Bayesian Analysis of Piping Leak Frequency Using OECD/NEA Data

Min Wang

Mahesh D. Pandey

Department of Civil Engineering,
University of Waterloo,
Waterloo, ON, N2L 3G1, Canada

Jovica R. Riznic

Canadian Nuclear Safety Commission,
Ottawa, ON, K1P 5S9, Canada

The estimation of piping failure frequency is an important task to support the probabilistic risk analysis and risk-informed in-service inspection of nuclear power plant systems. This paper describes a hierarchical or two-stage Poisson-gamma Bayesian procedure and applies this to estimate the failure frequency using the Organization for Economic Co-operation and Development/Nuclear Energy Agency pipe leakage data for the United States nuclear plants. In the first stage, a generic distribution of failure rate is developed based on the failure observations from a group of similar plants. This distribution represents the interplant (plant-to-plant) variability arising from differences in construction, operation, and maintenance conditions. In the second stage, the generic prior obtained from the first stage is updated by using the data specific to a particular plant, and thus a posterior distribution of plan specific failure rate is derived. The two-stage Bayesian procedure is able to incorporate different levels of variability in a more consistent manner. [DOI: 10.1115/1.4000343]

1 Introduction

Since the failure of piping system can have adverse effect on the safety and reliability of a nuclear power plant, the piping failure frequency is an important input parameter to probabilistic safety assessment (PSA) and risk-informed in service inspection (RI-ISI) of systems important to safety. In the design stage, reliability and safety analyses conducted are generally based on generic failure rates given in the codes or standards. As far as an operating plant is concerned, it is of interest to evaluate the plant-specific failure rates and to investigate their departure from generic reference values.

The estimation of the piping failure rate based on probabilistic methods using in-service data or relying on expert opinions has been described in various studies, such as Refs. [1–3]. More recently, U.S. Nuclear Regulatory Commission (USNRC) has published a comprehensive handbook of parameter estimation for probabilistic risk assessment [4] in which both frequentist and Bayesian inference methods are discussed. Because the nuclear power plants are highly reliable systems, piping failures tend to be relatively rare events. In other words, empirical data for failure rate estimation are quite sparse. In this situation, the Bayesian method is preferred over the classical statistical methods.

This paper describes a two-stage Bayesian procedure to estimate the plant-specific piping failure rate. The first step is to establish a common population variability curve (PVC) using the industry wide data. The second step is to customize this distribution to a specific plant of interest using the data collected from this specific plant. The proposed approach is applied to estimate the leak rate using the Organization for Economic Co-operation and Development/Nuclear Energy Agency (OECD/NEA) pipe leakage data for the United States nuclear plants.

2 Literature Review

The use of Bayesian method in PSA started since mid-1970s [5,6]. A key obstacle in statistical analysis is the lack of data regarding the failure of highly reliable nuclear power plant (NPP) equipment. In such cases, the Bayesian method is quite useful as it provides a mechanism for incorporating other sources of information as prior belief. Furthermore the Bayesian framework also allows a propagation of basic event uncertainties through a logical model. Siu and Kelly [7] presented a detailed tutorial on Bayesian parameter estimation especially in the context of PSA.

In the plant-specific risk studies, the idea of using a generic distribution reflecting the plant-to-plant variability as prior distribution can be found in the reliability literature. Apostolakis et al. [8,9] suggested that a generic distribution be constructed by using expert judgment or standards. Vaurio [10] proposed a gamma distribution conjugate to the Poisson-distributed failure rate as a prior distribution. The data from other plants were used to determine this chosen prior by moment matching method.

The formulation of hierarchical (multistage) model is central to modern Bayesian statistics due to flexibility to construct hierarchical priors and combine different levels of information. A general purpose of this formulation is the assimilation of date from different sources. The two-stage Bayesian approach can be considered as a special case of the hierarchical Bayes, and it was initially developed by Kaplan and coworker [11,12] and Fröhner [13,14].

Kaplan [12] suggested lognormal distribution as the prior PVC of failure rate in the first stage and a uniform distribution as hyperprior of the hyperparameters of the lognormal distribution. Pörn [15] applied an idea of the two-stage Bayesian method to estimate the component failure rates presented in the T-Book, a handbook concerning reliability data of components in Nordic nuclear power plants. This study adopted a contamination class distribution as the prior of the failure rate, and then the uncertainty associated with the contamination parameter is modeled by a noninformative hyperprior. Reader can refer to Ref. [16] for more detailed discussion about the contamination class of the priors. Pörn's model was reviewed by Cooke et al. [17] and further discussed by Meyer and Hennings [18].

Hora and Iman [19] proposed a super population model in which the failure rate for each plant is unique but related to the failure rate of all other plants through the super population. Hofer and co-workers [20–22] however pointed out the wrong order of integration of improper integrals in the study of Hora and Iman [19]. Readers are referred to Ref. [23] for a more detailed discussion of this topic. Later Bunea et al. [24] applied this method to pipe failure data obtained from German nuclear plants (ZEDB database).

Contributed by the Nuclear Division of ASME for publication in the Journal of Engineering for Gas Turbines and Power. Manuscript received July 21, 2009; final manuscript received July 28, 2009; published online April 26, 2010. Editor: Dilip R. Ballal.

3 Bayesian Analysis

3.1 Problem Statement. The occurrence of the pipe failure event is modeled as a homogeneous Poisson process with failure rate λ. The number of failures, x, in an interval $(0-t)$ are given by the Poisson distribution

$$f(x|\lambda) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \tag{1}$$

The basic idea is that the failure rate for a specific plant, λ_i, is a fixed but unknown value. The failure rate in another similar plant is λ_k, which is different from λ_i and an unknown constant.

Therefore, to model uncertainty associated with the failure rate λ, it is treated as a random variable with gamma distribution

$$f(\lambda|\alpha,\beta) = \frac{\lambda^{\alpha-1}\exp(-\beta\lambda)\beta^{\alpha}}{\Gamma(\alpha)} \tag{2}$$

where $\Gamma(\cdot)$ denotes the gamma function. The average and coefficient of variation (COV) of λ are given as α/β and $1/\sqrt{\alpha}$, respectively. The gamma distribution is a conjugate prior.

Thus, in a fleet of if n plants, the underlying failure rates, $\lambda_1,\lambda_2,\ldots,\lambda_n$, constitute a random sample from the prior distribution, Eq. (2). The prior is also referred to as the population variability curve, and it models the plant to plant variability in the failure rate.

Ideally if in a large fleet of nuclear plants large amount of pipe failure data were available, a histogram of λ could have been generated and fitted with a smooth probability distribution to derive the prior PVC. However, in reality such large data do not exist, which complicates the modeling of PVC. This second level of uncertainty is modeled by treating the parameters α and β as random variables. This is the essence of the two-stage procedure.

Suppose pipe failure data are available from $(n+1)$ plants. The failure data from a plant are summarized as the doublet (x_i,t_i), $i=1,2,\ldots,n+1$. Note that x_j is the number of failures during the observation t_j in an ith plant. The objective is to estimate the failure rate in $(n+1)$th plant using the data from n plants and also data available from the current plant of interest.

3.2 Hierarchical Model. The hierarchical Bayes approach is based on the use of hierarchical priors. Consider a random variable Y with distribution $f(y|\theta_1)$, where θ_1 is a distribution parameter. A prior distribution of the uncertain parameter θ_1 is specified as $f(\theta_1|\theta_2)$, which is indexed by another parameter θ_2, referred to as the hyperparameter. Instead of assigning a fixed value to θ_2, another prior distribution can be assigned to θ_2, which, in turn, may be indexed by a hyperparameter θ_3 and so on. Thus, in this approach the prior distribution is specified in a hierarchical fashion in a multiple stages, or hierarchies. The parameter θ_i $(i=2,3,\ldots,k)$ is referred to as ith hyperparameter in the prior distribution and θ_i could be a vector.

The mathematical structure of a general k-level hierarchical model is described as follows. Given a vector of the observed data $\mathbf{y}$, the conditional updated distribution or posterior of θ_1 is written as

$$f(\theta_1|\mathbf{y}) = c^{-1}\int_{\theta_2}\cdots\int_{\theta_k} f(\mathbf{y}|\theta_1)f(\theta_1|\theta_2)\cdots f(\theta_{k-1}|\theta_k)f(\theta_k)d\theta_k\cdots d\theta_2 \tag{3}$$

where c is normalizing constant and must be chosen to satisfy $\int_{\theta 1} f(\theta_1|\mathbf{y})d\theta_1=1$. In case of a two-stage model, Eq. (3) reduces to

$$f(\theta_1|\mathbf{y}) = c^{-1}\int_{\theta_2} f(\mathbf{y}|\theta_1)f(\theta_1|\theta_2)f(\theta_2)d\theta_2 \tag{4}$$

Although there is no theoretical reason for limiting hierarchical priors to just two stages, more than two stages rarely used in practice (e.g., Refs. [25,26]).

3.3 Two-stage Bayesian Procedure

3.3.1 Assumptions.

(1) The hyperparameter θ has density $f(\theta)$.
(2) Given θ, the parameters $\{\lambda_i\}$ are identically and independently distributed (iid) with density $f(\lambda|\theta)$.
(3) Given $\{\lambda_i\}$, the data $\{x_i\}$ are independent of density $f(x_i|\lambda_i)$ and independent of θ and of all λs other than λ_i.

3.3.2 Calculation Procedure. Using the law of total probability, the mixed distribution of X has density, given θ as

$$f(x|\theta) = \int_{\lambda} f(x|\lambda)g(\lambda|\theta)d\lambda \tag{5}$$

and x_i is iid with this density. Likewise, we write $f(\lambda_{n+1}|\mathbf{x_{n+1}})$ to include the hyperparameter θ

$$f(\lambda_{n+1}|\mathbf{x_{n+1}}) = \int_{\theta} f(\lambda_{n+1}|\mathbf{x_{n+1}},\theta)f(\theta|\mathbf{x_{n+1}})d\theta \tag{6}$$

Separating the evidence $\mathbf{x_{n+1}}$ of the right side into $\mathbf{x_n}$ and x_{n+1}, Eq. (6) is rewritten as

$$f(\lambda_{n+1}|\mathbf{x_{n+1}}) = \int_{\theta} f(\lambda_{n+1}|x_{n+1},\mathbf{x_n},\theta)f(\theta|x_{n+1},\mathbf{x_n})d\theta \tag{7}$$

Applying the Bayes' theorem on both terms in the integral and using the assumptions 2 and 3, the first term of integrand in Eq. (7) turns out to be

$$f(\lambda_{n+1}|x_{n+1},\mathbf{x_n},\theta) = \frac{f(x_{n+1}|\lambda_{n+1})f(\lambda_{n+1}|\theta)}{f(x_{n+1}|\mathbf{x_n},\theta)} \tag{8}$$

and the second term of integral in Eq. (7) is

$$f(\theta|x_{n+1},\mathbf{x_n}) = \frac{f(x_{n+1}|\mathbf{x_n},\theta)f(\theta|\mathbf{x_n})}{f(x_{n+1}|\mathbf{x_n})} \tag{9}$$

Substituting Eqs. (8) and (9) into Eq. (7), we get

$$f(\lambda_{n+1}|\mathbf{x_{n+1}}) = \int_{\theta} \frac{f(x_{n+1}|\lambda_{n+1})f(\lambda_{n+1}|\theta)f(\theta|\mathbf{x_n})}{f(x_{n+1}|\mathbf{x_n})}d\theta \tag{10}$$

Ignoring the normalization constant $f(x_{n+1}|\mathbf{x_n})$, Eq. (10) can be simplified to

$$f(\lambda_{n+1}|\mathbf{x_{n+1}}) \propto f(x_{n+1}|\lambda_{n+1})\int_{\theta} f(\lambda_{n+1}|\theta)f(\theta|\mathbf{x_n})d\theta \tag{11}$$

where $f(\theta|\mathbf{x_n})$ is obtained, by using Bayes' theorem and ignoring the normalizing factor, as (see Appendix A for derivation)

$$f(\theta|\mathbf{x_n}) \propto \left[\prod_{i=1}^{n}\int_{\lambda_i} f(x_i|\lambda_i)f(\lambda_i|\theta)d\lambda_i\right]f(\theta) \tag{12}$$

Now we can see that Eq. (11) is important for two reasons. First, it separates the role of x_{n+1} and $\mathbf{x_n}$. Second, it shows how the inference procedure is worked in two stages. The integral part of the Eq. (11) is obtained as population distribution (PVC) in the hierarchical Bayesian model. However, it is actually the predictive distribution of λ_{n+1}, denoted as $f(\lambda_{n+1}|\mathbf{x_n})$, based on Eq. (12) because it has the form

$$f(\lambda_{n+1}|\mathbf{x_n}) = E_{\theta|\mathbf{x_n}}[f(\lambda_{n+1}|\theta)] \tag{13}$$

Equation (11) clearly defines the two stages in the Bayesian inference for λ_{n+1}. In the first stage, the supplementary information $\mathbf{x_n}$ is used to generate a posterior distribution of hyperparameter θ, called the hyperposterior distribution, from Eq. (12). Then a population distribution or predictive distribution $f(\lambda_{n+1}|\mathbf{x_n})$ of λ_{n+1} is obtained from Eq. (13) based on the hyperposterior distribution.

The procedures for inference λ_{n+1} using the two-stage Bayesian analysis can be summarized as follows.

(1) Establish the hyperprior $f(\theta)$ based on the information available. Noninformative prior is commonly used.
(2) Evaluate the likelihood function of $\mathbf{x_n}$, $f(\mathbf{x_n}|\theta)$, which is the part enclosed in the square brackets of the Eq. (12).
(3) Evaluate the hyperposterior distribution $f(\theta|\mathbf{x_n})$ from Eq. (12) using the evidence from other similar plants.
(4) Evaluate the population variability distribution $f(\lambda_{n+1}|\mathbf{x_n})$ from Eq. (13) for stage 2.
(5) Evaluate the likelihood function of the plant-specific observed data, $f(x_{n+1}|\lambda_{n+1})$.
(6) Evaluate the posterior distribution $f(\lambda_{n+1}|\mathbf{x_{n+1}})$ from Eq. (11) using the population distribution as the prior distribution obtained from step 3.

3.3.3 Discussion. Equation (3) illustrates the general formulation of the hierarchical Bayesian model. This section briefly discusses the relationship of the hierarchical Bayes with the classical Bayesian method and empirical Bayesian (EB) method.

3.3.3.1 Single stage Bayesian method. When only one-stage prior information is available, the hierarchical Bayesian model reduces to the simple and well known Bayes model for the failure rate λ_{n+1} of the specific plant $n+1$.

$$f(\lambda_{n+1}|x_{n+1}) \propto f(x_{n+1}|\lambda_{n+1})f(\lambda_{n+1}) \tag{14}$$

The prior distribution for λ_{n+1} can be selected as a gamma distribution with parameters α_S and β_S, $f(\lambda_{n+1}|\alpha_S,\beta_S)$. Because the mean and coefficient of variation of λ_{n+1} are given as α_S/β_S and $1/\sqrt{\alpha_S}$, respectively, α_S and β_S can be estimated as $[\mathrm{COV}(\lambda_{n+1})]^{-2}$ and $\alpha_S/E[\lambda_{n+1}]$, respectively. Because the gamma distribution and Poisson distribution are conjugate, the posterior distribution of λ_{n+1} turns out to be gamma distribution with parameters $(x_{n+1}+\alpha_S, t_{n+1}+\beta_S)$, in which x_{n+1} and t_{n+1} are number of failures and operating time observed in the specific plant. The posterior mean of λ_{n+1} can simply be calculated as $(x_{n+1}+\alpha_S)/(t_{n+1}+\beta_S)$. We refer to Eq. (14) as simple Bayes because it does not use the information for other plants.

3.3.3.2 Empirical Bayesian method. This method involves the use of classical techniques to fit the prior distribution on the basis of data. In EB method, the prior distribution of λ_{n+1} can be selected as a gamma distribution with parameters α_E and β_E, as in the case of the simple Bayesian method. Instead of assigning the prior mean and COV to λ_{n+1}, the parameters α_E and β_E are estimated by using the data from other plants in maximum likelihood method. As in the simple Bayesian method, the posterior distribution of λ_{n+1} is obtained as also gamma distribution with parameters $(x+\alpha_E, t+\beta_E)$. This method uses fixed point estimates α_E and β_E.

3.3.3.3 Hierarchical Bayesian method. Instead of using a deterministic prior distribution, in the simple one-stage Bayesian and empirical Bayesian methods, the prior $f(\lambda_{n+1}|\alpha,\beta)$ in the two-stage Bayesian method is uncertain. That is, instead of using the point estimate for α and β, the parameters α and β in the prior are treated as random variables. The two-stage hierarchical Bayesian method is illustrated in Eq. (11) where the prior of λ_{n+1} is obtained as average over the hyperposterior distribution of α and β, which is updated by using the information for other plants.

3.3.4 Implementation. An advantage though controversial is that the Bayes' theorem can incorporate the prior knowledge from other sources before making observation as prior distribution. The effect of the prior distribution on the estimation decreases with increasing observation. However, if this prior distribution can well represent the reality, we can make estimations with a higher degree of belief, especially for a small size of sample. When there is no prior information available, a noninformative prior is common.

The first step in the two-stage Bayesian analysis is to assign a hyperprior distribution to α and β in the population variability distribution. The choice of a form for the second stage or high stage prior seems to have relatively little effect. Because we have no readily available information regarding the hyperprior, noninformative prior based on the concept of data translated likelihood [27] is adopted for them.

Jeffreys' rule [28] requires that the prior distribution is invariant under one-to-one transformation. This invariant prior is proportional to the square root of the determinant of Fisher's (information) matrix of the population variability distribution. Application of Jeffreys' rule to the gamma distribution in Eq. (2) yields the joint hyperprior of α and β as

$$f(\alpha,\beta) \propto \frac{1}{\sqrt{\alpha}\beta} \tag{15}$$

There are two likelihood functions in the first and second stages. In the first stage, the likelihood of the observations $\mathbf{x_n}$, $f(\mathbf{x_n}|\alpha,\beta)$, from other plants is used to estimate the hyperposterior distribution. As mentioned in the step 2 of the two-stage Bayesian procedure, $f(\mathbf{x_n}|\alpha,\beta)$ is the part enclosed in the square brackets of the Eq. (12). Replacing the θ with (α,β), we have

$$f(\mathbf{x_n}|\alpha,\beta) = \prod_{i=1}^{n} \int_{\lambda_i} f(x_i|\lambda_i) f(\lambda_i|\alpha,\beta) d\lambda_i \tag{16}$$

Substituting the Eqs. (1) and (12) into (16), the likelihood of $\mathbf{x_n}$ turns out (Appendix B gives the derivation)

$$f(\mathbf{x_n}|\alpha,\beta) = \prod_{i=1}^{n} \frac{\Gamma(x_i+\alpha)}{x_i!\,\Gamma(\alpha)}\left(\frac{\beta}{\beta+t_i}\right)^{\alpha}\left(\frac{t_i}{\beta+t_i}\right)^{x_i} \tag{17}$$

The likelihood function of x_{n+1} in the second stage comes from the Poisson distribution

$$f(x_{n+1}|\lambda_{n+1}) = \frac{(\lambda_{n+1}t_{n+1})^{x_{n+1}}\exp(-\lambda_{n+1}t_{n+1})}{x_{n+1}!} \tag{18}$$

Incorporating the hyperprior distribution in Eq. (15) and the likelihood function in Eq. (19), the hyperposterior distribution of α and β is

$$f(\alpha,\beta|\mathbf{x_n}) \propto \prod_{i=1}^{n} \frac{\Gamma(x_i+\alpha)}{x_i!\,\Gamma(\alpha)}\left(\frac{\beta}{\beta+t_i}\right)^{\alpha}\left(\frac{t_i}{\beta+t_i}\right)^{x_i} f(\alpha,\beta) \tag{19}$$

The population variability distribution is mathematically a prediction distribution of λ_{n+1} over the hyperposterior distribution of α and β. From Eq. (13), the PVC can be obtained

$$f(\lambda_{n+1}|\mathbf{x_n}) = \int_{\alpha}\int_{\beta}\left[\frac{\lambda_{n+1}^{\alpha-1}\exp(-\beta\lambda_{n+1})\beta^{\alpha}}{\Gamma(\alpha)} f(\alpha,\beta) \times \prod_{i=1}^{n} \frac{\Gamma(x_i+\alpha)}{x_i!\,\Gamma(\alpha)}\left(\frac{\beta}{\beta+t_i}\right)^{\alpha}\left(\frac{t_i}{\beta+t_i}\right)^{x_i}\right] d\alpha d\beta \tag{20}$$

4 Simulation Study

A simulation study is present to illustrate the two-stage Bayesian procedure for estimating the plant-specific failure rate. Suppose we have 31 plants and the 31st plant is of interest.

Table 1 Simulated failure data

i	λ_i	x_i
1	0.3906	6
2	0.3850	3
3	0.0925	0
4	0.2947	3
5	0.0307	1
6	0.1026	0
7	0.2099	3
8	0.1132	0
9	0.3787	6
10	0.1006	1
11	0.0847	2
12	0.2999	5
13	0.3689	1
14	0.0302	0
15	0.0884	0
16	0.1754	2
17	0.3204	4
18	0.0849	1
19	0.6011	6
20	0.2177	3
21	0.5146	6
22	0.0215	0
23	0.3952	1
24	0.0843	1
25	0.1734	2
26	0.0366	0
27	0.2397	2
28	0.1244	0
29	0.4795	4
30	0.0935	2

(1) Simulate the hyperparameters α (=1.84) and β (=8.83) from a joint lognormal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$

$$\boldsymbol{\mu} = \begin{bmatrix} \ln 2 \\ \ln 10 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.6^2 & 0.8 \times 0.6 \times 0.9 \\ 0.8 \times 0.6 \times 0.9 & 0.9^2 \end{bmatrix} \tag{21}$$

(2) Simulate the failure rates λ_i for each plant $i=1,\ldots,31$, from the gamma distribution with parameters α and β. The λ_{31}, $\lambda_{31}=0.2733$, will be again estimated using proposed method.

(3) Simulate the failure data from the Poisson distribution for each λ_i. For simplicity, the observation time of each plant is assumed to be the same, 10 years. Therefore, we simulate 31 vectors of failure data x_i for each plant and $x_{31}=3$.

Table 1 summarizes the failure data generated for the other 30 plants and is used to estimate the PVC. For the 31st plant, we have $\lambda_{31}=0.2733$ and $x_{31}=3$.

4.1 Stage 1. The noninformative hyperprior of α and β is adopted in the inference. The marginal hyperposterior probability density functions (PDFs) of α and β are obtained and shown in Fig. 1 by using the data from 30 plants given in Table 1. It shows the distributions are close to the lognormal parent used in simulation. The statistics of α and β obtained from the hyperposterior distribution are compared with those of the lognormal distribution in Table 2. The estimated values from the two-stage Bayes methods are close to those of the parent lognormal distribution. Figure 2 compares the PVC of the failure rate with the assumed parent. It shows that the estimated PVC is close to the parent. The estimated mean value and standard deviation are 0.2189 and 0.159, respectively, which are close to the values 0.2081 and 0.153 calculated from the simulated values of α and β.

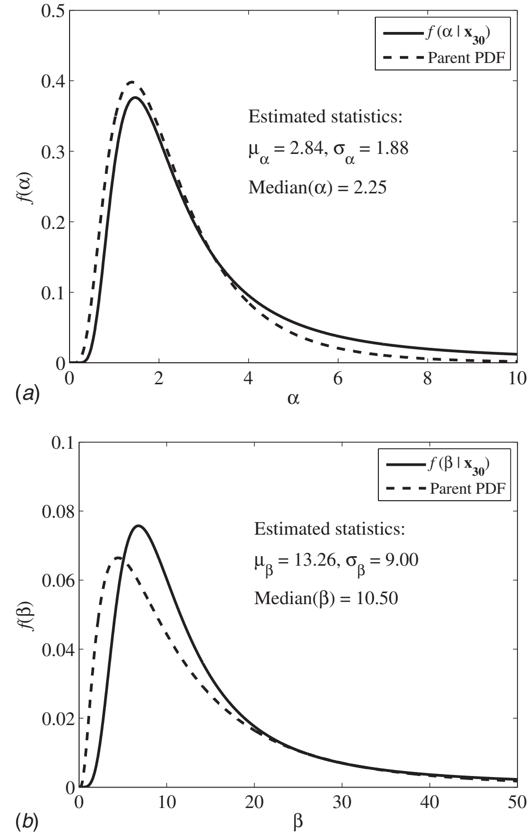

Fig. 1 Posterior distribution of hyperparameters (*a*) α and (*b*) β

4.2 Stage 2. In the stage 2, the PVC obtained from the stage 1 is adopted as the prior distribution of the failure rate λ_{31}. The posterior distribution of λ_{31} is obtained by updating the PVC using data from 31st plant. Both PVC and posterior distribution are shown in Fig. 3. The mean value and standard deviation of λ_{31} are

Table 2 Comparison of statistics

	Mean		Standard deviation		Median	
Parameter	LN	TB	LN	TB	LN	TB
α	3.19	2.84	2.10	1.88	2.00	2.25
β	11.49	13.26	12.85	9.00	10.00	10.50

LN: parameters of parent lognormal distribution.
TB: results obtained from the two-stage Bayes method.

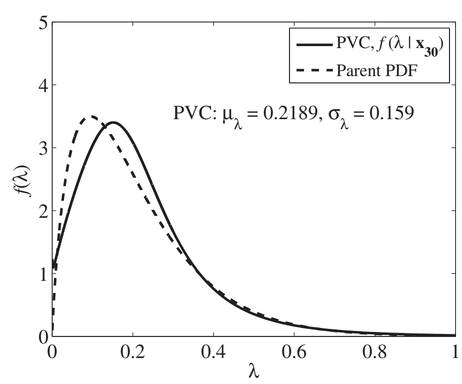

Fig. 2 Comparison of estimated and parent PVCs

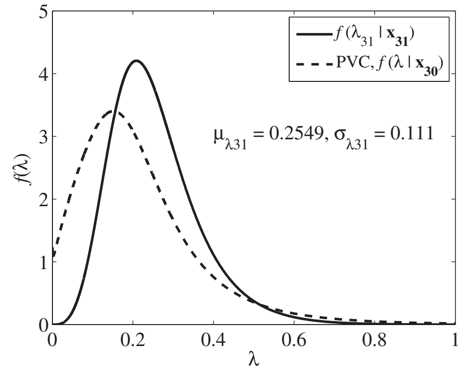

Fig. 3 Estimated PDF of failure rate (λ_{31})

calculated as 0.2549 and 0.111, respectively. The following observations are noteworthy: (1) the estimated mean failure rate is larger than that of PVC and close to 0.2733, which is used to simulate the failure data of the 31st plant. (2) The uncertainty of λ_{31} in terms of standard deviation is 0.111, which is smaller than that of PVC, due to the additional data available of the plant.

5 Application

In this section, we analyze piping leakage data obtained from nuclear power plants (pressurized water reactors (PWRs)). The piping leakage events have been recorded from 1970 up to June 2007. 29 PWR NPPs with initial reactor criticality in 1980s are selected as plants with similar conditions. There are 29 leakage events in 29 PWRs during a total of 652.08 observation years. Table 3 lists the observed number of failure, x_i, and observation period, t_i, in years, for 10 of 29 plants. If the data were from a homogeneous Poisson process, then the constant leak rate can be estimated as the ratio of the number of leakage events and the total exposure time. The leak rate under this assumption is calculated to be 0.045 per plant-year. Three types of Bayesian methods, simple Bayesian method, empirical Bayesian method, and two-stage Bayesian method, are used to estimate the failure rate.

5.1 Single Stage Bayesian Method. The prior distribution of λ can be selected as a gamma distribution with parameters α and β, as in Eq. (2). The mean leak rate is estimated as the ratio of the number of failure to the total operating time. The COV of λ is assigned from the engineering judgment, say, 1.0. For plant 10 in Table 4, x and t are 0 and 24.79, respectively, and α_S and β_S estimated from the other 28 plants are 1.00 and 21.63. Its posterior mean leak rate is therefore 0.0215 with a standard deviation of 0.0215 compared with the prior mean 0.0462, which is the number of leakage events divided by the total operating time of other 28 plants.

Table 3 Mean leak rate (λ) estimated from two-stage Bayes

i	x_i	t_i (year)	λ_i^{PVC}	σ_i^{PVC}	λ_i^{TB}	σ_i^{TB}
1	2	20.21	0.0435	0.0329	0.0573	0.0292
2	1	20.41	0.0452	0.0343	0.0458	0.0241
3	3	22.07	0.0421	0.0303	0.0670	0.0335
4	2	22.30	0.0437	0.0334	0.0558	0.0279
5	1	22.70	0.0453	0.0345	0.0447	0.0233
6	3	23.71	0.0423	0.0308	0.0656	0.0322
7	1	25.53	0.0455	0.0345	0.0433	0.0223
8	2	26.09	0.0440	0.0340	0.0534	0.0260
9	0	19.10	0.0467	0.0311	0.0351	0.0211
10	0	24.79	0.0471	0.0307	0.0332	0.0199

PVC: results of PVC obtained from the two-stage Bayes.
TB: results of specific plant obtained from the two-stage Bayes.

Table 4 Mean leak rate (λ) estimated from simple and empirical Bayes

i	x_i	t_i (year)	λ_i^S	σ_i^S	λ_i^{EB}	σ_i^{EB}
1	2	20.21	0.0688	0.0397	0.0731	0.0448
2	1	20.41	0.0465	0.0329	0.0462	0.0355
3	3	22.07	0.0864	0.0432	0.0938	0.0489
4	2	22.30	0.0658	0.0380	0.0693	0.0425
5	1	22.70	0.0443	0.0313	0.0436	0.0335
6	3	23.71	0.0835	0.0418	0.0905	0.0472
7	1	25.53	0.0417	0.0295	0.0407	0.0312
8	2	26.09	0.0609	0.0352	0.0633	0.0388
9	0	19.10	0.0244	0.0244	0.0210	0.0241
10	0	24.79	0.0215	0.0215	0.0183	0.0209

S: results obtained from simple Bayes.
EB: results obtained from empirical Bayes.

5.2 Empirical Bayesian Method. The prior distribution of λ is selected as a gamma distribution with parameters α and β as did in the simple one-stage Bayesian method. The parameters are estimated using the optimistic method by Vaurio [10]. For plant 29, x and t are 0 and 24.79, respectively, and from the other 28 plants estimates are α_E=0.77 and β_E=17.28. Its posterior mean leak rate is therefore 0.0183 with a standard deviation of 0.0209, compared with the prior mean 0.0445, which is α_E/β_E, with a standard deviation of 0.0507.

5.3 Hierarchical Bayesian Method. Instead of using the point estimate for α and β, the hyperposterior distribution of α and β are estimated by the Bayesian method (Eq. (12)). For plant 29, the posterior mean leak rate is estimated as 0.0332 with a standard deviation of 0.0199, compared with the prior mean of 0.0471 with a standard deviation of 0.0307.

Three Bayesian methods described above are applied to the all 29 NPPs. For the length of the paper, only the results of ten plants are listed in Tables 3 and 4. λ_i^{PVC} in the table denotes the mean leak rate of PVC for ith plant obtained in the first stage, and λ_i^{TB} denotes the mean value of the plant-specific (ith) leak rate obtained by updating the average PVC in the second stage of two-stage Bayes. Note that when the inference is made for the plant i, the failure data (x_i, t_i) of the ith plant are excluded and only used in the second stage, while the data from the other 28 plants are used for estimation of PVC in the first stage. λ^S and λ^{EB} denote the estimates of the simple Bayes and empirical Bayes, respectively.

The results from 29 NPPs and Tables 3 and 4 show the following.

(1) The value of λ_i^{PVC} ranges from 0.0421 to 0.0471 with a mean of 0.0454 and a standard deviation of 0.0021. The small variation in the λ_i^{PVC} shows that the PVCs obtained for different plants are very similar, which implies that the estimated PVC is a good approximation.
(2) Compared with the results from simple Bayesian method, the estimated mean leak rate λ with two-stage Bayes is smaller when failure data are available, i.e., $x>0$. When x=0, two-stage λ is larger than that obtained from simple analysis. However, the uncertainty of λ estimated with two-stage Bayes is about 8–26% smaller, which means that the two-stage Bayes gives more precise estimates.
(3) Compared with the results from empirical Bayes, the estimated mean leak rate λ with two-stage Bayes is slightly smaller when $x>0$, while it is slightly larger when x=0. However, the uncertainty of λ estimated with two-stage Bayes is about 5–30% smaller, which means that the two-stage Bayes gives more precise estimates.

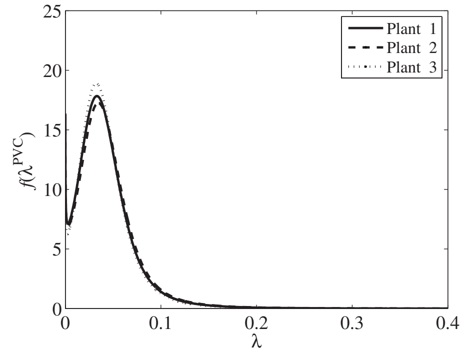

Fig. 4 PVCs for different NPPs

Figure 4 shows the PVCs obtained for plants 1, 2, and 3 with number of failures 2, 1, and 3, respectively. Although the number of failures varies from 1 to 3, the PVCs are quite similar, which implies that it is appropriate to select the PVC obtained in the first stage as the prior distribution in the second stage inference. Figure 5 shows PDFs of plant-specific leak rate. The estimated mean values of the leak rate for ten plants are listed in Table 3.

Figure 6 illustrates the posterior distribution of leak rate for plant 1 obtained from three Bayesian methods. It can be clearly seen that the uncertainty of the leak rate estimated from two-stage Bayes is quite smaller than the other methods. And as aforementioned, the posterior distribution estimated by simple Bayes is greatly affected by the prior because only the information from the plant itself is used.

6 Conclusions

This paper describes a two-stage Bayesian model for estimating the plant-specific failure rate. The general hierarchical Bayesian model is described and compared with the simple and empirical Bayesian models. A simulation study is presented to illustrate the use of the two-stage Bayesian procedure with the failure data generated from Poisson process and prior gamma density. The simulation results show that the two-stage Bayesian method works well in the inference.

The two-stage Bayesian method is applied to the pipe leak data collected from the U.S. PWRs. The results show that the population variability curves obtained in the first stage for different NPPs are very close to each other, which imply that the estimated PVC can serve as a prior in the plant-specific failure rate analysis. By using the pipe leak data of a specific plant, the PVC is updated and thus plant-specific distribution of the leak rate is obtained. The results of the simple Bayes and empirical Bayes are presented for comparison. The paper shows that the uncertainty associated with leak rate estimated from the two-stage Bayes is smaller than that associated with estimates of the single stage Bayesian method.

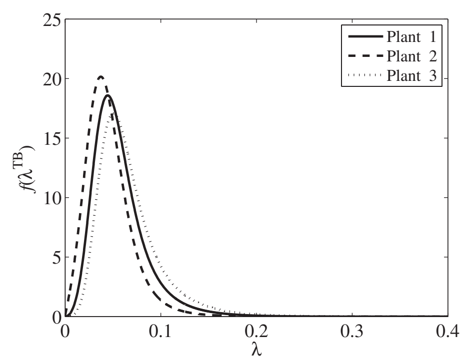

Fig. 5 Plant-specific pipe leak rate for different NPPs

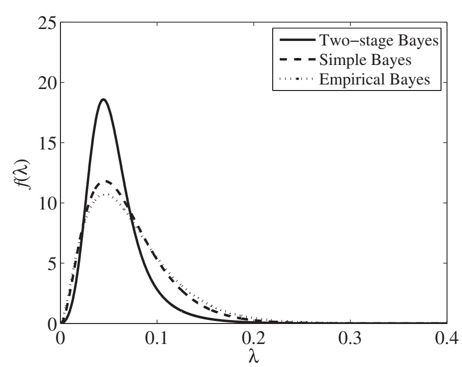

Fig. 6 Comparison of posterior distributions of λ for Plant 1 obtained from different Bayesian methods

Acknowledgment

The authors are grateful to the Canadian Nuclear Safety Commission (CNSC), Natural Sciences and Engineering Research Council of Canada (NSERC), and the University Network of Excellence in Nuclear Engineering (UNENE) for providing the financial support for this study. We would like to thank Mr. Richard Wiebe, a graduate student, for his technical assistance in FORTRAN programming.

Nomenclature

COV = coefficient of variance
EB = empirical Bayes
NEA = Nuclear Energy Agency
NPP = nuclear power plant
OECD = Organization for Economic Co-operation and Development
PDF = probability density function
PSA = probabilistic safety assessment
PVC = population variability curve
PWR = pressurized water reactor
RI-ISI = risk-informed in-service inspection
iid = identically and independently distributed

Appendix A: Derivation of Eq. (12)

By Bayes' theorem, the hyperposterior distribution $f(\theta|\mathbf{x_n})$ can be written as

$$f(\theta|\mathbf{x_n}) \propto f(\mathbf{x_n}|\theta)f(\theta) \tag{A1}$$

Extending the Eq. (5) for mixed distribution $f(\mathbf{x_n}|\theta)$, we get

$$f(\theta|\mathbf{x_n}) \propto \left[\int_{\lambda_1}\cdots\int_{\lambda_n} f(\mathbf{x_n}|\lambda_1,\cdots,\lambda_n,\theta)f(\lambda_1,\cdots,\lambda_n|\theta)d\lambda_1\cdots d\lambda_n\right] \times f(\theta) \tag{A2}$$

By Assumption 3 that given θ and $\{\lambda_i\}$, x_i have density $f(x_i|\lambda_i)$, independent of θ and of all λ's other than λ_i, we have

$$f(\mathbf{x_n}|\lambda_1,\ldots,\lambda_n,\theta) = \prod_{i=1}^{n} f(x_i|\lambda_i) \tag{A3}$$

and by Assumption 2 that given θ, $\{\lambda_i\}$ is iid, we get

$$f(\lambda_1,\ldots,\lambda_n|\theta) = \prod_{i=1}^{n} f(\lambda_i|\theta) \tag{A4}$$

Equation (A2) becomes

$$f(\theta|\mathbf{x_n}) \propto \left\{ \int_{\lambda_1} \cdots \int_{\lambda_n} \prod_{i=1}^{n} f(x_i|\lambda_i) f(\lambda_i|\theta) d\lambda_1 \cdots d\lambda_n \right\} f(\theta) \tag{A5}$$

Since the product now involved no "cross terms," we may interchange the product operator and the integral and get Eq. (12).

Appendix B: Derivation of Eq. (17)

The part enclosed in the square brackets of the Eq. (12) can be interpreted as the likelihood of $\mathbf{x_n}$ given fixed θ, $f(\mathbf{x_n}|\theta)$, where θ can be a vector, for example, $\boldsymbol{\theta}=(\alpha,\beta)$, then Eq. (12) becomes in the form of Eq. (16).

Suppose that the failure event follows Poisson distribution with density $f(x|\lambda)$ expressed in Eq. (1), and further suppose that λ follows gamma distribution $f(\lambda|\alpha,\beta)$ expressed in Eq. (2). With Eqs. (1) and (2), the integrant in Eq. (16) can be written as

$$\frac{(\lambda_i t_i)^{x_i} \exp(-\lambda_i t_i)}{x_i!} \cdot \frac{\lambda_i^{\alpha-1} \exp(-\beta\lambda_i)\beta^\alpha}{\Gamma(\alpha)} = \frac{\lambda_i^{x_i+\alpha-1} \exp[-\lambda_i(t_i+\beta)](t_i+\beta)^{x_i+\alpha}}{\Gamma(x_i+\alpha)} \cdot \frac{\Gamma(x_i+\alpha)}{x_i!\,\Gamma(\alpha)} \cdot \frac{t_i^{x_i}\beta^\alpha}{(t_i+\beta)^{x_i+\alpha}} \tag{B1}$$

The first term of the right side of Eq. (B1) is the gamma density with parameter $(x_i+\alpha, t_i+\beta)$. Because the second and third terms do not include λ_i, the integration with respect to λ_i is simply the product of the second and the third term. The likelihood function $f(\mathbf{x_n}|\alpha,\beta)$ in Eq. (16) reduces to Eq. (17).

References

[1] USNRC, 1975, "Reactor Safety Study: An Assessment of Accident Risk in U.S. Commercial Nuclear Power Plants," Wash-1400, U.S. Nuclear Regulatory Commission, Washington, DC, Report No. NUREG-75/014.

[2] USNRC, 1990, "Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants," U.S. Nuclear Regulatory Commission, Washington, DC, Report No. NUREG-1150.

[3] Tregoning, R., Abramson, L., and Scott, P., 2005, "Estimating Loss-of-Coolant Accident (LOCA) Frequencies Through the Elicitation Process-Draft Report for Comment," U.S. Nuclear Regulatory Commission, Washington, DC, Report No. NUREG-1829.

[4] Atwood, C. L., LaChance, J. L., Martz, H. F., Anderson, D. J., Englehardte, M., Whitehead, D., and Wheeler, T., 2003, "Handbook of Parameter Estimation for Probabilistic Risk Assessment," U.S. Nuclear Regulatory Commission, Washington, DC, Report No. NUREG/CR-6823.

[5] Apostolakis, G., 1978, "Probability and Risk Assessment: The Subjectivistic Viewpoint and Some Suggestions," Nuclear Safety, **19**, pp. 305–315.

[6] Parry, G. W., and Winter, P. W., 1981, "Characterization and Evaluation of Uncertainty in Probabilistic Risk Analysis," Nuclear Safety, **22**, pp. 28–42.

[7] Siu, N. O., and Kelly, D. L., 1998, "Bayesian Parameter Estimation in Probabilistic Risk Assessment," Reliab. Eng. Syst. Saf., **62**, pp. 89–116.

[8] Apostolakis, G., Kaplan, S., Garrick, B. J., and Duphily, R. J., 1980, "Data Specialization for Plant Specific Risk Studies," Nucl. Eng. Des., **56**, pp. 321–329.

[9] Apostolakis, G., 1982, "Data Analysis in Risk Assessments," Nucl. Eng. Des., **71**, pp. 375–381.

[10] Vaurio, J. K., 1987, "On Analytic Empirical Bayes Estimation of Failure Rates," Risk Anal., **7**, pp. 329–338.

[11] Kaplan, S., and Garrick, B. J., 1981, "On the Quantitative Definition of Risk," Risk Anal., **1**, pp. 11–27.

[12] Kaplan, S., 1983, "On a 'Two-Stage' Bayesian Procedure for Determining Failure Rates From Experimental Data," IEEE Trans. Power Appar. Syst., **PAS-102**, pp. 195–202.

[13] Fröhner, F. H., 1985, "Analytic Bayesian Solution of the Two-Stage Poisson-Type Problem in Probabilistic Risk Analysis," Risk Anal., **5**, pp. 217–225.

[14] Fröhner, F. H., 1985, "The Two-Stage Poisson-Type Problem in Probabilistic Risk Analysis," Risk Anal., **5**, pp. 231–234.

[15] Pörn, K., 1996, "The Two-Stage Bayesian Method Used for the T-Book Application," Reliab. Eng. Syst. Saf., **51**, pp. 169–179.

[16] Berger, J. O., 1985, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer-Verlag, New York.

[17] Cooke, R., Dorrepaal, J., and Bedford, T., 1995, "Review of SKI Data Processing Methodology," Delft University of Technology, SKI Report No. 95:2.

[18] Meyer, W., and Hennings, W., 1999, "Prior Distributions in Two-Stage Bayesian Estimation of Failure Rates," *Safety and Reliability: Proceedings of the ESREL '99—Tenth European Conference*, Vol. 2, P. Kafka and G. I. Schueller, eds., Balkema, Rotterdam, The Netherlands, pp. 893–898.

[19] Hora, S. C., and Iman, R. L., 1990, "Bayesian Modeling of Initiating Event Frequencies at Nuclear Power Plants," Risk Anal., **10**, pp. 103–109.

[20] Hofer, E., Hora, S. C., Iman, R. L., and Peschke, J., 1997, "On the Solution Approach for Bayesian Modeling of Initiating Event Frequencies and Failure Rates," Risk Anal., **17**, pp. 249–252.

[21] Hofer, E., 1999, "On Two-Stage Bayesian Modeling of Initiating Event Frequencies and Failure Rates," Reliab. Eng. Syst. Saf., **66**, pp. 97–99.

[22] Hofer, E., and Peschke, J., 1999, "Bayesian Modeling of Failure Rates and Initiating Event Frequencies," *Safety and Reliability: Proceedings of the ESREL '99—Tenth European Conference*, Vol. 2, P. Kafka and G. I. Schueller, eds., Balkema, Rotterdam, The Netherlands, pp. 881–886.

[23] Cooke, R., Bunea, C., Charitos, T., and Mazzuchi, T. A., 2002, "Mathematical Review of ZEDB Two-Stage Bayesian Models," Department of Mathematics, Delft University of Technology, Report No. 02-45.

[24] Bunea, C., Charitosb, T., Cookec, R. M., and Becker, G., 2003, "Two-Stage Bayesian Models—Application to ZEDB Project," *Safety and Reliability: Proceedings of the ESREL 2003*, T. Bedford and P. H. J. A. M. van Gelder, eds., Balkema, Lisse, The Netherlands, pp. 321–331.

[25] Goel, P. K., and DeGroot, M. H., 1981, "Information About Hyperparameters in Hierarchical Models," J. Am. Stat. Assoc., **76**, pp. 140–147.

[26] Goel, P. K., 1983, "Information Measure and Bayesian Hierarchical Models," J. Am. Stat. Assoc., **78**, pp. 408–410.

[27] Box, G. E. P., and Tiao, G. C., 1973, *Bayesian Inference in Statistical Analysis*, Wiley, New York.

[28] Jeffreys, H., 1961, *Theory of Probability*, 3rd ed., Clarendon, Oxford.