versão impressa ISSN 0101-7438 / versão online ISSN 1678-5142

BAYESIAN ASSESSMENT OF THE VARIABILITY OF RELIABILITY MEASURES

Enrique López Droguett *
Departamento de Engenharia de Produção
Universidade Federal de Pernambuco (UFPE)
Recife – PE
ealopez@ufpe.br

Frank J. Groen
Ali Mosleh
Reliability Engineering Program
University of Maryland
College Park – USA

* *Corresponding author* / autor para quem as correspondências devem ser encaminhadas

Recebido em 01/2005; aceito em 09/2005 após 1 revisão
Received January 2005; accepted September 2005 after one revision

Abstract

Population variability analysis, also known as the first stage in two-stage Bayesian updating, is an estimation procedure for the assessment of the variability of reliability measures among a group of sub-populations of similar systems. The estimated variability distributions are used as prior distributions in system-specific Bayesian updates. In this paper we present a Bayesian approach for population variability analysis involving the use of non-conjugate variability models that works over a continuous, rather than the discretized, variability model parameter space. The cases to be discussed are the ones typically encountered by the reliability practitioner: run-time data for failure rate assessment, demand-based data for failure probability assessment, and expert-based evidence for failure rate and failure probability analysis. We outline the estimation procedure itself as well as its link with conventional Bayesian updating procedures, describe the results generated by the procedures and their behavior under various data conditions, and provide numerical examples.

Keywords: Bayes' theorem; population variability; reliability.

Resumo

Análise de variabilidade populacional, também conhecida como o primeiro estágio no processo de atualização Bayesiana em dois estágios, é um procedimento de estimação utilizado para a quantificação da variabilidade de métricas de confiabilidade num conjunto de sub-populações de sistemas similares. As distribuições de variabilidade obtidas são usadas como distribuições a priori em atualizações Bayesianas específicas para um sistema. Neste artigo, apresenta-se um procedimento Bayesiano para a análise da variabilidade populacional envolvendo o uso de modelos de variabilidade não-conjugados que utilizam um espaço contínuo, ao invés de discreto, dos parâmetros do modelo de variabilidade. Discutem-se casos tipicamente encontrados na prática: dados de falha no tempo, dados de falha sob demanda e opiniões de especialistas para a estimação da taxa e probabilidade de falha. O procedimento de estimação é discutido, estabelece-se sua ligação com procedimentos convencionais de atualização Bayesiana e descrevem-se resultados gerados e o comportamento dos mesmos sob diferentes condições de dados, apresentando-se também exemplos numéricos.

Palavras-chave: teorema de Bayes; variabilidade populacional; confiabilidade.

1. Introduction

When samples are identical with respect to some reliability measure, we have the so-called homogeneous samples. These type of samples are typically obtained from components or systems that have been operating under the same environmental and operational conditions, and a Bayesian sequential updating can be performed such that, data from one subset is used to update the state-of-knowledge obtained based on data from another subset. The same result can be reached by aggregating all data sets and then applying the Bayes' theorem. For example, it is possible to sum up the number of failures and time in service in each sample, and then update the failure rate estimate via Bayes' theorem by using the aggregated number of failures and total time in service. For a detailed discussion on Bayesian parameter estimation, please refer to Siu & Kelly (1998).

However, in many situations we do not have homogeneous samples. Samples for a system or component submitted to different operational and environmental conditions, design and production differences, present different reliability measures. They may have different failure modes, failure rates, and repair rates. For instance, pressure-temperature sensors operating under different temperature and vibration conditions may have different failure rates. Similarly, variations in maintenance and reliability practices of different natural gas suppliers can lead to different failure rates of identical pressure control valves.

In such cases, it is not realistic to assume that all items of a population composed of different samples (sub-populations) have the same reliability measures (e.g., failure rate). We have an inherent variability of the reliability measures among the samples. We say that the population is non-homogeneous. In other words, the failure rate or any other reliability parameter is inherently different from sample to sample in the population. This is the so-called *population variability* of the reliability measure of interest (failure rate). It is important to stress that such variability is a characteristic of the system and it is not changed just with an increased amount of evidence. It will only change if the actual reliability characteristics of the item change (for a discussion on the concept of population variability, the reader may refer to Siu & Kelly, 1998).

In this context, population variability analysis is a technique for arriving at a prior distribution for Bayesian reliability parameter assessments based on partially relevant data. It concerns the assessment of the inherent variability in reliability measures such as failure rates among a number of sub-populations. The variability analysis is known as the first stage of a two-stage Bayesian parameter updating procedure introduced, in the context of reliability and risk analysis, by Kaplan (1983). The second stage of the procedure corresponds to the updating based on system specific data and having the parameter's population variability distribution as the prior.

Also known as non-homogeneous analysis and hierarchical Bayes, population variability analysis has been recognized as computationally demanding, Hora (1990). As a result, proposed solutions to the problem have introduced conceptual limitations in order to make the approaches computationally less demanding. Indeed, under the umbrella of Bayes Empirical Bayes methods, Kaplan (1983) introduces a discretization of the parameter space, while the approaches proposed by Hora (1990) and Pörn (1996) make use of conjugate variability models (Bernardo, 1994).

Other methods that are not fully Bayesian have also been proposed. Frohner (1985) approach consists in the construction of the variability distribution by superimposing posterior distributions obtained for the individual sub-populations. Vaurio (1987) presents a matching-moment procedure for the selection of a single Gamma distribution. However, it has been recognized that these procedures have severe conceptual limitations that handicap their applicability (Kaplan, 1985).

In order to circumvent these limitations of both Bayesian and non-fully Bayesian approaches, Droguett, Groen & Mosleh (2004) proposed a fully Bayesian population variability analysis methodology that makes possible the combined use of run-time data (e.g., failures and operational time) and expert opinions resulting in a set of mixed likelihood models. However, this approach is only discussed in the context of these mixed likelihood models and there is no attempt to develop models for cases where either run-time data or expert-based evidence is available.

Therefore, in this paper we present the building blocks of a fully Bayesian approach for population variability analysis involving the use of non-conjugate variability models, i.e., when the probability distribution for the variability and the likelihood function for the evidence do not result in a analytical model as the posterior distribution. Furthermore, the approach works over a continuous, and not discretized, variability model parameter space. The cases to be discussed are the ones frequently encountered by the reliability and risk practitioner: run-time data for failure rate (or repair rate) assessment, demand-based data for failure probability assessment, and expert-based evidence for failure rate and failure probability analysis. Some elements of the approach in this paper have previously been presented in Droguett & Groen (2004).

This paper also discusses the interpretation of the results generated by population variability analysis, as well as some behavioral properties, with the objective to illustrate the usefulness of this type of analysis in practical situations. The topics covered include lessons learned during the development of an algorithm for performing population variability assessment using a Markov Chain Monte Carlo method (Chib & Greenberg, 1995; Gilks, Richardson & Spiegelhalter, 1996), as well questions commonly asked by users of the software in which the algorithm is incorporated.

The paper starts out by discussing the population variability analysis procedure, including a presentation of the probabilistic model that forms the basis for the analysis construction. Various types of likelihood models are developed for situations that make use of partially relevant test and field data, as well as the use of engineering judgments. The concept of mixed likelihoods is also introduced. Next, the paper focuses on the form and interpretation of the results generated by a population variability analysis. These results include the population variability distribution, which consists of a weighted sum of distributions, and the estimation of variability measures. Then, it is shown how this distribution can be used as a prior distribution in a system-specific parameter estimation procedure. In the next section, a detailed discussion of the behavior of the population variability distribution under a variety of data conditions is illustrated. Examples of application are presented in sections 7 and 8. Concluding remarks are then provided.

2. Model-Based Variability Estimation

Let us now suppose that in estimating a reliability measure x of an item, the available evidence E forms a non-homogeneous population. We then need to assess the population variability distribution, $\varphi(x)$, of the reliability measure of interest. Let us also assume that the population variability is a member of a parametric family of distributions, and $\underline{\theta} = \{\theta_1, \ldots, \theta_n\}$ is the set of parameters, i.e., $\varphi(x) = \varphi(x \mid \theta_1, \ldots, \theta_n)$. We have that the uncertainty distribution over the space of $\varphi(x \mid \underline{\theta})$ is the same as the uncertainty over values of $\underline{\theta}$, as for each value of $\underline{\theta}$ there is a unique $\varphi(x \mid \underline{\theta})$ and vice-versa. Therefore, our goal of assessing $\varphi(x \mid \underline{\theta})$ is reduced to estimating $\underline{\theta}$.

By considering that our prior state-of-knowledge about $\underline{\theta}$ is represented by the probability distribution $\pi_o(\underline{\theta})$, and given the available evidence *E*, we use the Bayes' theorem to find the posterior probability distribution over $\underline{\theta}$:

$$\pi(\underline{\theta} \mid E) = \frac{L(E \mid \underline{\theta}) \cdot \pi_0(\underline{\theta})}{\int_{\underline{\theta}} L(E \mid \underline{\theta}) \cdot \pi_0(\underline{\theta}) \cdot d\underline{\theta}} \qquad (1)$$

where $L(E \mid \underline{\theta})$ is the likelihood of evidence *E* given $\underline{\theta}$, and $\pi(\underline{\theta} \mid E)$ is the posterior distribution of $\underline{\theta}$ given evidence *E*.

3. Bayesian Revision of Parameter Distributions

To perform a population variability analysis of a reliability measure we need to specify an appropriate probability distribution to describe the underlying variability of the measure of interest, $\varphi(x \mid \underline{\theta})$, as well as construct the likelihood function $L(E \mid \underline{\theta})$. The likelihood construction is obviously an evidence-driven process, i.e., it is dependent on the type of available evidence. We will consider two categories: (i) data-based likelihoods corresponding to evidence of type number of failures and exposure (time in service or number of demands); (ii) expert-based likelihoods that corresponds to estimates of possible values of a reliability measure. The specification of the probability distribution describing the variability may be guided by the nature of the reliability measure, for instance, a gamma distribution for failure rate, or a beta distribution in case of probability of failure.

3.1 Data-based likelihoods

Let us assume that we are interested in assessing the population variability of an item failure rate, λ, and the available evidence is $\{(k_i, T_i),\ i = 1, \ldots n\}$, where k_i is the number of failures and T_i is the time to observe the k_i failures in the *i* sample, and *n* is the total number of samples. If we know the failure rate $\lambda_i = \lambda$ of each sample, we can use the Poisson distribution to estimate the likelihood of observing k_i failures in T_i:

$$P(k_i, T_i \mid \lambda) = \frac{(\lambda T_i)^{k_i}}{k_i!} e^{-\lambda T_i} \qquad (2)$$

As we only know that λ is one of the possible values of the failure rate represented by its population variability distribution $\varphi(\lambda \mid \underline{\theta})$, we average the likelihood over all possible values of λ in order to calculate the probability of the data unconditional on the unknown value of λ:

$$L(k_i, T_i \mid \underline{\theta}) = \int_0^\infty \frac{(\lambda T_i)^{k_i}}{k_i!} e^{-\lambda T_i} \varphi(\lambda \mid \underline{\theta}) d\lambda \tag{3}$$

If the population variability distribution is a Gamma with parameters α and β, we have a *Gamma-Poisson likelihood*:

$$\begin{aligned} L(k_i, T_i \mid \alpha, \beta) &= \int_0^\infty \frac{(\lambda T_i)^{k_i}}{k_i!} e^{-\lambda T_i} \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)} \beta^\alpha e^{-\beta\lambda} d\lambda \\ &= \frac{T_i^{k_i} \Gamma(\alpha + k_i) \beta^\alpha}{k_i! \Gamma(\alpha)(\beta + T_i)^{\alpha + k_i}} \end{aligned} \tag{4}$$

Another possible choice for the variability distribution is a Lognormal with parameters μ and σ, resulting in the *Lognormal-Poisson likelihood* as follows:

$$\begin{aligned} L(k_i, T_i \mid \mu, \sigma) &= \int_0^\infty \frac{(\lambda T_i)^{k_i}}{k_i!} e^{-\lambda T_i} \frac{1}{\sqrt{2\pi}\lambda\sigma} e^{-\frac{1}{2}\left(\frac{\ln\lambda - \ln\mu}{\sigma}\right)^2} d\lambda \\ &= \frac{T_i^{k_i}}{k_i! \sqrt{2\pi}\lambda\sigma} \int_0^\infty \lambda^{k_i - 1} e^{-\frac{1}{2}\left(\frac{\ln\lambda - \ln\mu}{\sigma}\right)^2 - \lambda T_i} d\lambda \end{aligned} \tag{5}$$

Another important case is the assessment of the variability of the failure probability *p*. Let us assume that the non-homogeneous evidence is $\{(k_i, D_i),\ i = 1, \ldots, n\}$, where now D_i is the number of demands in the *ith* sample. We now use the Binomial distribution to estimate the likelihood of observing k_i failures in D_i:

$$P(k_i, D_i \mid p) = \binom{D_i}{k_i} p^{k_i} (1-p)^{D_i - k_i} \tag{6}$$

As before, to find the probability of the data unconditional on the unknown value of *p*, we average the likelihood over all possible values of *p*:

$$L(k_i, D_i \mid \underline{\theta}) = \int_0^1 \binom{D_i}{k_i} p^{k_i} (1-p)^{D_i - k_i} \varphi(p \mid \underline{\theta}) dp \tag{7}$$

where $\varphi(p \mid \underline{\theta})$ is the probability of failure population variability. If the variability is represented by a Beta distribution with parameters α and β, we have a *Beta-Binomial likelihood*:

$$\begin{aligned} L(k_i, D_i \mid \alpha, \beta) &= \int_0^1 \binom{D_i}{k_i} p^{k_i} (1-p)^{D_i - k_i} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} dp \\ &= \binom{D_i}{k_i} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + k_i)\Gamma(\beta + D_i - k_i)}{\Gamma(\alpha + \beta + D_i)} \end{aligned} \tag{8}$$

3.2 Expert-based likelihood

In this section we consider the situation where independent experts provide estimates about a reliability characteristic (e.g., an item failure rate or probability of failure) with underlying variability. As noted by Mosleh (1992), this variability reflects the experts' ability to provide assessments and the difference among them.

Therefore, assume that we are interested in assessing the variability of an item failure rate and the available evidence is in the form of estimates of possible values of the failure rate and the analyst's measure of confidence in each expert, which is considered to be represented by the multiplicative error model (Mosleh & Apostolakis, 1986). That is, the data is $\{(\lambda_i, \sigma_i),\ i = 1, \ldots, n\}$, where λ_i is the opinion provided by expert *i*, σ_i is the logarithmic standard deviation of λ_i representing the uncertainty of expert *i*, and *n* is the total number of experts. Note that σ_i can be interpreted as the analyst's confidence on the i^{th} expert. Provided that we know the failure rate λ_i of each sample, we use the Lognormal distribution with median $\ln \lambda$ to assess the likelihood of observing λ_i:

$$L(\lambda_i, \sigma_i \mid \lambda) = \frac{1}{\sqrt{2\pi}\lambda_i \sigma_i} e^{-\frac{1}{2}\left(\frac{\ln \lambda_i - \ln \lambda}{\sigma_i}\right)^2} \tag{9}$$

Provided that the failure rate has variability given by $\varphi(\lambda \mid \underline{\theta})$, the probability of data unconditional on the unknown value of λ is obtained by averaging the likelihood over all possible values of λ, i.e.,

$$L(\lambda_i, \sigma_i \mid \mu, \sigma) = \int_0^\infty \frac{1}{\sqrt{2\pi}\lambda_i \sigma_i} e^{-\frac{1}{2}\left(\frac{\ln \lambda_i - \ln \lambda}{\sigma_i}\right)^2} \frac{1}{\sqrt{2\pi}\lambda \sigma} e^{-\frac{1}{2}\left(\frac{\ln \lambda - \ln \mu}{\sigma}\right)^2} d\lambda \tag{10}$$

where we have considered that the population variability is represented by a Lognormal distribution with parameters μ and σ:

$$\varphi(\lambda \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\lambda \sigma} e^{-\frac{1}{2}\left(\frac{\ln \lambda - \ln \mu}{\sigma}\right)^2} \tag{11}$$

The Eq. (10) can be solved to provide the *Lognormal-Lognormal likelihood*:

$$L(\lambda_i, \sigma_i \mid \mu, \sigma) = \frac{1}{\sqrt{\frac{1}{\sigma_i^2} + \frac{1}{\sigma^2}}} \frac{e^{-\frac{1}{2}\frac{(\ln \lambda_i - \ln \mu)^2}{\sigma^2 + \sigma_i^2}}}{\sqrt{2\pi}\sigma_i^2 \sigma \lambda_i} \tag{12}$$

3.3 Mixed likelihoods

Each one of the preceding likelihood models were constructed assuming one source of evidence as an indication of a measure variability. In many situations, however, additional information might be available. For instance, we could have (k_i, T_i) data as well as experts' opinions (λ_i, σ_i) regarding an item failure rate variability. Different likelihoods are possible depending on the nature of evidence and the choice of the underlying variability distribution, which give rise to the following *Mixed Likelihood* models:

Table 1 – Mixed likelihood functions.

Prior/Evidence	(k_i,T_i)	(λ_i, σ_i)	(k_i,D_i)
Gamma	Gamma-Poisson-Lognormal		
LogNormal	Lognormal-Poisson-Lognormal		
		Lognormal-Binomial-Lognormal	
Beta		Beta-Binomial-Lognormal	

In this article, however, we focus on the building blocks of these mixed models, namely the gamma-poisson, lognormal-poisson, beta-binomial and lognormal-lognormal models, and their behavior under various data conditions as discussed in the next sections. For a detailed discussion on the mixed likelihood models see Droguett, Groen & Mosleh (2004).

4. Variability Measures

At this point, we assume that we have a model of the variability of x, $\varphi(x|\theta_1,...,\theta_n)$, as well as a distribution $\pi(\theta_1,...,\theta_n)$. We first consider the construction of the best estimate of the variability density function. This estimate is arrived at by averaging $\varphi(x|\theta_1,...,\theta_n)$ using $\pi(\theta_1,...,\theta_n)$ as a weighting function

$$\hat{p}(x) = \int \cdots \int_{\theta_1,...,\theta_n} \phi(x \mid \theta_1,...,\theta_n) \cdot \pi(\theta_1,...,\theta_n) \cdot d\theta_1 \cdots d\theta_n \tag{13}$$

The estimated density function therefore consists of a weighted mix of distributions of the chosen model, as opposed to being formed by a single 'best' distribution chosen from the set of distributions possible within the definition of the model, e.g., a Maximum Likelihood estimator. The estimated mean of the population variability distribution is obtained by computing the mean of the estimated density function

$$\hat{\mu}_x = \int_x x \cdot \hat{p}(x) \cdot dx \tag{14}$$

Similarly, the estimated variance is defined as

$$\sigma_x^2 = \int_x (x - \hat{\mu}_x)^2 \cdot \hat{p}(x) \cdot dx \tag{15}$$

In addition to these point estimators it is possible to estimate uncertainty distributions for measures such as the cumulative variability density function $P(x)$, z-percentile x_z, mean μ_x, and variance σ_x^2. These distributions provide an indication of the degree of uncertainty associated with the estimated variability distribution, and thus due to the limited amount of knowledge about the populations under consideration.

The concept of these uncertainty distributions is shown in Figure 1. The figures shows a joint density expressing the likelihood that a particular distribution $\varphi(x| \theta_1, \theta_2)$ represents the variability among the subpopulations. For instance, consider that we are assessing the variability of the failure rate among a number of populations, and that we believe the variability to be distributed lognormally. The joint density distribution will then be defined over the parameter space (ν, τ). The higher the value of $\pi(\nu, \tau)$, the better the match between the distribution $\mathrm{LN}(x|\nu, \tau)$ and the distribution of failure rates indicated by the data.

Figure 1(b) shows a number of the results in the density space. Each thin solid lines represents a unique distribution $\varphi(x|\theta_1, \theta_2)$. The thicker solid line represents the estimated population variability distribution $\hat{p}(x)$, computed according to Eq. (13) referred to as the 'mean distribution'. In contrast, the distribution of the means is represented by the dashed line. This distribution does not express a degree of variability, but rather the extent of uncertainty about the mean of the variability distribution.

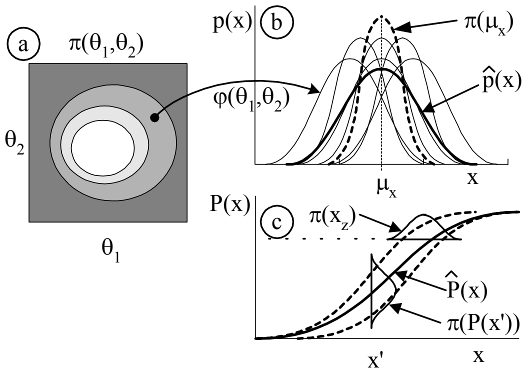

Figure 1 – Plots of the various estimation results.

The solid line in Figure 1(c) represents the estimated cumulative distribution $\hat{P}(x)$. This value is interpreted as saying that the estimated fraction of the populations for which the value of X is smaller than x equals $\hat{P}(x)$. The dashed lines in this plot show the degree of uncertainty about this estimate in the form of percentile curves $P_\alpha(x)$, constructed by plotting the α-percentile of distributions $\pi(P(x))$, evaluated for each value x. Simultaneously, these bounds can be interpreted as the α uncertainty bounds on the z-percentiles of the variability distribution.

5. Model of Probabilistic Dependencies

The primary application of population variability analysis in the risk and reliability context is that of the construction of generic prior distributions for system-specific analyses. In Kaplan (1983), the combination of population variability analysis and system-specific update is referred to as a two-stage Bayesian procedure. The result of a population variability analysis consists of an estimate of the distribution of parameters such as failure rates, among a number of populations. In case of system-specific analyses, this distribution serves as a prior to estimate that parameter for a single population. In other words, the two-stage Bayesian updating procedure is considered to be a procedure that in fact involves two separate updating rounds.

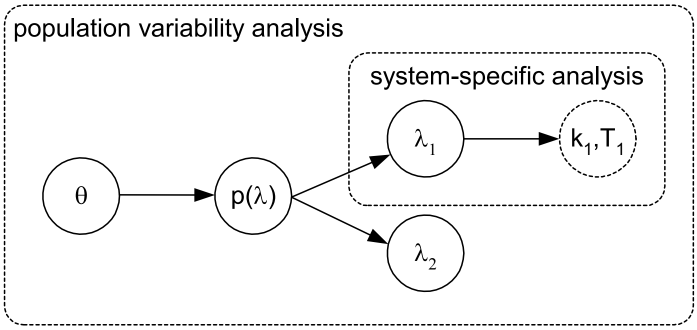

Figure 2 – Model of probabilistic dependences in combined population variability – system specific analysis.

By considering the probabilistic dependencies involved in the problem, it can be shown that both stages can be integrated into a single problem. In fact, a model of probabilistic dependencies in the combined problem, in the form of a belief network (Pearl, 1998), is shown in Figure 2. A similar illustration was presented earlier by Pörn (1996). Given a probability distribution $\pi(\alpha,\beta)$, we can find a prior $\pi(\lambda_i)$ using Eq. (13), which is based on the dependencies between (α,β) and $p(\lambda)$, and $p(\lambda)$ and λ_i, that is:

$$\pi(\lambda_i) = \iint_{\alpha,\beta} \pi(\lambda_i \mid \alpha,\beta)\cdot\pi(\alpha,\beta)\cdot d\alpha\cdot d\beta \tag{16}$$

The probability distribution $\pi(\lambda_i)$ is referred to as the generic prior and represents the uncertainty about both λ_1 and λ_2 in case no other information about either population was available.

Suppose now that for population 1, k_1 failures were observed during a total time of operation T_1. In a system-specific analysis, we would use the dependency between λ_1 and the observation as a likelihood function in Bayes Theorem

$$\pi(\lambda_1 \mid k_1,T_1) = \frac{\Pr(k_1,T_1 \mid \lambda_1)\cdot\pi(\lambda_1)}{\Pr(k_1,T_1)} \tag{17}$$

However, the evidence could simultaneously be used to update the distribution over α and β

$$\begin{aligned}\pi(\alpha,\beta,\lambda_1 \mid k_1,T_1) &= \frac{\Pr(k_1,T_1 \mid \lambda_1,\alpha,\beta)\cdot\pi(\lambda_1,\alpha,\beta)}{\Pr(k_1,T_1)}\\ &= \frac{\Pr(k_1,T_1 \mid \lambda_1)\cdot\pi(\lambda_1 \mid \alpha,\beta)\cdot\pi(\alpha,\beta)}{\Pr(k_1,T_1)}\end{aligned} \tag{18}$$

If we marginalize this distribution to $\pi(\alpha,\beta \mid k_1,T_1)$ by integrating over λ, we obtain precisely the expression of Eq. (4). This illustrates that k_1 and T_1 theoretically affects both the system-specific distribution $\pi(\lambda_1)$ as well as the population variability distribution in the form of $\pi(\alpha,\beta)$ and thus derived distributions such as $\pi(\lambda_2)$.

This model illustrates that when performing a two-stage procedure, data used in the second, system-specific, stage, should not be included in the first stage of the analysis, in order to avoid 'double-counting' of the data.

6. Behavior under Various Data Conditions

The amount of data available to perform a population variability analysis can be characterized based on the number of subpopulations as well as the exposure level, in terms of the number of demands or total time in operation (exposure) for individual populations. Larger amounts of exposure provide better estimates of the reliability parameter for individual populations; higher numbers of sub-populations provide more information about the variability of the reliability parameters, see Figure 3.

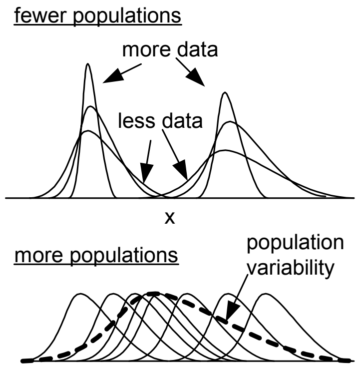

Figure 3 – Characterization of the impact of the number of populations and the amount of data available for individual populations in population variability estimations.

In order to arrive at a good, and ultimately exact, estimate, it is not necessary to have infinite amounts of data for individual populations, as long as data is available for a sufficiently large number of populations. This can be shown for the case where data for each population is available in the form of the number of failures and total time in operation or total number of demands. We use *k* to denote the number of failures, and *t* to denote the total exposure (operating time or number of demands). The likelihood function for *n* populations is

$$\Pr(\langle k_1,t_1\rangle,\ldots,\langle k_n,t_n\rangle \mid \theta) = \prod_{i=1}^{n} \Pr(\langle k_i,t_i\rangle \mid \theta) \tag{19}$$

where

$$\Pr(\langle k_i,t_i\rangle \mid \theta) = \int_{\lambda} \Pr(\langle k_i,t_i\rangle \mid \lambda)\cdot \pi(\lambda \mid \theta)\cdot d\lambda \tag{20}$$

Here, λ represents failure rate or failure probability, depending on the interpretation of *t*.

If we assume that the exposure *t* for all populations is constant, Eq. (19) can be rewritten as

$$\Pr(\langle k_1,t_1\rangle,\ldots,\langle k_n,t_n\rangle \mid \theta) = \prod_{k=0}^{\infty} \Pr(\langle k,t\rangle \mid \theta)^{n_k} \tag{21}$$

where n_k is the number of populations in which k failures were observed. Assuming that the population truly has a population variability governed by $\varphi(x \mid \tilde{\theta})$, the expected values of n_k, $k = 0,\ldots,\infty$, are given by $n \cdot \Pr(\langle k,t \rangle \mid \tilde{\theta})$. As the number of populations n increases, the likelihood function therefore approaches

$$\Pr(\langle k_1,t_1 \rangle,\ldots,\langle k_n,t_n \rangle \mid \theta) \overset{n\to\infty}{=} \prod_{k=0}^{\infty} \Pr(\langle k,t \rangle \mid \theta)^{n\cdot\Pr(\langle k,t \rangle \mid \tilde{\theta})} \tag{22}$$

Regardless of the variability model $\varphi(x|\theta)$, this likelihood function has a maximum for

$$\Pr(\langle k,t \rangle \mid \theta) = \Pr(\langle k,t \rangle \mid \tilde{\theta}),\ k = 0,\ldots,\infty \tag{23}$$

which is the case when $\theta = \tilde{\theta}$, i.e., when the probability of observing k failures to the observed fraction of subpopulations for which k failures were actually observed. As the number of populations n increases, the posterior distribution obtained using the likelihood function in Eq. (22) will converge towards its maximum at $\theta = \tilde{\theta}$. It follows that the uncertainty about the population variability can be completely removed based on limited amounts of information for individual populations, as long as the number of populations is sufficiently high, the chosen variability model is a good representation of the true population variability distribution, and the exposure levels of the subpopulation are constant.

If the actual population variability distribution is not well captured by any of the distributions $\varphi(x \mid \theta)$, the procedure will generally still converge towards a single value $\hat{\theta}$ of the variability model parameter as the number of subpopulations increases.

This behavior is best illustrated by considering the case where the data strongly indicate two prevalent failure rates, λ_1 and λ_2, equally distributed among the populations. Given n populations, the expected likelihood function, based on the expected number of populations for which k failures were observed, equals

$$\Pr(\langle k_1,t_1 \rangle,\ldots,\langle k_n,t_n \rangle \mid \theta) = \prod_{k=0}^{\infty} \Pr(\langle k,t \rangle \mid \theta)^{\frac{n}{2}\cdot\left(\Pr(\langle k,t \rangle \mid \lambda_1)+\Pr(\langle k,t \rangle \mid \lambda_2)\right)} \tag{24}$$

assuming again equal exposure t for all populations. This function, with λ_1 = 1E-3, λ_2 = 1E-2, and t = 1,000, and a Gamma variability distribution model

$$\varphi(\lambda \mid \alpha,\beta) = \frac{\beta^{\alpha} e^{-\beta\cdot\lambda} \cdot \lambda^{\alpha-1}}{\Gamma(\alpha)} \tag{25}$$

is plotted as a function of α and β for n_1 = 10, n_2 = 50, and n_3 = 100 in Figure 4. The plot shows that for increasing n, the significant part of the likelihood function converges towards a single point, despite the fact that the Gamma density is a poor fit for the actual distribution. This type of behavior is generally observed in case of a poor match between model and true variability, including cases where the exposure t_i, $i = 1,\ldots,n$, is not constant. Note however that the likelihood function, with a maximum at α = 0.898 and β = 163.3 indicates a best fit Gamma distribution with a mean (0.055) equal to that of the true mean of the variability distribution. This correspondence is not found between the estimated variance (3.38e-5) and the actual variance (2.03e-5).

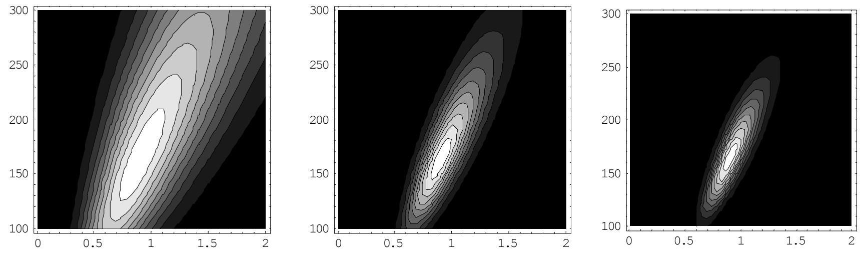

Figure 4 – Contour plots of likelihood function for increasing amounts of data drawn from a truly bimodal variability distribution.

The behavior of the procedure as a function of the amount of data obtained for each population is characterized in Table 2. The table presents the properties of the estimated variability distribution obtained for nine datasets. Each set contains data for 200 populations, the failure rates of which were generated based on a Gamma population variability distribution with parameters $\alpha = 2$ and $\beta = 100$. For cases A.1, A.2, and A.3, the exposure of the individual populations was systematically varied such that populations with lower failure rates had a larger exposure. In cases B.1 and B.2, the exposure for each population was constant. Finally, in cases C.1 and C.2, the exposure was random (uniformly distributed). In each case, the number of failures was randomly sampled from the Poisson distribution.

The Table 2 lists the estimated value for each case, along with the theoretical values ("Actual" column). The table indicates that in case A.1, a bias exists towards lower failure rates levels, where the data had systematically larger exposure levels. This bias disappears however when the overall degree of exposure increases (cases A.2 and A.3). The bias is not present in cases with constant exposure (B.1 and B.2) or random exposure (C.1 and C.2). In case B.1 however, the variance of the variability distribution is underestimated. The conclusion with respect to the impact of population sizes on the estimate is that populations with larger exposure levels carry a larger weight in the estimate.

Table 2 – Estimated population densities under various data conditions.

	Actual	Case A.1	Case A.2	Case A.3	Case B.1	Case B.2	Case C.1	Case C.2
t_i	-	$2 / \lambda$	$20 / \lambda$	$200 / \lambda$	100	500	250-750	2500-7500
Mean	2.00E-02	1.45E-02	1.92E-02	1.96E-02	1.85E-02	2.02E-02	2.02E-02	2.00E-02
Variance	2.00E-04	1.25E-04	1.93E-04	1.85E-04	1.44E-04	1.97E-04	1.96E-04	1.87E-04
1st	1.49E-03	7.22E-04	1.29E-03	1.55E-03	1.71E-03	1.55E-03	1.59E-03	1.66E-03
5th	3.55E-03	2.07E-03	3.22E-03	3.65E-03	3.97E-03	3.70E-03	3.77E-03	3.84E-03
10th	5.32E-03	3.30E-03	4.90E-03	5.41E-03	5.76E-03	5.51E-03	5.60E-03	5.65E-03
50th	1.68E-02	1.18E-02	1.60E-02	1.66E-02	1.61E-02	1.70E-02	1.71E-02	1.70E-02
90th	3.89E-02	2.92E-02	3.77E-02	3.78E-02	3.43E-02	3.89E-02	3.89E-02	3.82E-02
95th	4.74E-02	3.63E-02	4.62E-02	4.60E-02	4.15E-02	4.74E-02	4.74E-02	4.64E-02
99th	6.64E-02	5.28E-02	6.54E-02	6.43E-02	5.80E-02	6.64E-02	6.64E-02	6.47E-02

7. Example: Estimating the Failure Probability from Engineering Judgments

This section illustrates and discusses the methodology through an example. In particular, the assessment of the variability of a reliability measure is analyzed when the available evidence is in the form of engineering judgments. Indeed, let us suppose that we are interested in assessing the population variability of the failure probability on demand, *p*, of a certain type of system. Following the reasoning presented in section 3, it is considered that the variability of the probability of failure is given by a Lognormal distribution. With the aim of having a reference to validate the behavior of the model, expert-based evidence was generated from a Lognormal distribution, Eq. (11), with median equals to 1×10^{-4} and $\sigma = 1.4$. In the context of expert opinions, it is a common practice to express the analyst's confidence on the i^{th} expert, σ_i, in terms of the error factor $EF_i = e^{1.645\sigma_i}$ (see Cooke, 1991). Therefore, the data used for this example consists of a series of expert estimates of the system's failure probability and associated error factors, as shown in Table 3. This data set was analyzed using the Lognormal-Lognormal model.

Table 3 – Engineering judgments used in the example.

Expert	Estimate	EF	Expert	Estimate	EF
1	4.6E-05	10	14	1.0E-04	10
2	5.7E-05	10	15	1.1E-04	10
3	6.4E-05	10	16	1.1E-04	10
4	6.8E-05	10	17	1.1E-04	10
5	7.3E-05	10	18	1.2E-04	10
6	7.6E-05	10	19	1.2E-04	10
7	8.0E-05	10	20	1.3E-04	10
8	8.3E-05	10	21	1.3E-04	10
9	8.6E-05	10	22	1.4E-04	10
10	8.9E-05	10	23	1.5E-04	10
11	9.3E-05	10	24	1.6E-04	10
12	9.6E-05	10	25	1.9E-04	10
13	9.9E-05	10	26	2.2E-04	10

The first step in the analysis is the estimation of the joint probability density function of the population variability parameters. In this example, this distribution $\pi(\mu,\sigma \mid E)$ is obtained by solving Eq. (1) with the likelihood function given by Eq. (12), where E is the available evidence in Table 3. The resulting $\pi(\mu,\sigma \mid E)$ is shown in

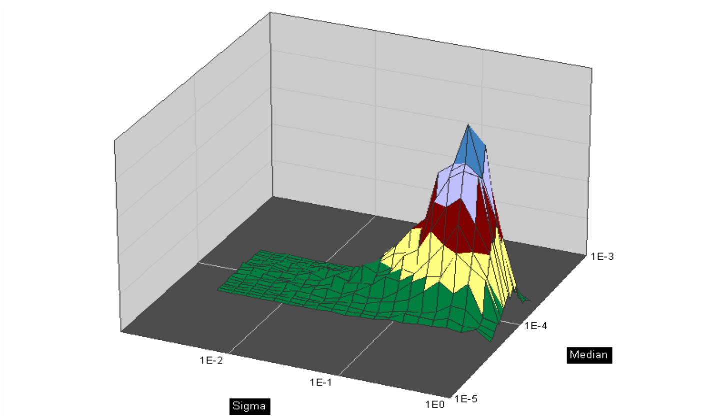

Figure 5 – Posterior distribution of the failure probability variability parameters based on engineering judgments.

Next, we proceed to estimate the population variability distribution of the failure probability. With the posterior joint density of the variability distribution parameters, $\pi(\mu,\sigma \mid E)$, and the Lognormal variability distribution $\varphi(p \mid \mu,\sigma)$, the expected population variability distribution for the failure probability is estimated by solving Eq. (13). The resulting estimated cumulative population variability density function is shown in Figure 6, along with the theoretical distribution from which the data set was generated. In order to have a measure of the uncertainty surrounding the estimated failure probability variability distribution, Figure 6 also shows the corresponding 5% and 95% uncertainty bounds. Figure 6 shows that the estimated expected variability distribution is a reasonable approximation of the true distribution. The deviations at lower and higher values of the failure probability as well as the wide uncertainty bounds are a result of the limited amount of evidence used in the variability analysis and from the high value of the error factor assigned to the engineering judgments, reflecting the analyst low confidence level in the experts.

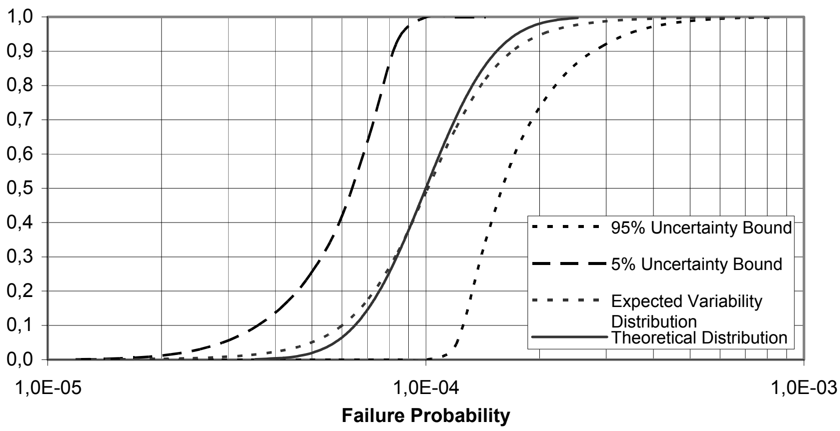

Figure 6 – Cumulative population variability density of the failure probability based on engineering judgments.

8. Example Application: Assessment of the Failure Rate Variability Distribution of Motor-Operated Valves

In this section we discuss the estimation of the failure rate variability distribution of motor-operated valves (MOV) based on real data. Table 4 shows the actual run-time data gathered for different motor-operated valves from eight plants. The data is for the failure mode 'transfer open/closed during operation'.

Table 4 – Run time data used in estimating the failure rate variability distribution for motor-operated valves.

Data Source	Component	Failures	Operating Time (hours)
Plant A	Manual and MOV	0	1.10×10^7
Plant B	MOV (transfer open/leakage)	0	6.95×10^5
Plant C	Manual and MOV	0	9.12×10^2
Plant D	Manual (transfer open/leakage)	0	6.00×10^6
Plant E	Manual and MOV	0	3.70×10^6
Plant F	Manual (transfer open/leakage)	0	2.90×10^2
Plant G	MOV (transfer open/leakage)	0	1.89×10^6
Plant H	MOV	1	3.10×10^7

The objective is to estimate the population variability distribution of the failure rate for motor-operated valves given the available evidence in the form of failures and operating time from MOV operating in different plants and likely under different operating and maintenance policies. Therefore, it is not realistic to assume that the samples are homogeneous with regards to the failure rate. In other words, we cannot consider that the different motor-operated valves have the same failure rate. Indeed, the failure rate is different from MOV to MOV observed in operation, and this variability is represented by the population variability distribution of the failure rate.

Given the type of available evidence, we will employ the Lognormal-Poisson model, i.e., the population variability is assumed to follow a Lognormal distribution and the likelihood of observing k_i failures in T_i is provided by a Poisson distribution. Solving the Lognormal-Poisson model given the available evidence in Table 4, the resulting cumulative density function and probability density function of the expected failure rate variability are shown in Figure 7 and Figure 8, respectively. Also plotted in Figure 7 are the 5% and 95% percentile curves.

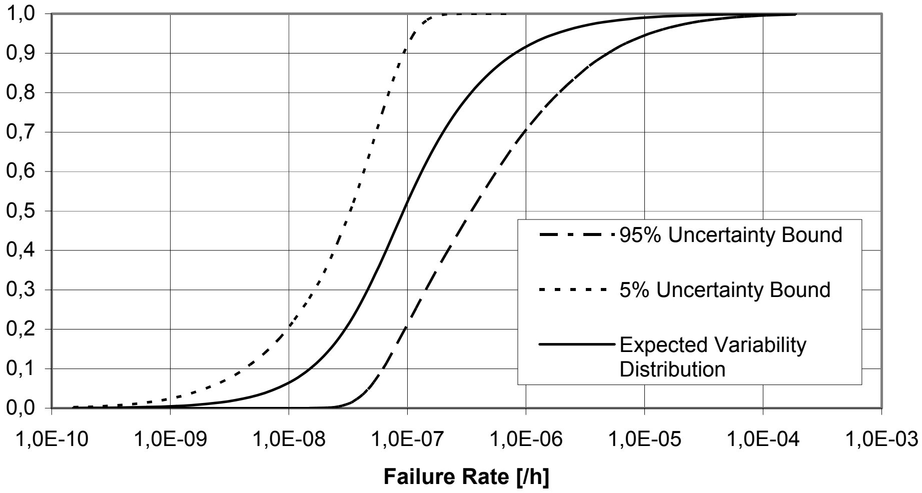

Figure 7 – Expected failure rate variability distribution and 5th and 9th uncertainty bounds for motor-operated valves.

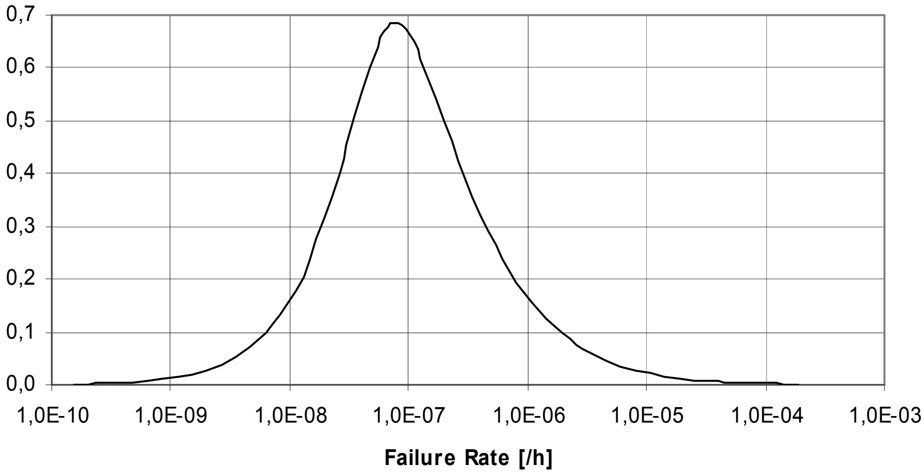

Figure 8 – Probability density function of the expected failure rate variability for motor-operated valves.

The Table 5 shows the mean, 5^{th}, 50^{th}, and 95^{th} percentile curves of the failure rate population variability. Also shown in Table 5 are the several percentiles for each of these curves.

Table 5 – Mean and percentile curves of the failure rate variability distribution for motor-operated valves.

Value	Mean Curve	5th Curve	50th Curve	95th Curve
Mean	1.25E-06			
Variance	0.004001			
1st	1.93E-09	3.10E-08	4.49E-09	4.69E-10
5th	7.94E-09	4.73E-08	1.13E-08	2.00E-09
10th	1.51E-08	6.22E-08	1.84E-08	4.24E-09
50th	9.25E-08	3.48E-07	9.95E-08	3.30E-08
90th	8.12E-07	4.92E-06	5.58E-07	9.22E-08
95th	1.80E-06	1.12E-05	9.09E-07	1.14E-07
99th	9.88E-06	5.24E-05	2.31E-06	1.60E-07

Note that the resulting expected failure rate variability distribution has a mean of 1.25×10^{-6}/h with 5th and 95th percentiles of 7.94×10^{-9}/h and 1.80×10^{-6}/h, respectively. This wide uncertainty bounds, also shown in Table 5, are due to not only the limited amount of data but also the scarce number of failure: only one failure was observed in a total operating time of 5.43×10^{7} hours. The uncertainty bounds tend to get narrower as additional operating experience is gained, for instance, by increasing the number of populations (see discussion in section 7). Another possibility to reduce the level of uncertainty in the failure rate assessment is the combined used of run-time data and expert-based evidence.

9. Concluding Remarks

In this paper we have discussed a Bayesian approach to assess the variability of reliability measures based on partially relevant data. The so-called population variability analysis has a very strong appeal from a practical point of view as it allows for the quantification of reliability metrics such as an item's failure rate or probability of failure based on evidence obtained from not only equivalent or similar items but also under only similar operating conditions and maintenance policies. The mathematical framework was then presented and discussed where emphasis was given to the likelihood construction process for various types of information: exposure-based data and expert-based evidence.

In this context, the modeling of mixed evidence was introduced. Mixed evidence modeling concerns the incorporation of exposure data and engineering judgments into the body of knowledge about some reliability measure of interest. It allows for both the coupled and decoupled modeling of exposure and expert information sources, where the former corresponds to cases where the expert provides estimates about a reliability measure of an item for which exposure evidence is also available. The latter situation occurs when an expert's opinion is about an item different from the one for which exposure evidence is available. This class of mixed models is topic of current research and a detailed discussion can be found in Droguett, Groen & Mosleh (2004).

The utilization of a population variability distribution as a prior in a two-stage variability analysis was then presented. It was demonstrated by considering the probabilistic dependencies involved in the Bayesian two-stage updating procedure, that both stages can be integrated into a single problem.

The behavior of the population variability modeling under different data conditions was analyzed. Provided that available evidence can be characterized based on the number of sub-populations and the exposure level for individual populations, it was shown that to arrive at a good estimate it is not necessary to have infinite amounts of data for individual populations as long as data is available for a sufficiently large number of populations. This was then illustrated by means of an example based on simulated data from a Gamma population variability distribution.

The Bayesian variability analysis approach was then illustrated by means of two examples. In the first of them, the probability of failure variability distribution of a system was assessed when only expert opinions were available. To work as a validation case, the data set was generated from a Lognormal distribution. From the results, it was observed that the estimated expected variability distribution is a reasonable approximation to the theoretical distribution.

In the second example the approach was employed in the failure rate variability analysis of motor-operated valves based on real data. Primarily due to the limited number of observed failures (only one failure in a total of 5.43×10^7 operating hours for eight units), the results showed a somewhat wide uncertainty bounds (5^{th} and 95^{th} percentile curves). However, as illustrated in both examples, the approach allows for the explicit quantification of the uncertainty in a population variability analysis, which is a valuable aid in the process of decision making under uncertainty.

References

(1) Bernardo, J.M. & Smith, A.F.M. (1994). *Bayesian Theory*. John Wiley & Sons, New York.

(2) Cooke, R.M. (1991). *Experts in Uncertainty*. Oxford University Press, New York.

(3) Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**(4), 327-335.

(4) Droguett, E.L.; Groen, F. & Mosleh, A. (2004). The combined use of data and expert estimates in population variability analysis. *Reliability Engineering & System Safety*, **83**, 311-321.

(5) Droguett, E.L. & Groen, F. (2004). Assessment of the variability of reliability measures. XXXVI SBPO – Simpósio Brasileiro de Pesquisa Operacional. São João del Rei, MG.

(6) Frohner, F. (1985). Analytic Bayesian solution of the two-stage Poisson-type problem in probabilistic risk analysis. *Risk Analysis*, **5**, 217-234.

(7) Gilks, W.R.; Richardson, W.R. & Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

(8) Hora, S.C. & Iman, R.L. (1990). Bayesian modeling of initiating event frequencies at nuclear power plants. *Risk Analysis*, **10**, 102-109.

(9) Kaplan, S. (1985). The two-stage Poisson-type problem in probabilistic risk analysis. *Risk Analysis*, **5**, 227-230.

(10) Kaplan, S. (1983). On a 'two-stage' Bayesian procedure for determining failure rates from experimental data. *IEEE Transactions on Power Apparatus and Systems*. **PAS-102**, 195-199.

(11) Mosleh, A. & Apostolakis, G. (1986). The assessment of probability distributions from expert opinions with an application to seismic fragility curves. *Risk Analysis*, **6**, 447-461.

(12) Mosleh, A. (1992). Bayesian modeling of expert-to-expert variability and dependence in estimating rare event frequencies. *Reliability Engineering & System Safety*, **38**.

(13) Pearl, J. (1998). *Probabilistic Reasoning in Intelligent Systems*. Second edition. Morgan Kaufmann Publishers, San Francisco.

(14) Pörn, K. (1996). The two-stage Bayesian method used for the t-book application. *Reliability Engineering & System Safety*, **51**, 169-179.

(15) Siu, N.O. & Kelly, D.L. (1998). Bayesian parameter estimation in probabilistic risk assessment. *Reliability Engineering & System Safety*, **62**, 89-116.

(16) Vaurio, J. (1987). On analytic empirical Bayes estimation of failure rates, plus discussion. *Risk Analysis*, **7**, 329-338.