posterior point and credible interval estimates of  $\alpha$ ,  $\beta$ , and each  $p_i$ .

In addition, the marginal posterior mean and standard deviation of  $\alpha$  are calculated to be 0.188 and 0.136, respectively, whereas those for  $\beta$  are computed to be 46.4 and 32.4. A hierarchical Bayes 90% credible interval for  $\alpha$  is [0.046, 0.442], while the corresponding interval for  $\beta$  is [9.95, 109.5]. Note the large uncertainties associated with the hierarchical Bayes estimates of  $\alpha$  and  $\beta$  because of the sparseness of the data. Table 8.6 shows that the empirical Bayes point estimates of  $\alpha$  and  $\beta$  are 0.137 and 36.34, respectively, which are well within the 90% credible intervals of the corresponding hierarchical Bayes estimates.

Figure 8.11 illustrates the hierarchical Bayes results. For each plant, the posterior mean and 90% credible interval are shown. The mean and 90% credible

interval for the population-variability distribution are also shown, and a vertical dashed line is drawn through the mean.

Table 8.8 contains the same portion of the numerical hierarchical Bayes analysis results for Example 8.2 as given in Table 8.6. The results are presented to only two significant digits, because the Monte Carlo errors reported by BUGS show that the third significant digit is not meaningful.



**Figure 8.11** Plant-specific posterior means and 90% credible intervals for  $p$ , from hierarchical Bayes analysis with beta population-variability distribution.

The point and interval estimates in Table 8.8 are all in reasonably close agreement with the empirical Bayes estimates in Table 8.6.

**Table 8.8** Portion of hierarchical Bayes analysis results using beta prior for Example 8.2.

Plant	Posterior mean and 90% credible interval <sup>a</sup>
-------	---

Industry	(1.6E-9, 4.0E-3, 2.1E-2)
Indian Point 3	(5.5E-2, 3.2E-2, 7.9E-2)
Millstone 2	(1.6E-2, 2.6E-2, 8.3E-2)
Crystal R. 3	(1.5E-2, 2.3E-2, 7.1E-2)
Indian Point 2	(1.3E-2, 1.9E-2, 5.9E-2)
Robinson 2	(1.2E-2, 1.8E-2, 5.4E-2)
Prairie Isl. 1	(1.5E-15, 4.1E-3, 2.1E-2)
Vogtle 1	(2.5E-16, 1.2E-3, 6.4E-3)
a. Format is (lower bound, mean, upper bound).	

### 8.2.3.5.2 Analysis with Logistic-Normal Prior

One of the primary advantages in using the hierarchical Bayes method is the ability to consider non-conjugate population-variability (or prior) distributions. We now illustrate this for Example 8.2.

The previous analysis considered a conjugate beta prior in this example. Table 8.8 shows that, for  $x = 0$ , the use of a beta prior produces lower 5% credibility limits on the order of  $10^{-15}$  or  $10^{-16}$ , which are unrealistically small. This result is a consequence of the HB-fitted L-shaped beta prior distribution with high density close to  $p = 0$ .

We can avoid such unrealistic results by using a non-conjugate logistic-normal prior distribution (see Section 6.3.2.4.2) in the hierarchical Bayes approach. Recall that, while such a prior is extremely difficult to consider in an empirical Bayes approach, it is extremely easy to do in hierarchical Bayes.

Figure 8.12 contains the WinBUGS model specification for using a logistic-normal prior in Example 8.2. Observe that this is no more difficult than using the hierarchical Bayes model based on a conjugate beta prior shown in Figure 8.10. Note that BUGS parameterizes the normal distribution in terms of  $\tau = 1/\sigma^2$ . A commonly used prior distribution for  $\tau$  is gamma, and that choice is used here. Thus, assigning  $\mu$  a prior precision of 0.0001 is equivalent to assigning it a prior variance of 10000, or a prior standard deviation of 100.

```

model
{
  for (i in 1:M) {
    x[i] ~ dbin(p[i],n[i])
    y[i] ~ dnorm(mu,tau)
    p[i] <- exp(y[i])/(1 + exp(y[i]))
  }
  mu ~ dnorm(-5,0.0001)
  tau ~ dgamma(1,7)
}

```

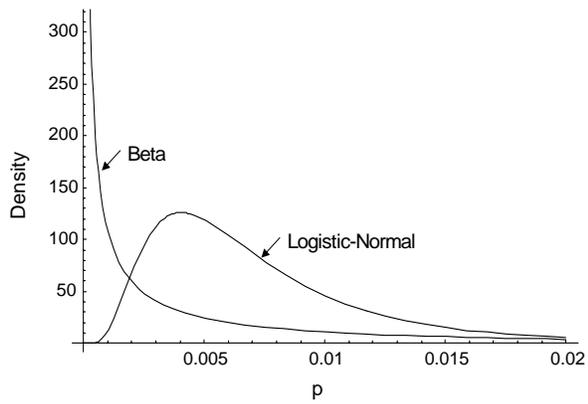
**Figure 8.12** WinBUGS model specification for a logistic-normal prior in Example 8.2.

Again using 10,000 burn-in iterations and 90,000 replications of the Gibbs sampler for the model in Figure 8.12, WinBUGS likewise calculated posterior means and 90% credibility intervals for  $\mu$ ,  $\sigma$ , and each  $p_i$ .

The marginal posterior mean and standard deviation of  $\mu$  are calculated to be  $-5.097$  and  $0.09517$ , respectively, whereas those for  $\sigma$  are computed to be  $0.640$  and  $0.238$ . A hierarchical Bayes 90% credible interval for  $\mu$  is  $[-5.253, -4.939]$ , while the corresponding interval for  $\sigma$  is  $[0.322, 1.08]$ .

Figure 8.13 shows the two estimated population-variability distributions, when the form is assumed to be beta (the conjugate distribution) or logistic-normal. The mean of the beta prior is  $0.004$  and the mean of the logistic-normal prior is  $0.007$ , nearly twice as large. Note that, unlike the beta prior, the logistic-normal prior avoids the high probability density close to  $p = 0$ .

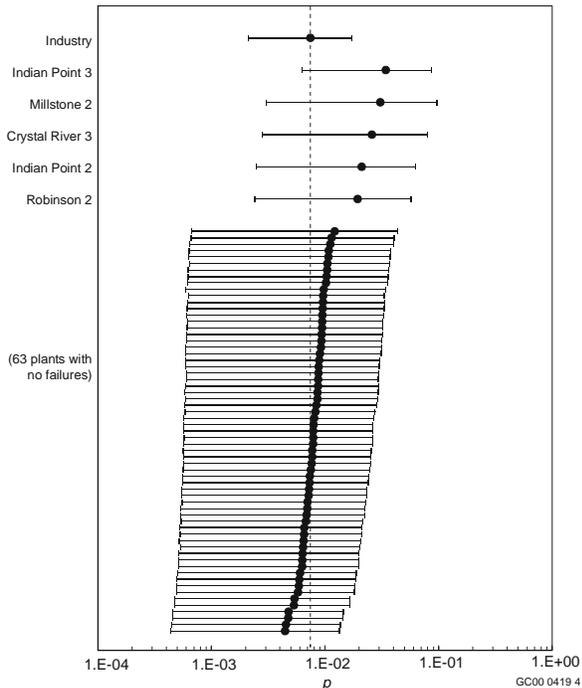
8. Parameter Estimation Using Existing Data Sources DRAFT



**Figure 8.13** Fitted population-variability distributions in Example 8.2.

Figure 8.14 illustrates the hierarchical Bayes results using the logistic-normal prior. As in Figure 8.11, the posterior mean and 90% credible interval are shown for each plant. The mean and 90% credible interval for the population-variability distribution are also shown, and a vertical dashed line is drawn through the mean. This plot may be directly compared with Figure 8.11.

Table 8.9 contains the same portion of the numerical hierarchical Bayes analysis results for the logistic-normal as given in Table 8.8 for the beta prior.



**Figure 8.14** Plant-specific posterior means and 90% credible intervals for  $p$ , from hierarchical Bayes analysis with logistic-normal population-variability distribution.

Note that the posterior means and 90% credible intervals in Table 8.9 are all larger than those in Table 8.8. As stated above, this is a direct consequence of using the logistic-normal prior in Figure 8.13. Observe also that the lower 90% credibility limits for plants with no failures, such as Prairie Island 1 and Vogtle 1, are now much more realistic than the corresponding limits in Table 8.8.

## 8.2.4 Discussion

### 8.2.4.1 Hierarchical Bayes Is Still Bayes

The hierarchical Bayes model is a special case of the familiar Bayesian model. It is not some new kind of construction. To see this, consider Levels 0 and 1 together in Figure 8.6. The prior parameter is a vector  $\theta$ , consisting of the hyperparameters and the  $\lambda_i$ s. Thus, the prior  $\theta$  is a vector with dimension  $2+m$ . The prior distribution on  $\theta$  is specified: the joint distribution of the hyperparameters  $\alpha$  and  $\beta$  is given by the hyperprior in Level 0, and the conditional distributions of the  $\lambda_i$ s

**Table 8.9** Portion of hierarchical Bayes analysis results using logistic-normal prior for Example 8.2.

Plant	Posterior mean and 90% credible interval <sup>a</sup>
Industry	(2.1E-3, 7.4E-3, 1.7E-2)
Indian Point 3	(6.2E-3, 3.4E-2, 8.7E-2)
Millstone 2	(3.0E-3, 3.1E-2, 9.7E-2)
Crystal R. 3	(2.8E-3, 2.6E-2, 8.0E-2)
Indian Point 2	(2.5E-3, 2.1E-2, 6.2E-2)
Robinson 2	(2.4E-3, 1.9E-2, 5.7E-2)
Prairie Isl. 1	(6.6E-4, 1.2E-2, 4.4E-2)
Vogtle 1	(4.3E-4, 4.E-3, 1.3E-2)

a. Format is (lower bound, mean, upper bound).

are independent, and specified by  $g$ , conditional on  $\alpha$  and  $\beta$ . Thus, Levels 0 and 1 together specify the prior parameter vector and its prior distribution. The posterior distribution is therefore given by Bayes' Theorem:

$$g_{\text{post}}(\theta | \text{data}) \propto \Pr(\text{data} | \theta) \times g_{\text{prior}}(\theta).$$

This differs from the applications of Bayes' Theorem elsewhere in this handbook in only two ways: the parameter is a high-dimensional vector, and the prior distribution has a lot of structure, as shown in Figure 8.6.

A practical consequence of the high dimension of  $\theta$  is that the tools of Chapter 6, numerical integration and simple random sampling methods, do not work well. More recently developed methods, versions of Markov chain Monte Carlo sampling, must be used. Conceptually, however, hierarchical Bayes modeling fits perfectly within the framework of Bayes' Theorem. In particular, everything is legal, with no double counting of data.

### 8.2.4.2 The "Two-Stage" Bayesian Method

Kaplan (1983) introduced a "two-stage" Bayesian method, which has sometimes been used in PRA work. It is described here in terms of Figure 8.6. The method singles out the plant of special interest, say Plant 1. It

then estimates the hyperparameters,  $\alpha$  and  $\beta$ , in a Bayesian way, using the data from *all the plants except Plant 1*. It then uses the estimated  $g(\lambda | \alpha, \beta)$  as a prior, combining it with the data from Plant 1 to estimate  $\lambda_1$  in the usual Bayesian way.

The originally intended reason for not using Plant 1 in the first stage was to avoid double counting. As mentioned in Section 8.2.4.1, the hierarchical Bayes method is based directly on Bayes' Theorem, and therefore does not involve double counting. Therefore, the two-stage Bayesian method should no longer be used, but should be replaced by the conceptually cleaner hierarchical Bayes method. Now that numerical algorithms have been developed to sample from the posterior distributions, this is completely feasible.

#### 8.2.4.3 Lower Bounds on Parameters

Example 8.2 illustrated that different population-variability distributions (prior distributions) can lead to radically different lower percentiles of the parameters' posterior distributions. This occurred in that example for those plants that had experienced no failures. A beta prior led to 5th percentiles for  $p$  on the order of  $1E-15$ , whereas a logistic-normal prior led to 5th percentiles on the order of  $5E-4$ . No one believes the first answers, but many people could believe the second answers. Does that mean that the answers from the logistic-normal prior are "right"?

In fact, the lower bounds are an artifact of the model, in both cases. The plants that experienced no failures reveal only that  $p$  is "small" at those plants. They do not give information about how small  $p$  might be. If many plants have no failures, as in Example 8.2, then we have very little information about the lower end of the population-variability distribution. In contrast to this, the plants that experienced one or more failures convey much more information, revealing both how large and how small  $p$  could plausibly be at those plants. Therefore, the 95th percentile of  $p$  at any plant is *somewhat* dependent on the assumed form of the population-variability distribution (beta, logistic-normal, or whatever). But when many plants have no observed failures, the 5th percentile of  $p$  at any of those

plants is *extremely* dependent on this assumed form.

And why was a particular form assumed for the population-variability distribution? For convenience only! Thus, even if the answers from a logistic-normal prior look credible, we do not "know" that they are correct. We may choose to discard the results from using a beta prior, because we do not want to publish 5th percentiles that could be ridiculed. We might also choose to publish the results from using a logistic-normal prior, knowing that the 5th percentiles appear credible. But it is a delusion to think that we "know" lower bounds on  $p$  at the plants with no observed failures. The calculated lower bounds remain an artifact of the assumed model.

Fortunately, lower bounds are not a concern for risk. Means and upper bounds are the important quantities, and they can be estimated with much less dependence on the model.

#### 8.2.4.4 Empirical Bayes as an Approximation to Hierarchical Bayes

As remarked elsewhere, Figures 8.3 and 8.9 are very similar to each other, and Figures 8.5 and 8.11 are similar to each other. That is, in both Examples 8.1 and 8.2, the empirical Bayes results are numerically close to the hierarchical Bayes results, when (a) the empirical Bayes method includes the Kass-Steffey adjustment, and (b) both methods use the conjugate population-variability distribution, a gamma distribution for Poisson data and a beta distribution for binomial data. This agreement between the methods is more than coincidence. Kass and Steffey (1989) developed their method specifically with this intent: to make the empirical Bayes approach give a first-order approximation to the hierarchical Bayes approach with a diffuse hyperprior. The method does this well in the two examples. Of course, when the hierarchical Bayes method does not use a conjugate population-variability distribution, as in Section 8.2.3.5.2, there is no corresponding empirical Bayes method.