

Bayesian updating and model class selection with Subset Simulation

Siu-Kui Au¹, Francisco Alejandro DiazDelaO² and Ikumasa Yoshida³

^{1,2}Institute for Risk and Uncertainty, School of Engineering, University of Liverpool, UK

³Department of Urban and Civil Engineering, Tokyo City University, Japan

Abstract

Identifying the parameters of a model and rating competitive models based on measured data has been among the most important but challenging topics in modern science and engineering, with great potential of application in structural system identification, updating and development of high fidelity models. These problems in principle can be tackled using a Bayesian probabilistic approach, where the parameters to be identified are treated as uncertain and their inference information are given in terms of their ‘posterior’ (i.e., given data) probability distribution. For complex models encountered in applications, efficient computational tools robust to the number of uncertain parameters in the problem are required for computing the ‘posterior statistics’, which can generally be formulated as a multi-dimensional integral over the space of the uncertain parameters. Subset Simulation (SuS) has been developed for solving reliability problems involving complex systems and it is found to be robust to the number of uncertain parameters. An analogy has been recently established between a Bayesian updating problem and a reliability problem, which opens up the possibility of efficient solution by SuS. The formulation, called BUS (Bayesian Updating with Structural reliability methods), is based on conventional rejection principle. Its theoretical correctness and efficiency requires the prudent choice of a multiplier, which has remained an open question. Motivated by the choice of the multiplier and its philosophical role, this paper presents a study of BUS. The work leads to a revised formulation that resolves the issues regarding the multiplier so that SuS can be

¹ Member, ASCE. Professor, Chair of Uncertainty, Reliability and Risk. School of Engineering, University of Liverpool. Harrison Hughes Building, Brownlow Hill, L69 3GH, Liverpool, UK. E-mail: siukuiau@liverpool.ac.uk. Office phone: +44 (0)151 794 5217.

² Assistant Professor. School of Engineering, University of Liverpool. Harrison Hughes Building, Brownlow Hill, L69 3GH, Liverpool, UK. E-mail: f.a.diazdelao@liverpool.ac.uk. Office phone: +44 (0)151 794 5235.

³ Professor. Department of Urban and Civil Engineering, Tokyo City University. 1-28-1 Tamazutsumi Setagaya-ku, Tokyo 158-8557, Japan. E-mail: iyoshida@tcu.ac.jp. Office phone: +81-3-5707-1156.

implemented without knowing the multiplier. Examples are presented to illustrate the theory and applications.

Keywords:

Bayesian inference, BUS, Subset Simulation, Markov Chain Monte Carlo, model class selection, model updating

Introduction

Making inference about the parameters of a mathematical model based on observed measurements of the real system is one of the most important problems in modern science and engineering. The Bayesian approach provides a fundamental means to do this in the context of probability logic (Malakoff 1999, Cox 1961, Jaynes 2003), where the parameters are viewed as uncertain variables and the inference results are cast in terms of their probability distribution after incorporating information from the observed data. In engineering dynamics, for example, vibration data from a structure is collected from sensors and used for identifying the modal properties (e.g. natural frequencies, damping ratios, mode shapes) and structural model properties (e.g. stiffness, mass) (Hudson 1977, Ewins 2000). This has been formulated in a Bayesian context (Beck & Katafygiotis 1998, Beck 2010), which resolved a number of philosophically challenging issues of the inverse problem, such as the treatment of multiple sets of parameters giving the same model fit to the data, an issue known as ‘identifiability’.

Let $\Theta \in R^n$ be a set of parameters of a model \mathcal{M} , based on which a probabilistic prediction of the data D can be formulated through the ‘likelihood function’ $P(D|\Theta, \mathcal{M})$. Clearly, the probability distribution of Θ depends on the available information. Based only on knowledge in the context of \mathcal{M} , the distribution is described by the ‘prior distribution’ $p(\Theta|\mathcal{M})$. When data about the system is available, it can be used to update the distribution. Using Bayes’ Theorem, the ‘posterior distribution’ that incorporates the data information in the context of \mathcal{M} is given by

$$p(\Theta|D, \mathcal{M}) = P(D|\mathcal{M})^{-1} P(D|\Theta, \mathcal{M}) p(\Theta|\mathcal{M}) \quad (1)$$

where

$$P(D|\mathcal{M}) = \int P(D|\boldsymbol{\theta}, \mathcal{M}) p(\boldsymbol{\theta}|\mathcal{M}) d\boldsymbol{\theta} \quad (2)$$

is a normalizing constant. Future predictions of a response quantity of interest $r(\boldsymbol{\theta})$ (say) can be updated by incorporating data information, through the posterior expectation (Papadimitriou et al. 2001):

$$E[r(\boldsymbol{\theta})|D, \mathcal{M}] = \int r(\boldsymbol{\theta}) p(\boldsymbol{\theta}|D, \mathcal{M}) d\boldsymbol{\theta} \quad (3)$$

As far as the posterior distribution of $\boldsymbol{\theta}$ for a given model \mathcal{M} is concerned, the constant in (2) is immaterial because it does not change the distribution. However, it is the primary quantity of study in Bayesian model class selection problems where competing models are compared based on the value of $P(\mathcal{M})P(D|\mathcal{M})$ (Carlin & Chib 1995, Chen et al. 2000, Beck & Yuen 2004). In that context, $P(D|\mathcal{M})$ is often called the ‘evidence’ (the higher the better).

Capturing efficiently essential information about the posterior distribution, i.e., posterior statistics, and calculating the posterior expectation is a highly non-trivial problem, primarily resulting from the complexity of the likelihood function. In many applications, the likelihood function is only implicitly known, i.e., its value can be calculated point-wise but its dependence on the model parameters is mathematically intractable. This renders analytical solutions infeasible and conventional numerical techniques inapplicable. In this case, Markov Chain Monte Carlo (MCMC) (Metropolis et al. 1953, Hastings 1970, Robert & Casella 2004, Fishman 1996) is found to provide a powerful computational tool. MCMC allows the samples of an arbitrarily given distribution to be efficiently generated as the samples of a specially designed Markov chain. In MCMC, candidate samples are generated by a ‘proposal distribution’ (chosen by the analyst) and they are adaptively accepted based on ratios of the target distribution value at the candidate and the current sample. If the candidate is rejected, the current sample is taken as the next sample. In the context of the Bayesian updating problem, samples following the posterior distribution are generated using MCMC and they are used for estimating posterior statistics by means of statistical averaging.

While MCMC in principle provides a powerful solution for Bayesian computation, difficulties are encountered in applications, motivating different variants of the algorithm. For example, in

problems with a large amount of data, the posterior distribution takes on significant values only in a small region of the parameter space, whose size generally shrinks in an inverse square root law with the data size. Depending on sufficiency or relevance of the data for the model parameters, the regions of significant probability content can be around a set of isolated points (globally or locally identifiable) or a lower dimensional manifold (unidentifiable) with non-trivial geometry (Katafygiotis & Beck 1998, Katafygiotis & Lam 2002). To the least extent this causes efficiency problems, making the choice of the proposal distribution difficult and leading to high rejection rate of candidates and hence poor efficiency. When the issue is not managed properly, significant bias can result in the statistical estimation based on the samples. Strategies similar to simulated annealing have been proposed to convert the original difficult updating problem effectively into a sequence of more manageable problems with less data, thereby allowing the samples to adapt gradually (Beck & Au 2002, Cheung & Beck 2009, Ching & Chen 2007). Another issue is ‘dimension sustainability’, i.e., whether the algorithm remains applicable when the number of variables (i.e., dimension) of the problem increases. This imposes restrictions on the design of MCMC algorithms so that quantities such as the ratio of likelihood functions involved in the simulation process do not ‘degenerate’ as the dimension of the problem increases.

Application robustness and dimension sustainability are well-recognized in the engineering reliability method literature (Au & Beck 2003, Schueller et al. 2004, Katafygiotis & Zuev 2008). In this area, the general objective is to determine the failure probability that a scalar response of interest exceeds a specified threshold value, or equivalently to determine its complementary cumulative distribution function (CCDF) near the upper tail (i.e., large thresholds). Subset Simulation (SuS) (Au & Beck 2001, Au & Wang 2014) has been developed as an advanced Monte Carlo strategy that is efficient for small failure probabilities (rare events) but still retain a reasonable robustness similar to the Direct Monte Carlo method. In SuS, samples conditional on a sequence of intermediate failure events are generated by MCMC and they gradually populate towards the target failure region. These ‘conditional samples’ provide information for estimating the whole CCDF of the response quantity of interest. SuS typically does not make use of any problem-specific information, treating the input-output relationship between the response and the

uncertain parameters as a ‘black box’. Based on an independent-component MCMC strategy, it is applicable for an arbitrary (potentially infinite) number of uncertain variables in the problem.

By establishing an analogy with the reliability problem that SuS is originally designed to solve, it is possible to adapt SuS to provide an efficient solution for another class of problems. For example, by considering an ‘augmented reliability problem’ where deterministic design parameters are artificially considered as uncertain, SuS has been applied to investigate the sensitivity of the failure probability with respect to the design parameters and their optimal choice without repeated simulation runs (Au 2005, Ching & Hsieh 2007, Song et al. 2009, Taflanidis & Beck 2009). Another example can be found in constrained optimization problems, where an analogy was established between rare failure events in reliability problems and extreme events in optimization problems, allowing SuS to be applied to solving complex problems with nonlinear objective functions and potentially a large number of inequality constraints and optimization variables (Li & Au 2010, Wang et al. 2011).

In view of the application robustness and dimension sustainability, it would be attractive to adapt SuS for Bayesian computations. This is not trivial since the problem contexts are different. One major difference is that in the reliability problem the uncertain parameters follow standard classes of distributions (e.g., Gaussian, exponential) specified by the analyst; while in the Bayesian updating problem the uncertain parameters follow the posterior distribution, which generally does not belong to any standard distribution because the likelihood function is problem-dependent.

Recent developments have shown promise for adapting SuS to Bayesian updating problems. In the context of Approximate Bayesian Computation (ABC), Chiachio et al. (2014) built an analogy with the reliability problem so that the posterior samples in the Bayesian updating problem can be obtained as the conditional samples in SuS at the highest simulation level determined by a tolerance parameter that gradually diminishes. The latter controls the approximation of the likelihood function through a proximity model (a feature of ABC) between the measured and simulated data for a given value of model parameter.

Along another line of thought, Straub and Papaioannou (2014) recently provided a formulation called BUS (Bayesian Updating using Structural reliability methods) that opens up the possibility of Bayesian updating using Subset Simulation. It combined an earlier idea (Straub 2011) with the standard rejection principle to establish an analogy between a Bayesian updating problem and a reliability problem, or more correctly a ‘probabilistic failure analysis’ problem (Au & Beck 2003, Au 2004, Au & Wang 2014). Through the analogy, the samples following the posterior distribution in the Bayesian updating problem can be obtained as the conditional samples in the reliability problem. Unlike ABC, the formulation is exact as it respects fully the original likelihood function; and in this sense it is more fundamental. One outstanding problem, however, is the choice of the ‘likelihood multiplier’, or ‘multiplier’ in short, in the context of rejection principle. To guarantee the theoretical correctness of the analogy, it must be less than the reciprocal of the maximum value of the likelihood function, which is generally unknown especially before the problem is solved. Some suggestions have been given in Straub & Papaioannou (2014) based on inspection of the likelihood function. An adaptive choice was suggested based empirically on the generated samples (Betz et al. 2014). It is more robust to applications as it does not require prior input from the analyst. However, the problem remains open, since Betz et al. conjecture in their paper that the samples produced by their method are from the posterior distribution. They point out that more theoretical analysis of their results is needed.

This work is motivated by the choice of the multiplier and its mathematical and philosophical role in the BUS formulation. A rigorous mathematical study is carried out to provide understanding of the multiplier, which leads to a revised formulation, allowing SuS to be implemented independent of the choice of the multiplier and convergence of results to be checked formally. Essentially, by defining the failure event in the BUS formulation, we show that SuS can in fact be implemented *without the multiplier* and the samples beyond a certain simulation level all have the same target posterior distribution.

This paper is organized as follows. We first give an overview of Subset Simulation and the original BUS formulation. The mathematical role of the multiplier and its bias effect arising from inappropriate choice are then investigated. A revised formulation is then proposed and associated

theoretical issues are investigated, followed by a discussion on the application of SuS under the revised formulation. Examples are presented to explain the theory and illustrate its applications.

Subset Simulation

We first briefly introduce Subset Simulation (SuS) to facilitate understanding its application in the context of Bayesian model updating and model class selection later. SuS is an advanced Monte Carlo method for reliability and failure analysis of complex systems, especially for rare events. It is based on the idea that a small failure probability can be expressed as a product of larger conditional failure probabilities, effectively converting a rare simulation problem into a series of more frequent ones.

Reliability and failure analysis problem

Despite the variety of failure events in applications, they can often be formulated as the exceedance of a critical response over a specified threshold. Let $Y = h(\Theta)$, be a scalar response quantity of interest that depends on the set of uncertain parameters Θ distributed as the parameter probability density function (PDF) $q(\Theta)$. The function h represents the relationship between the uncertain input parameters and the output response. The parameter PDF q is specified by the analyst from standard distributions. Without loss of generality, the uncertain parameters are assumed to be continuous-valued and independent, since discrete-valued variables or dependent variables can be obtained by mapping continuous-valued independent ones.

The primary interest of reliability analysis is to determine the ‘failure probability’ $P(Y > b)$ for a specified threshold value b :

$$P(Y > b) = \int q(\Theta) I(\Theta \in F) d\Theta \quad (4)$$

where

$$F = \{Y > b\} = \{\Theta \in R^n : h(\Theta) > b\} \quad (5)$$

denotes the failure event or the failure region in the parameter space, depending on the context; $I(\cdot)$ is the indicator function, equal to 1 if its argument is true and zero otherwise. Probabilistic failure analysis on the other hand is concerned with what happens when failure occurs, which often involves investigating the expectation of some response quantity $r(\Theta)$ (say) conditional on the failure event, i.e.,

$$E[r(\Theta) | F] = \int r(\Theta) q(\Theta | F) d\Theta \quad (6)$$

where

$$q(\Theta | F) = P_F^{-1} q(\Theta) I(\Theta \in F) \quad (7)$$

is the PDF of Θ conditional on failure.

When the relationship between Y and Θ , i.e., the function h , is complicated, analytical or conventional numerical integration is not feasible for computing $P(Y > b)$ or $E[r(\Theta) | F]$. Advanced computational methods are then required for their efficient determination. SuS offers an efficient solution by generating a sequence of sample populations of Θ conditional on increasingly rare failure events $\{Y > b_i\}$, where $\{b_i : i = 1, 2, \dots\}$ is an increasing sequence of threshold values adaptively determined during the simulation run. These ‘conditional samples’ provide information for estimating the CCDF of Y , i.e., $P(Y > b)$ versus b from the frequent (left tail) to the rare (right tail) regime. When the right tail covers the threshold value associated with the target failure event, the required failure probability can be obtained from the estimate of the CCDF. The conditional samples can also be used for estimating the conditional expectation in probabilistic failure analysis, a feature not shared by conventional variance reduction techniques. As we shall see in the next section, through the analogy between the reliability and Bayesian updating problem, the conditional samples provide the posterior samples required for Bayesian model updating. The failure probability provides the information for estimating the evidence for Bayesian model class selection.

Subset Simulation procedure

A typical SuS algorithm is presented as follows (Au & Beck 2001, Au & Wang 2014). Two parameters should be set before starting a simulation run: 1) the ‘level probability’ $p_0 \in (0,1)$ and 2) the ‘number of samples per level’ N . It is assumed that $p_0 N$ and p_0^{-1} are positive integers. As will be seen shortly, these are respectively equal to the number of chains and the number of samples per chain at a given simulation level. In the reliability literature, a prudent choice is $p_0 = 0.1$. The number of samples N controls the statistical accuracy of results (the higher the better), generally in an inverse square root manner. Common choice ranges from a few hundreds to over a thousand, depending on the required accuracy of the target failure probability estimate.

A simulation run starts with Level 0 (unconditional), where N i.i.d. (independent and identically distributed) samples of Θ are generated from q , i.e., Direct Monte Carlo. The corresponding values of Y are computed and arranged in ascending order, giving an ordered list denoted by $\{b_k^{(0)} : k = 1, \dots, N\}$. The value $b_k^{(0)}$ gives the estimate of b corresponding to the exceedance probability $p_k^{(0)} = P(Y > b)$ where

$$p_k^{(0)} = \frac{N - k}{N} \quad k = 1, \dots, N \quad (8)$$

The next level, i.e., Level 1, is conditional on the intermediate failure event $\{Y > b_1\}$, where b_1 is determined as the $(p_0 N + 1)$ -th largest sample value of Y at Level 0, i.e.,

$$b_1 = b_{N(1-p_0)}^{(0)} \quad (9)$$

By construction, the $p_0 N$ samples of Θ corresponding to $\{b_{N(1-p_0)+j}^{(0)} : j = 1, \dots, p_0 N\}$ are conditional on $\{Y > b_1\}$. These conditional samples are used as ‘seeds’ for generating additional samples conditional on $\{Y > b_1\}$ by means of MCMC. A MCMC chain of p_0^{-1} samples is generated from each seed, giving a total a population of $p_0 N \times p_0^{-1} = N$ samples conditional on $\{Y > b_1\}$ at Level 1.

During MCMC the values of Y of the conditional samples at Level 1 have been calculated. They are arranged in ascending order, giving the ordered list denoted by $\{b_k^{(1)} : k = 1, \dots, N\}$. The value $b_k^{(1)}$ gives the estimate of b corresponding to exceedance probability $p_k^{(1)} = P(Y > b)$ where

$$p_k^{(1)} = p_0 \frac{N-k}{N} \quad k = 1, \dots, N \quad (10)$$

The next level, i.e., Level 2, is conditional on $\{Y > b_2\}$ where b_2 is determined as the $(p_0 N + 1)$ -th largest sample value of Y at Level 1, i.e.,

$$b_2 = b_{N(1-p_0)}^{(1)} \quad (11)$$

The above process of generating additional MCMC samples and moving up simulation levels is repeated until the target threshold level or probability level has been reached. In general, at Level i ($i = 1, 2, \dots$), in the ordered list of sample values of Y denoted by $\{b_k^{(i)} : k = 1, \dots, N\}$, the value $b_k^{(i)}$ gives the estimate of b corresponding to exceedance probability $p_k^{(i)} = P(Y > b)$ where

$$p_k^{(i)} = p_0^i \frac{N-k}{N} \quad k = 1, \dots, N \quad (12)$$

Several features of SuS are worth-mentioning. It is population-based in the sense that the samples at a given level are generated from multiple ($p_0 N$) chains, making it robust to ergodic problems. An independent-component MCMC algorithm is used, which is the key to be sustainable for high dimensional problems (Au & Beck 2001, Schueller et al. 2004, Haario et al. 2005). The conditional samples at each level all have the target conditional distribution and there is no ‘burn-in’ problem commonly discussed in the MCMC literature. This is because the MCMC chains are all started with a seed distributed as the target distribution (conditional on that level), and so they are stationary right from the start.

Variants of the SuS algorithm have been proposed to improve efficiency, e.g., Papadopoulos et al. (2012), Zuev & Katafygiotis (2011), Bourinet et al. (2011). See also the review in Section 5.9 of Au & Wang (2014). The algorithm can even be implemented as a VBA (Visual Basic for Applications) Add-In in a spreadsheet (Au et al. 2010, Wang et al. 2010).

BUS formulation

In this section we review the BUS formulation (Straub & Papaioannou 2014) that builds an analogy between the Bayesian updating problem and a reliability problem, thereby allowing SuS to be applied to the former. For mathematical clarity and to simplify notation, in the Bayesian updating problem we use $q(\boldsymbol{\theta})$ to denote the prior PDF $p(\boldsymbol{\theta} | \mathcal{M})$, $L(\boldsymbol{\theta})$ to denote the likelihood function $p(D | \boldsymbol{\theta}, \mathcal{M})$, P_D to denote the normalizing constant $P(D | \mathcal{M})$, and $p_D(\boldsymbol{\theta})$ to denote the posterior PDF $p(\boldsymbol{\theta} | D, \mathcal{M})$. The same symbol $q(\boldsymbol{\theta})$ is used for the prior PDF in the Bayesian updating problem and the parameter PDF in the reliability problem, as it has the same mathematical property (chosen from standard distributions by the analyst) and role (the distribution to start the SuS run) in both problems. In a Monte Carlo approach the primary target in Bayesian model updating is to generate samples according to the posterior PDF $p_D(\boldsymbol{\theta})$ (rewritten from (1)):

$$p_D(\boldsymbol{\theta}) = P_D^{-1} q(\boldsymbol{\theta}) L(\boldsymbol{\theta}) \quad (13)$$

Rejection principle

The BUS formulation is based on the conventional rejection principle. Let c , called the ‘likelihood multiplier’ in this work, or simply ‘multiplier’, be a scalar constant such that for all $\boldsymbol{\theta}$ the following inequality holds:

$$cL(\boldsymbol{\theta}) \leq 1 \quad (14)$$

Also, assume that i.i.d. samples can be efficiently generated from the prior PDF $q(\boldsymbol{\theta})$. This is a reasonable assumption because the prior PDF is often chosen from a standard class of distributions (e.g., Gaussian, exponential). In the above context, a sample $\boldsymbol{\Theta}$ distributed as the posterior PDF $p_D(\boldsymbol{\theta}) \propto q(\boldsymbol{\theta})L(\boldsymbol{\theta})$ in (13) can be generated from the following straightforward application of the rejection principle:

Step 1. Generate U uniformly distributed on $[0,1]$ and $\boldsymbol{\Theta}$ distributed with the prior PDF $q(\boldsymbol{\theta})$.

Step 2. If $U < cL(\boldsymbol{\Theta})$, return $\boldsymbol{\Theta}$ as the sample. Otherwise go back to Step 1.

The following standard proof shows that the sample Θ returned from the above algorithm is distributed as $p_D(\Theta)$ in (13). First note in Step 1 that Θ is distributed as $q(\Theta)$ and U as $I(0 \leq u \leq 1)$, and so their joint PDF is $q(\Theta)I(0 \leq u \leq 1)$. In Step 2, since (Θ, U) are only returned when $U < cL(\Theta)$, their joint PDF is simply the joint PDF in Step 1 conditional on this event. That is,

$$p_{\Theta, U}(\Theta, u) = P_A^{-1} q(\Theta) I(0 \leq u \leq 1) I[u < cL(\Theta)] \quad (15)$$

where

$$P_A = \int \int q(\Theta) I(0 \leq u \leq 1) I[u < cL(\Theta)] du d\Theta \quad (16)$$

is the probability that the sample is accepted in Step 2, i.e., ‘acceptance probability’.

The marginal PDF of Θ returned by the rejection algorithm can be obtained by integrating out the uncertainty of U :

$$\begin{aligned} p_{\Theta}(\Theta) &= \int_0^1 p_{\Theta, U}(\Theta, u) du \\ &= P_A^{-1} q(\Theta) \int_0^1 I(0 \leq u \leq 1) I[u < cL(\Theta)] du \\ &= P_A^{-1} q(\Theta) cL(\Theta) \\ &\propto p_D(\Theta) \end{aligned} \quad (17)$$

where the third equality follows from the fact that $cL(\Theta) \leq 1$ from (14). This shows that the sample Θ returned by the rejection algorithm indeed follows the posterior PDF $p_D(\Theta)$. Integrating (13) and (17) over the whole space and noting that the integral on the LHS is 1 gives

$$P_A = cP_D \quad (18)$$

Although the above rejection algorithm is theoretically viable, the acceptance probability and hence efficiency is often very low in typical updating problems with a reasonable amount of data. This is because a typical sample drawn from the prior PDF $q(\Theta)$ often has a low likelihood value $L(\Theta)$ when the data is informative about the uncertain parameters, leading to significant change from the prior to the posterior PDF.

Equivalent reliability problem

Recognizing the high rejection rate when the rejection principle is directly applied, BUS transforms the problem into a reliability problem. The premise is that this will allow the existing algorithms developed in the reliability method literature to be applied to Bayesian updating problems, especially those are that capable of generating samples from the frequent (safe) region to the rare (failure) region, such as SuS. The reliability problem analogy of the Bayesian updating problem is constructed as follows. Consider a reliability problem with uncertain parameters (Θ, U) having the joint PDF $q(\Theta)I(0 \leq u \leq 1)$, where the ‘failure event’ is defined as

$$F = \{U < cL(\Theta)\} \quad (19)$$

Suppose that by some means (e.g., SuS) we can obtain a ‘failure sample’ distributed as $q(\Theta)I(0 \leq u \leq 1)$ and conditional on the failure event F . The PDF of the failure sample, denoted by (Θ', U') , is given by

$$p_{\Theta', U'}(\Theta, u) = P_F^{-1} q(\Theta) I(0 \leq u \leq 1) I[u < cL(\Theta)] \quad (20)$$

where

$$P_F = \int \int q(\Theta) I(0 \leq u \leq 1) I[u < cL(\Theta)] du d\Theta \quad (21)$$

is the ‘failure probability’ of the reliability problem. Except for the normalizing constant (which does not affect the distribution), the expression in (20) is the same as that in (15). Following exactly the same steps in (17), the marginal PDF of the failure sample is shown to be equal to the posterior PDF:

$$\begin{aligned} p_{\Theta'}(\Theta) &= \int_0^1 p_{\Theta', U'}(\Theta, u) du \\ &= P_F^{-1} q(\Theta) \int_0^1 I(0 \leq u \leq 1) I[u < cL(\Theta)] du \\ &= P_F^{-1} q(\Theta) cL(\Theta) \\ &\propto p_D(\Theta) \end{aligned} \quad (22)$$

Integrating (13) and (22) over the whole space and noting that the integral on the LHS is 1 gives

$$P_F = P_A \quad (23)$$

Posterior samples by Subset Simulation

The theory described in the last subsection is the essence of the formulation proposed in Straub & Papaioannou (2014), called BUS (Bayesian Updating using Structural reliability methods). It converts the generation of samples from the posterior PDF in Bayesian updating problems into the generation of failure samples in a reliability problem, or more specifically, a ‘probabilistic failure analysis’ problem (Au & Beck 2003, Au 2004). The significance of the BUS analogy is that it generalizes the idea of rejection principle so that the posterior samples need not come from a standard rejection algorithm. Any algorithm that can generate the failure samples in the equivalent reliability problem can be used. SuS is a natural choice because it not only provides the estimate for failure probability but also the failure samples, which are not provided by conventional reliability methods.

In Straub & Papaioannou (2014), the driving response variable was defined as (see a quick remark after this paragraph)

$$Y = cL(\Theta) - U \quad (24)$$

so that the failure event corresponds to

$$F = \{Y > 0\} \quad (25)$$

Populations of failure samples conditional on the intermediate failure events $F_i = \{Y > b_i\}$ for adaptively increasing b_i ($i = 1, 2, \dots$) are then generated until they pass the target failure event $F = \{Y > 0\}$, from which the samples conditional on F are collected as the posterior samples.

As a remark, in Straub & Papaioannou (2014) the driving response variable was in fact defined in a reverse manner as $Y = U - cL(\Theta)$. The presentation in (24) is adopted here so that it is consistent with the conventional SuS literature, where the intermediate threshold level increases rather than decreases as the simulation level ascends.

Likelihood multiplier

One issue of concern in the BUS formulation is the choice of the multiplier c satisfying the inequality in (14), which is not always trivial. Some suggestions were given, e.g., by inspecting

the mathematical structure of the likelihood function (Straub & Papaioannou 2014); or by adaptively using empirical the information from the generated samples (Betz et al. 2014). The latter is more robust as it does not require preliminary analysis, but, as stated by the authors, in order to guarantee that it satisfies the inequality, more theoretical analysis is needed.. In this section we rigourously investigate the role of the multiplier and its effect on the results if it is not properly chosen. The investigation leads to a revised formulation to be proposed in the next section.

In the context of BUS, the multiplier needs to be chosen before starting a SuS run as it affects the definition of the driving variable $Y = cL(\boldsymbol{\theta}) - U$ in (24). Clearly, the multiplier affects the distribution of the driving variable as well as the generated samples. Recall that only those samples conditional on $Y = cL(\boldsymbol{\theta}) - U > 0$ are collected as the posterior samples. The larger the value of c the more efficient the SuS run, because this will increase Y and the failure probability $P(Y > 0)$, thereby reducing the number of simulation levels required to reach the target failure event.

Largest admissible value

From the inequality in (14), the choice of the multiplier is governed by the region in the parameter space of $\boldsymbol{\theta}$ where the value of $L(\boldsymbol{\theta})$ is large. The largest admissible value of c is given by

$$c_{\max} = [\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta})]^{-1} \quad (26)$$

This result is well-known in the rejection principle literature. Clearly, this value is not known before computation. While using a value smaller than c_{\max} will be less efficient but still give the correct distribution in the samples, using a value larger than c_{\max} will lead to bias in the distribution of the samples. In some problems it is possible to investigate the mathematical structure of $L(\boldsymbol{\theta})$ and derive inequalities to propose a choice of c that guarantees $cL(\boldsymbol{\theta}) \leq 1$. In such cases, it is computationally beneficial to use that value. However, in general it is difficult by numerical means to have choice of c that guarantees the inequality.

Truncation effect of inadmissible multiplier

When an inadmissible (too large) value of the multiplier is used, the resulting distribution of the failure samples will be truncated, leading to bias in the posterior statistical estimates based on them. To see this, note that the inequality (14) was used in establishing the third equality in (22) (and (17) for the rejection algorithm). Suppose this inequality is violated, say, within some region B :

$$B = \{\boldsymbol{\theta} \in R^n : cL(\boldsymbol{\theta}) > 1\} \quad (27)$$

Then for any $\boldsymbol{\theta} \in B$, $I[u < cL(\boldsymbol{\theta})] = 1$ for $0 \leq u \leq 1$ and so (22) implies

$$p_{\boldsymbol{\Theta}'}(\boldsymbol{\theta}) = P_F^{-1} q(\boldsymbol{\theta}) \int_0^1 I[u < cL(\boldsymbol{\theta})] du = P_F^{-1} q(\boldsymbol{\theta}) \quad \boldsymbol{\theta} \in B \quad (28)$$

For those $\boldsymbol{\theta}$ not in B , the inequality is satisfied and the PDF value $p_{\boldsymbol{\Theta}'}(\boldsymbol{\theta})$ remains to be the correct posterior PDF $p_D(\boldsymbol{\theta})$ as in (22):

$$p_{\boldsymbol{\Theta}'}(\boldsymbol{\theta}) = P_F^{-1} q(\boldsymbol{\theta}) cL(\boldsymbol{\theta}) \propto p_D(\boldsymbol{\theta}) \quad \boldsymbol{\theta} \notin B \quad (29)$$

Thus, an inadmissible (too large) value of c introduces bias in the problem by truncating the posterior PDF to be the prior PDF in the region of $\boldsymbol{\theta}$ where the inequality is violated. Intuitively, in the context of rejection principle, if the multiplier is not small enough, the samples drawn from the prior PDF are accepted (incorrectly) ‘too often’, rendering their distribution closer to the prior PDF than they should be.

The truncation effect is illustrated in Figure 1, where the shaded interval denotes the region B . The prior PDF $q(\boldsymbol{\theta})$ has been taken to be constant (for simplicity in illustration) and so $p_D(\boldsymbol{\theta}) \propto cL(\boldsymbol{\theta})$. Instead of the target posterior PDF, the resulting distribution of the sample takes the shape of the center line. Within the region B it is truncated to the shape of $q(\boldsymbol{\theta})$.

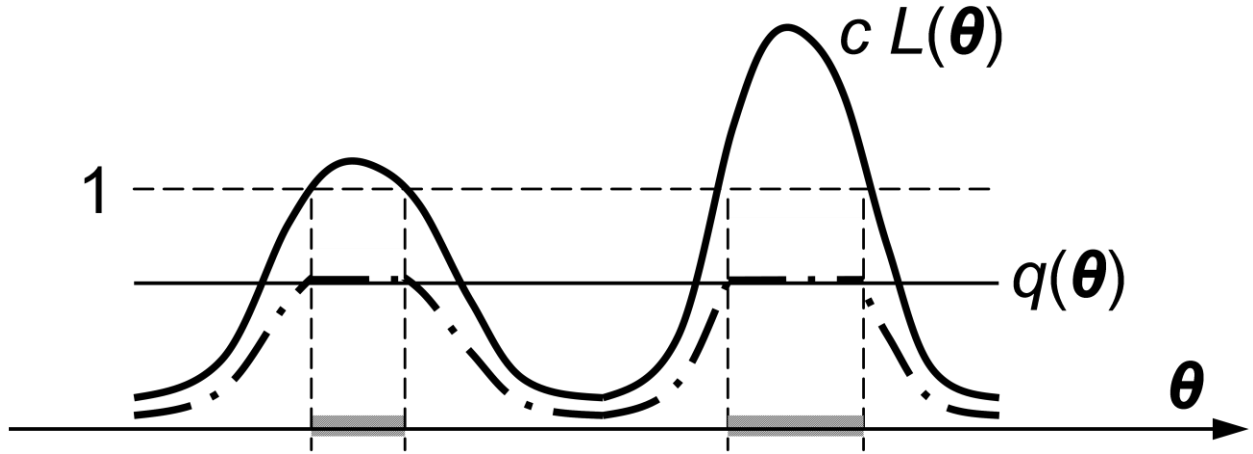


Figure. 1. Truncation of distribution in rejection algorithm. Center line – resulting distribution (short of the constant P_F^{-1}); shaded interval – truncation region B where $cL(\theta) > 1$.

Distribution invariance to admissible multiplier

As long as the multiplier satisfies the inequality in (14), it is completely arbitrary and it does not affect the distribution of the resulting samples, which is equal to the correct posterior PDF. This observation is trivial but has important implications. In the original BUS context, for example, it implies that the samples generated in different simulation runs with different admissible values of the multiplier can be simply averaged for estimating posterior statistics, because they all have the same correct posterior distribution. This fact shall also be used later when developing the proposed algorithm in this work.

Alternative BUS formulation

Having clarified the role of the multiplier, we now present a modification of the original BUS formulation that isolates the effect of the multiplier in a fundamental manner. This leads to a formulation where SuS can be performed without having to choose the multiplier before the simulation run; and where the effect of the multiplier appears clearly in the accuracy of the posterior distribution.

The modification is based on the simple observation that the failure event in (19) can be rewritten as

$$F = \left\{ \ln \left[\frac{L(\boldsymbol{\Theta})}{U} \right] > -\ln c \right\} \quad (30)$$

This means that the driving variable in SuS can be defined as

$$Y = \ln \left[\frac{L(\boldsymbol{\Theta})}{U} \right] \quad (31)$$

and the target failure event can now be written as

$$F = \{Y > b\} \quad (32)$$

where

$$b = -\ln c \quad (33)$$

The base of the logarithm is arbitrary but we choose to use natural logarithm here to facilitate the mathematical analysis.

Despite the apparently slight change in definition of the driving variable, the setup above changes the philosophy behind the multiplier and the way SuS is implemented to produce the posterior samples. The driving variable no longer depends on the multiplier and so the choice of the latter is no longer needed to start the SuS run. The multiplier only affects the target threshold level b beyond which the samples can be collected as posterior samples. As remarked at the end of the last section, as long as the multiplier is sufficiently small to satisfy the inequality in (14), the distribution of the samples conditional on the failure event $F = \{U < cL(\boldsymbol{\Theta})\}$ is invariably equal to the posterior distribution. This implies that in the proposed formulation the distribution of the samples conditional on $\{Y > b\}$ will settle (remain unchanged) for sufficiently large b . In the original BUS formulation where the driving variable is defined as $Y = cL(\boldsymbol{\Theta}) - U$ in (24) for a particular value of c (assumed to be admissible), only the samples conditional on the failure event $F = \{Y > 0\}$, i.e., for a threshold value of exactly zero, have the posterior distribution.

Substituting $b = -\ln c$ from (33) into (14) and rearranging, the inequality constraint in terms of b is given by, for all $\boldsymbol{\theta}$,

$$b > \ln L(\boldsymbol{\theta}) \quad (34)$$

From (26), the maximum admissible value of c is $c_{\max} = [\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta})]^{-1}$. Correspondingly the minimum value of b beyond which the distribution of samples will settle at the posterior PDF is

$$b_{\min} = -\ln c_{\max} = \ln[\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta})] \quad (35)$$

Similar to c_{\max} , the value of b_{\min} is generally unknown but this does not affect the SuS run. Under the proposed formulation, one can simply perform SuS with increasing levels until one determines that the threshold level of the highest level has passed b_{\min} . Despite not knowing b_{\min} , this turns out to be a more well-defined task as it is shown later that the CCDF of Y , i.e., $P(Y > b)$ versus b , has characteristic behavior for $b > b_{\min}$.

The authors believe that, while respecting the originality of BUS, the proposed formulation resolves the issue with the multiplier, as the requirement of choosing it a priori in the original formulation has been eliminated. The theoretical foundation of the proposed formulation is encapsulated in the following theorem. The first claim of the theorem has been established in (22). The second claim will be shown in (40).

Theorem

Let $\boldsymbol{\Theta} \in R^n$ be distributed as $q(\boldsymbol{\theta})$, U be uniformly distributed on $[0,1]$; $\boldsymbol{\Theta}$ and U be independent. Let $L(\boldsymbol{\theta})$ be a non-negative scalar function of $\boldsymbol{\theta} \in R^n$. Define $Y = \ln[L(\boldsymbol{\Theta})/U]$. Then for any $b > \ln[\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta})]$,

1) the distribution of $\boldsymbol{\Theta}$ conditional on $\{Y > b\}$ is $p_D(\boldsymbol{\theta}) = P_D^{-1} q(\boldsymbol{\theta}) L(\boldsymbol{\theta})$ where

$P_D = \int q(\mathbf{z}) L(\mathbf{z}) d\mathbf{z}$ is a normalizing constant;

2) $P_D = e^b P(Y > b)$.

■

Remarks on the definition of driving variable

From first glance it appears that the driving variable could have been defined without the logarithm, without affecting the canonical nature of the proposed formulation. The logarithm is

introduced for analytical and computational reasons, so that the driving variable is a well-defined random variable. In particular, if the driving variable is defined without the logarithm, i.e., $Y = L(\Theta)/U$, then, since Θ and U are independent, its expectation is given by

$$E\left[\frac{L(\Theta)}{U}\right] = E[L(\Theta)]E[U^{-1}] \quad (36)$$

However, $E[U^{-1}] = \int_0^1 u^{-1} du = \ln u \Big|_0^1$ is unbounded and so is $E[L(\Theta)/U]$. Similarly, the second moment $E[L(\Theta)^2 U^{-2}] = E[L(\Theta)^2]E[U^{-2}]$ is unbounded because $E[U^{-2}] = \int_0^1 u^{-2} du = -u^{-1} \Big|_0^1$ is unbounded.

On the other hand, for the driving variable proposed in (31),

$$Y = \ln\left[\frac{L(\Theta)}{U}\right] = \ln L(\Theta) + \ln(U^{-1}) \quad (37)$$

For U uniformly distributed on $[0,1]$, $\ln(U^{-1})$ is exponentially distributed with mean 1. For a well-posed likelihood function $L(\Theta)$ one can expect that $\ln L(\Theta)$ is a well-defined random variable when Θ is distributed as q , and so is the driving variable Y . In particular, if the first two moments of $\ln L(\Theta)$ are bounded, then the same is also true for the first two moments of Y because

$$\begin{aligned} E[Y] &= E[\ln L(\Theta)] + E[\ln(U^{-1})] \\ &= E[\ln L(\Theta)] + 1 \end{aligned} \quad (38)$$

$$\begin{aligned} E[Y^2] &= E\{[\ln L(\Theta) + \ln(U^{-1})]^2\} \\ &= E\{[\ln L(\Theta)]^2\} + 2E[\ln L(\Theta)]E[\ln(U^{-1})] + E\{[\ln(U^{-1})]^2\} \\ &= E\{[\ln L(\Theta)]^2\} + 2E[\ln L(\Theta)] + 2 \end{aligned} \quad (39)$$

since $E[\ln(U^{-1})] = 1$ and $E\{[\ln(U^{-1})]^2\} = 2$ (properties of exponential variable $\ln(U^{-1})$).

Bayesian model class selection

In addition to providing the posterior distribution and estimating the updated expectation in (3), the posterior samples can be used for estimating the normalizing constant P_D in (2). This is the

primary target of computation in Bayesian model class selection problems, where competing models are rated. In this section we show how this can be done using the conditional samples generated by SuS in the context of the proposed formulation.

Let b be an admissible threshold level, i.e., $b > b_{\min}$, so that the samples conditional on $\{Y > b\}$ have the correct posterior distribution $p_D(\boldsymbol{\theta})$. Consider the failure probability $P(Y > b)$, which can be estimated using the samples in SuS. Since $Y = \ln[L(\boldsymbol{\theta})/U]$ and $(\boldsymbol{\theta}, U)$ has a joint parameter PDF $q(\boldsymbol{\theta})I(0 < u < 1)$, $P(Y > b)$ is given by

$$\begin{aligned} P(Y > b) &= \int \int q(\boldsymbol{\theta})I(0 < u < 1)I\left\{\log\left[\frac{L(\boldsymbol{\theta})}{u}\right] > b\right\}dud\boldsymbol{\theta} \\ &= \int q(\boldsymbol{\theta})\int_0^1 I[u < e^{-b}L(\boldsymbol{\theta})]dud\boldsymbol{\theta} \\ &= e^{-b} \int q(\boldsymbol{\theta})L(\boldsymbol{\theta})d\boldsymbol{\theta} \end{aligned} \tag{40}$$

since $\int_0^1 I[u < e^{-b}L(\boldsymbol{\theta})]du = e^{-b}L(\boldsymbol{\theta})$ when $e^{-b}L(\boldsymbol{\theta}) < 1$ for all $\boldsymbol{\theta}$ (b is admissible). Observe from (2) that P_D is simply the last integral in (40). Thus,

$$P_D = e^b P(Y > b) \quad b > b_{\min} \tag{41}$$

That is, when $b > b_{\min}$, P_D can be obtained as the product of e^b and the failure probability $P(Y > b)$ it corresponds to.

Equation (41) can be rewritten as

$$P(Y > b) = e^{-b} P_D \quad b > b_{\min} \tag{42}$$

Since P_D is constant for a given problem, this suggests that for sufficiently large b , $P(Y > b)$ will decay exponentially with b . Interpreting $P(Y > b)$ as the CCDF of Y , this exponential decay gives a picture similar to a typical CCDF encountered in reliability analysis. This is another (though secondary) merit of introducing the logarithm in the definition of the driving variable Y in (31).

Characteristic trends

As shown in the last section, when $b > b_{\min}$ the failure probability $P(Y > b)$ in the proposed context is theoretically related to the evidence P_D through (41). In the actual implementation b_{\min} is not known and so it is necessary to determine whether $b > b_{\min}$ so that the samples conditional on $\{Y > b\}$ can be confidently collected as the correct posterior samples. Below we argue that the variation of $P(Y > b)$ with b takes on different characteristics on two different regimes of b . This may be utilized to judge in a SuS run whether the threshold value of a particular simulation level has already passed b_{\min} , thereby determining the stopping criterion.

First, note that $P(Y > b)$ is a non-increasing function of b . When b is at the left tail of the CCDF, $P(Y > b) \approx 1$ and it typically decreases with b , equal to P_D at $b = b_{\min}$. When $b > b_{\min}$ we know from (42) that $P(Y > b) = P_D e^{-b}$ and so it decays exponentially with b . We can thus expect that, as b increases from the left tail and passes b_{\min} , the CCDF of Y typically changes from a decreasing function to a fast (exponentially) decaying function. Correspondingly, the function $\ln P(Y > b)$ changes from a slowly decreasing function to a straight line with a slope of -1 .

On the other hand, consider the following function:

$$V(b) = b + \ln P(Y > b) \quad (43)$$

This function can be used for computing the log-evidence $\ln P_D$ as it can be readily seen that

$$V(b) = \ln P_D \quad b > b_{\min} \quad (44)$$

When b is at the left tail of the CCDF, $\ln P(Y > b) \approx 0$ and so $V(b) \approx b$ increases linearly with b . The above means that as b increases from the left tail of the CCDF of Y the function $V(b)$ increases linearly, going through a transition until it settles (remains unchanged) at $\ln P_D$ after $b > b_{\min}$. The characteristic behavior of $\ln P(Y > b)$ and $V(b)$ are depicted in Figure 2.

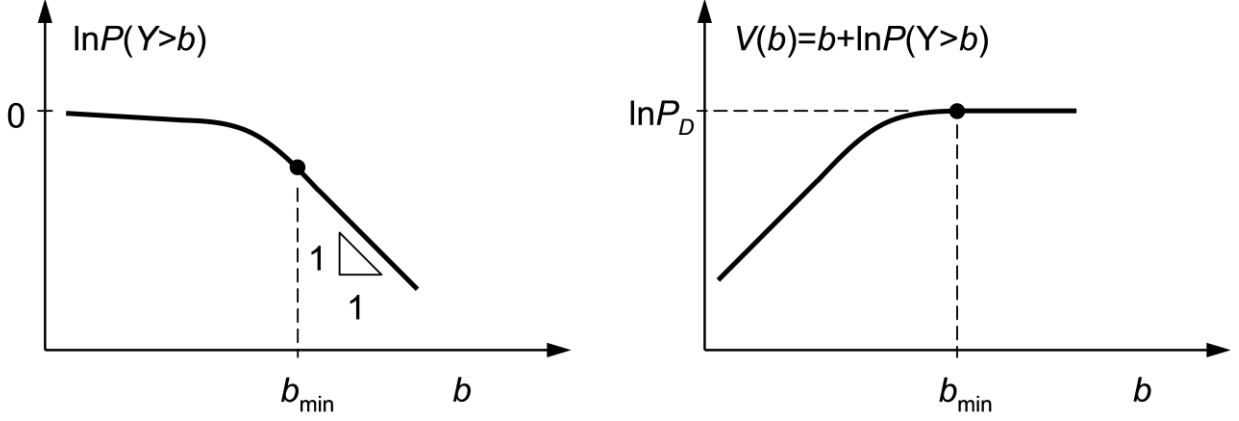


Figure 2. Characteristic trends of $\ln P(Y > b)$ and $V(b)$.

Strictly speaking, the above arguments only apply to the theoretical quantities. In a SuS run the quantities $\ln P(Y > b)$ and $V(b)$ as a function of b can only be estimated on a sample basis. The resulting estimated counterparts will exhibit random deviation from the theoretical trends due to statistical estimation error, whose extent depends on the number of samples used in the simulation run (the larger the number of samples, the smaller the error). Nevertheless, the above arguments and Figure 2 provide the basis for determining the simulation level to stop and to collect the posterior samples. That is, one stops at the simulation believed to have passed the transition.

Subset Simulation under the new framework

In the proposed context, the posterior samples can be obtained from the conditional samples in a straightforward manner from a SuS run. No modification of SuS is necessary. Below we outline how this can be done, focusing only on issues directly related to the Bayesian updating problem.

Recall the Bayesian updating problem where the primary target is to generate posterior samples of $\Theta \in R^n$ distributed as the posterior PDF $p_D(\Theta) \propto q(\Theta)L(\Theta)$, where $q(\Theta)$ is the prior distribution assumed to be chosen from a standard class of distributions (e.g., Gaussian, exponential); and $L(\Theta)$ is the likelihood function for a given set of data. As reviewed in the section “Subset Simulation”, a SuS run produces the estimate of the CCDF of the driving

variable Y , i.e., $P(Y > b)$ versus b . The posterior samples for Bayesian model updating can be obtained as the conditional samples in a SuS run for the reliability problem with driving variable $Y = \ln[L(\Theta)/U]$, where Θ is distributed as $q(\Theta)$ and U is uniformly distributed on $[0,1]$; Θ and U are independent. The conditional samples are collected from the level whose threshold level is determined to be greater than b_{\min} .

Stopping criteria

During SuS, suppose simulation levels $i = 0, 1, \dots, m$ has been performed, resulting in the estimates $\{(b_k^{(i)}, p_k^{(i)}) : k = 1, \dots, N(1 - p_0)\}$ (Levels $i = 0, 1, \dots, m-1$) and $\{(b_k^{(m)}, p_k^{(m)}) : k = 1, \dots, N\}$ (Level m) for the CCDF of Y . Plot the estimate of $\ln P(Y > b)$ and $V(b) = e^b P(Y > b)$ versus b . Examine these two curves and determine whether the threshold level b_m of Level m has passed the point b_{\min} depicted in Figure 2. If this is not the case then proceed to the next level. Otherwise, there is no need to proceed to the next level and the simulation can be stopped. In this case, take the samples $\{\Theta_k^{(m)} : k = 1, \dots, N\}$ at Level m (the highest level performed, conditional on $\{Y > b_m\}$) as the posterior samples.

The number of simulation levels m to reach the target level (i.e., with $b_m > b_{\min}$) depends on $P(Y > b_{\min})$. The smaller the $P(Y > b_{\min})$ the larger the m required. Assume for simplicity that $b_m = b_{\min}$. Since $P(Y > b_{\min}) = p_0^m$, the number of levels required is

$$m = \frac{\ln P(Y > b_{\min})}{\ln p_0} \quad (45)$$

Evaluating (42) at $b = b_{\min}$ gives $P(Y > b_{\min}) = e^{-b_{\min}} P_D$. Substituting $b_{\min} = \max_{\Theta} L(\Theta)$ from (35) and noting that $P_D = \int L(\Theta) q(\Theta) d\Theta = E[L(\Theta)]$ gives

$$P(Y > b_{\min}) = \frac{E[L(\Theta)]}{\max_{\Theta} L(\Theta)} \quad (46)$$

and so

$$m = (\ln p_0)^{-1} \ln \left\{ \frac{\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta})}{E[L(\boldsymbol{\theta})]} \right\} \quad (47)$$

Equations (46) and (47) are invariant to the scaling of the likelihood function. They show that the number of levels to reach the target level depends on how high the likelihood function can reach $\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta})$ compared to its average value $E[L(\boldsymbol{\theta})]$. In this sense, problems requiring more computational effort are those when there is a large disparity between the highest value of the likelihood function and typical values of the likelihood function sampled from the prior PDF. Those problems are precisely the ones that require more advanced strategies than Direct Monte Carlo (i.e., sampling from the prior). This is very intuitive and reveals fundamentally one computational challenge of Bayesian updating problems. Note that the same conclusions can also be deduced from the original BUS formulation by following a similar argument.

Posterior statistical estimation

The posterior samples $\{\boldsymbol{\Theta}_k^{(m)} : k = 1, \dots, N\}$ obtained from simulation level m for which $b_m > b_{\min}$ can be used for estimating posterior statistics in Bayesian updating problem and the evidence for Bayesian model class section. For the former, the posterior expectation in (3) is estimated by simple averaging:

$$E[r(\boldsymbol{\theta}) | D, \mathcal{M}] \approx \frac{1}{N} \sum_{k=1}^N r(\boldsymbol{\Theta}_k^{(m)}) \quad (48)$$

On the other hand, based on (41), the evidence can be estimated by

$$P(D | \mathcal{M}) = P_D \approx \tilde{P}_D = e^{b_m} p_0^m \quad (49)$$

Taking logarithm, the log-evidence is estimated by

$$\ln P(D | \mathcal{M}) = \ln P_D \approx \ln \tilde{P}_D = b_m + m \ln p_0 \quad (50)$$

Standard Gaussian space

As a remark on implementation, it is recommended to perform SuS in standard Gaussian space. That is, one starts with the standard Gaussian variables $\{Z_1, \dots, Z_{n+1}\}$ as the uncertain parameters in the SuS run and map them to the required variables in the problem, by $\boldsymbol{\Theta} = T(Z_1, \dots, Z_n)$ and

$U = \Phi(Z_{n+1})$. Here $T(\cdot)$ is the Rosenblatt transformation that gives Θ the desired distribution $q(\Theta)$; and $\Phi(\cdot)$ denotes the standard Gaussian CDF responsible for mapping a the standard Gaussian variable Z_{n+1} to the uniformly distributed variable U on $[0,1]$. Performing SuS in the standard Gaussian space is preferred because it is easier to choose the one-dimensional proposal distribution for the independent-component MCMC algorithm. This is especially relevant for the variable U uniformly distributed on $[0,1]$, whose bounded domain can lead to frequent rejection near the boundaries. In the standard Gaussian space, one typical choice for the one-dimensional proposal PDF is the standard Gaussian PDF or a uniform PDF centered at the current sample. Operating in the standard Gaussian space was also mentioned in the original BUS formulation (Straub & Papaioannou 2014).

Statistical error assessment

Some comments are in order regarding the statistical error of the results, in terms of the quality of the posterior samples and the statistical variability of the log-evidence estimator. Provided that the threshold value of the simulation level is greater than b_{\min} , its conditional samples are always distributed as the target posterior PDF $p_D(\Theta)$. As MCMC samples they are correlated, however. When used for statistical estimation they will give less information compared to if they were independent. Typically their correlation tends to increase with the simulation level. In view of this, it is not necessary to perform more simulation levels than is necessary to pass b_{\min} . This is the reason behind the stopping criterion.

For the evidence estimate in (49), it should be noted that its statistical variability arises from b_m .

By taking small random perturbation of the estimation formula, it can be reasoned that

$$\text{c.o.v. of } \ln \tilde{P}_D \approx \text{std. of } \ln \tilde{P}_D \approx \text{std. of } b_m \quad (51)$$

where ‘std.’ is an abbreviation for ‘standard deviation’. An estimation formula for the c.o.v. of b_m based on samples in a single SuS run is not available, however. Conventionally only the c.o.v. of the estimate \tilde{P}_b (say) for $P(Y > b)$ for fixed b is available, rather than the c.o.v. of the quantile value b_m for fixed exceedance probability. It can be reasoned, however, that the c.o.v.

of \tilde{P}_D (where b_m is random) can be approximated by the c.o.v. of $e^b \tilde{P}_b$ for fixed b (then taking $b = b_m$ obtained in a simulation run). The latter is equal to the c.o.v. of \tilde{P}_b , for which standard estimation formula is available (Au & Beck 2001, Au & Wang 2014).

Comparison with original BUS formulation

Table 1 provides a comparison between BUS and the proposed formulation. Implementing SuS under the proposed framework has several advantages over the original BUS, stemming mainly from the treatment of the multiplier in the former. First of all, there is no need to determine the appropriate value of the multiplier to start the simulation run. The definition of the driving variable is more intrinsic as it only depends on the likelihood function and not on the multiplier. In the BUS context, if the chosen value of the multiplier is not small enough, it will lead to bias in the distribution of the samples, unfortunately in the high likelihood region of the posterior distribution that is most important. If it is chosen too small it will result in lower efficiency, as it requires more simulation levels to reach the target event from which the samples can be taken as posterior samples. In both cases if it is found after a SuS run that the choice of the multiplier is not appropriate, one needs to perform an additional run with a (hopefully) better choice of the multiplier. These issues are all irrelevant in the proposed context because the problem specification of the SuS run does not depend on the multiplier.

Table 1. Comparison of BUS and proposed formulation

	BUS*	Proposed
Driving variable	$Y = cL(\Theta) - U$ for any $c < [\max_{\Theta} L(\Theta)]^{-1}$	$Y = \ln[L(\Theta)/U]$
Target failure event	$F = \{Y > 0\}$	$F = \{Y > b\}$ for any $b > \ln[\max_{\Theta} L(\Theta)]$
Evidence calculation	$P_D = cP(Y > 0)$	$P_D = e^b P(Y > b)$ for any $b > \ln[\max_{\Theta} L(\Theta)]$

Stopping criterion	When threshold value of simulation level is equal to zero	After log-evidence becomes flat with threshold level; or after log-failure probability displays a slope of -1
--------------------	---	---

* Original definition is $Y = U - cL(\Theta)$. Changed here for consistency with SuS literature

On the other hand, in the BUS context the posterior samples must be obtained as those conditional on the target failure event $\{Y > 0\}$ where $Y = cL(\Theta) - U$. For example, samples conditional on $\{Y > 0.1\}$ cannot be directly used. Since the threshold values $\{b_1, b_2, \dots\}$ generated adaptively in different simulation levels of SuS are random, they generally do not coincide with 0, i.e., the target threshold value of interest. In this case, not all samples can be used directly as conditional samples. In the original BUS algorithm (Straub & Papaioannou 2014), if the threshold level of the next level determined adaptively from the samples of the current level is greater than zero, it is set equal to zero so that the next (and final) level is exactly conditional on $\{Y > 0\}$. In the proposed context, the posterior samples can be directly collected from the samples generated in SuS. This is because any sample conditional on $\{Y > b\}$ with $b > b_{\min}$ ($Y = \log[L(\Theta)/U]$) can be taken as a posterior sample. The value of b_{\min} is unknown but whether $b > b_{\min}$ can be determined from the sample estimates of $P(Y > b)$ or $V(b) = b + \ln P(Y > b)$ versus b , based on their characteristic behavior (see Figure 2).

Illustrative examples

We now present two examples that illustrate the applicability of the proposed methodology. The first one is the locally identifiable case of a two-degree-of-freedom shear building model originally presented in Beck and Au (2002). The second example is the unidentifiable case of the same model.

Example1. Two-DOF shear frame: locally identifiable case

Consider a two-storied building structure represented by a two-degree-of-freedom shear building model. The objective here is to identify the interstory stiffnesses which allow the structural

response to be subsequently updated. The first and second story masses are given by 16.5×10^3 kg and 16.1×10^3 kg respectively. Let $\boldsymbol{\theta} = [\theta_1, \theta_2]$ be the stiffness parameters to be identified. The interstory stiffnesses are thus parameterized as $k_1 = \theta_1 \bar{k}_1$ and $k_2 = \theta_2 \bar{k}_2$, where the nominal values for the stiffnesses are given by $\bar{k}_1 = \bar{k}_2 = 29.7 \times 10^6$ N/m. The joint prior distribution q for θ_1 and θ_2 is assumed to be the product of two Lognormal distributions with most probable values 1.3 and 0.8 respectively and unit standard deviations. For further details on the assumptions behind the parameterization and the choice of nominal values, refer to Beck and Au (2002). Let $D = \{\tilde{f}_1, \tilde{f}_2\}$ be the modal data used for the model updating, where $\tilde{f}_1 = 3.13$ Hz and $\tilde{f}_2 = 9.83$ Hz are the identified natural frequencies. The posterior PDF is formulated following Vanik et al. (2000) as

$$p_D(\boldsymbol{\theta}) \propto \exp[-J(\boldsymbol{\theta}) / 2\varepsilon^2] q(\boldsymbol{\theta}) \quad (52)$$

where ε is the standard deviation of the prediction error and $J(\boldsymbol{\theta})$ is a modal measure-of-fit function given by

$$J(\boldsymbol{\theta}) = \sum_{j=1}^2 \lambda_j^2 [f_j^2(\boldsymbol{\theta}) / \tilde{f}_j^2 - 1]^2 \quad (53)$$

Here, λ_1 and λ_2 are weights and $f_1(\boldsymbol{\theta})$ and $f_2(\boldsymbol{\theta})$ are the modal frequencies predicted by the corresponding finite element model.

For the implementation of SuS, a conventional choice of algorithm parameters in the reliability literature is adopted in this study. The level probability is chosen to be $p_0 = 0.1$ and the number of samples per level N is fixed at 10,000. In the standard Gaussian space, the one-dimensional proposal PDF is chosen to be uniform with a maximum step width of 1. A relatively large number of samples per level is been chosen in this study to illustrate the theoretical aspects of the proposed method. Strategies for efficiency improvement such as adaptive proposal PDF or likelihood function can be explored but are not further investigated here.

Figures 3 and 4 show the Markov chain samples for $\boldsymbol{\theta} = [\theta_1, \theta_2]$ at six consecutive simulation levels. The results are shown respectively in the standard Gaussian space, where the simulation was carried out, and the Lognormal space after the application of the relevant transformation.

Level 0 corresponds to the unconditional case (i.e., Direct Monte Carlo), that is, the joint prior PDF. As the simulation level ascends, the distribution of the samples evolves from the prior distribution to the target posterior distribution, which is bimodal in the present example.

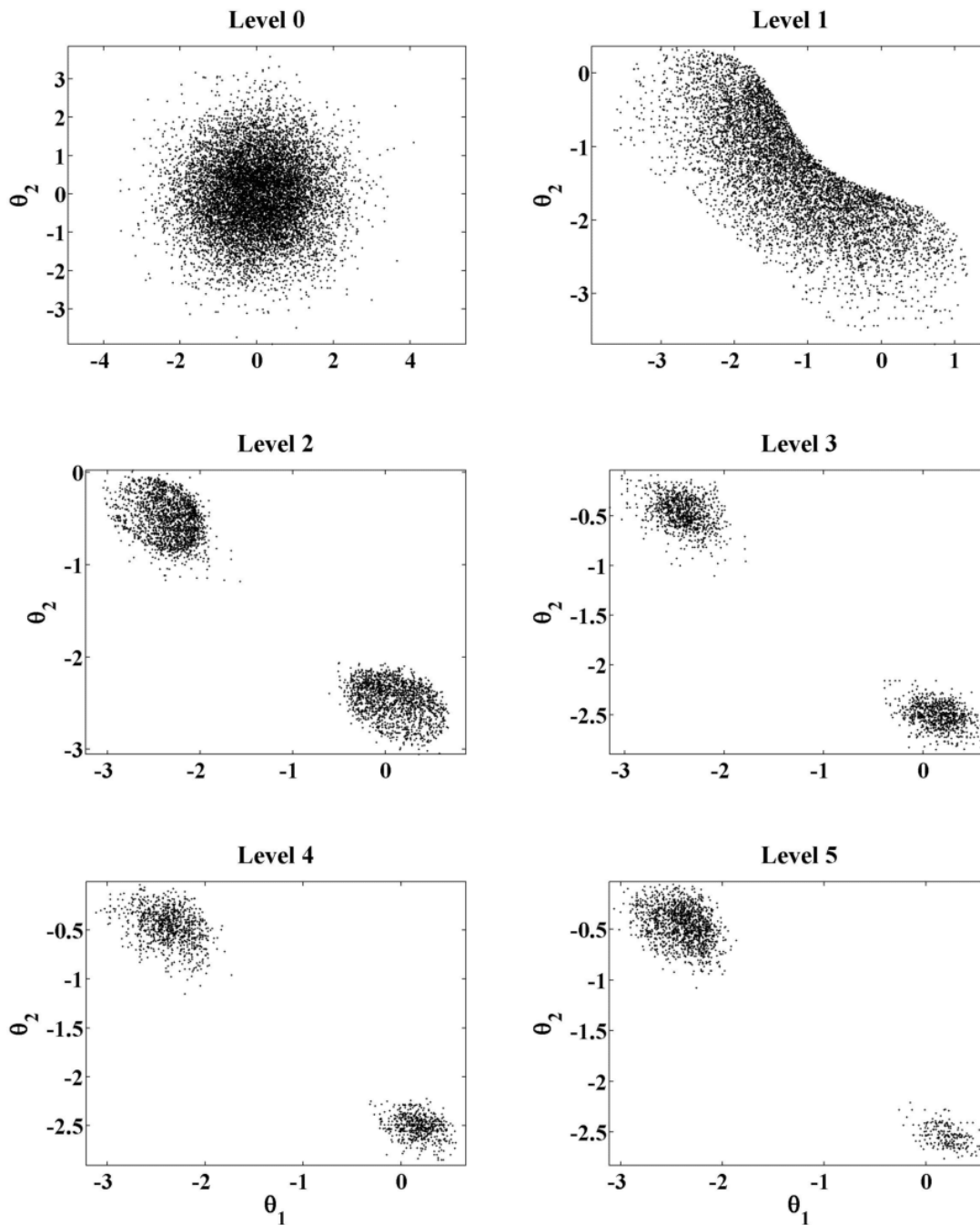


Figure 3. Markov chain samples in the standard Gaussian space for the stiffness parameters $\boldsymbol{\theta} = [\theta_1, \theta_2]$ from Level 0 (prior distribution) to Level 5.

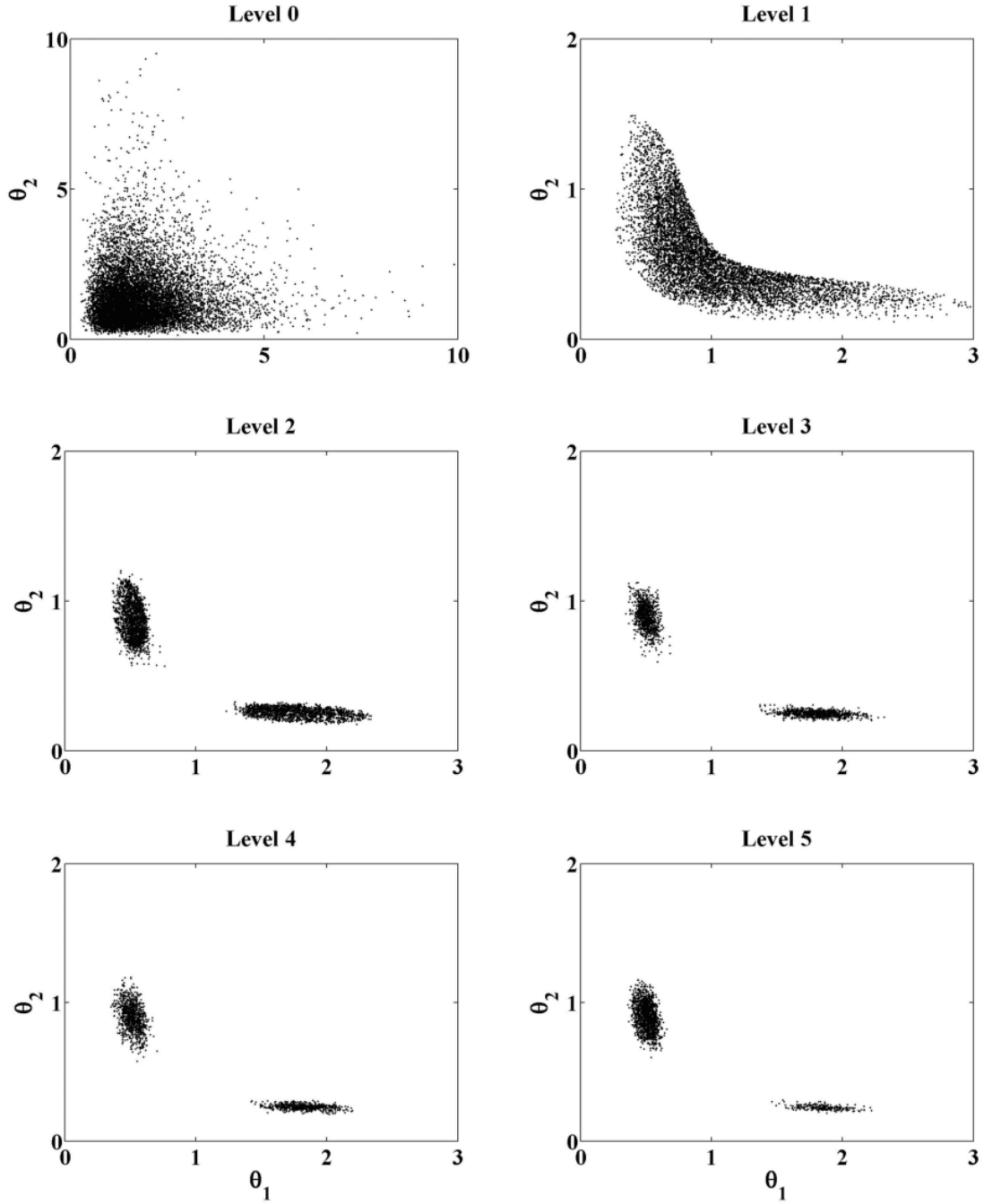


Figure 4. Markov chain samples in the Lognormal space for the stiffness parameters $\boldsymbol{\theta} = [\theta_1, \theta_2]$ from Level 0 (prior distribution) to Level 5.

Figures 5 and 6 show the marginal histograms for θ_1 and θ_2 corresponding to those samples in Figures 3 and 4. The bimodal nature of the marginal PDFs is clearly visible from Level 2. For comparison, the solid lines show the target marginal posterior distributions obtained by

numerically integrating the expression for the posterior PDF. It is apparent that the distribution of the samples has settled after Level 3. In reality, the exact target PDF is not available and so alternative means based on the simulated samples must be employed to determine whether the distribution of the samples has settled at the target one. Within the context of the proposed methodology, this is done through the plot of the log-failure probability and log-evidence versus the threshold level, which shall be discussed next.

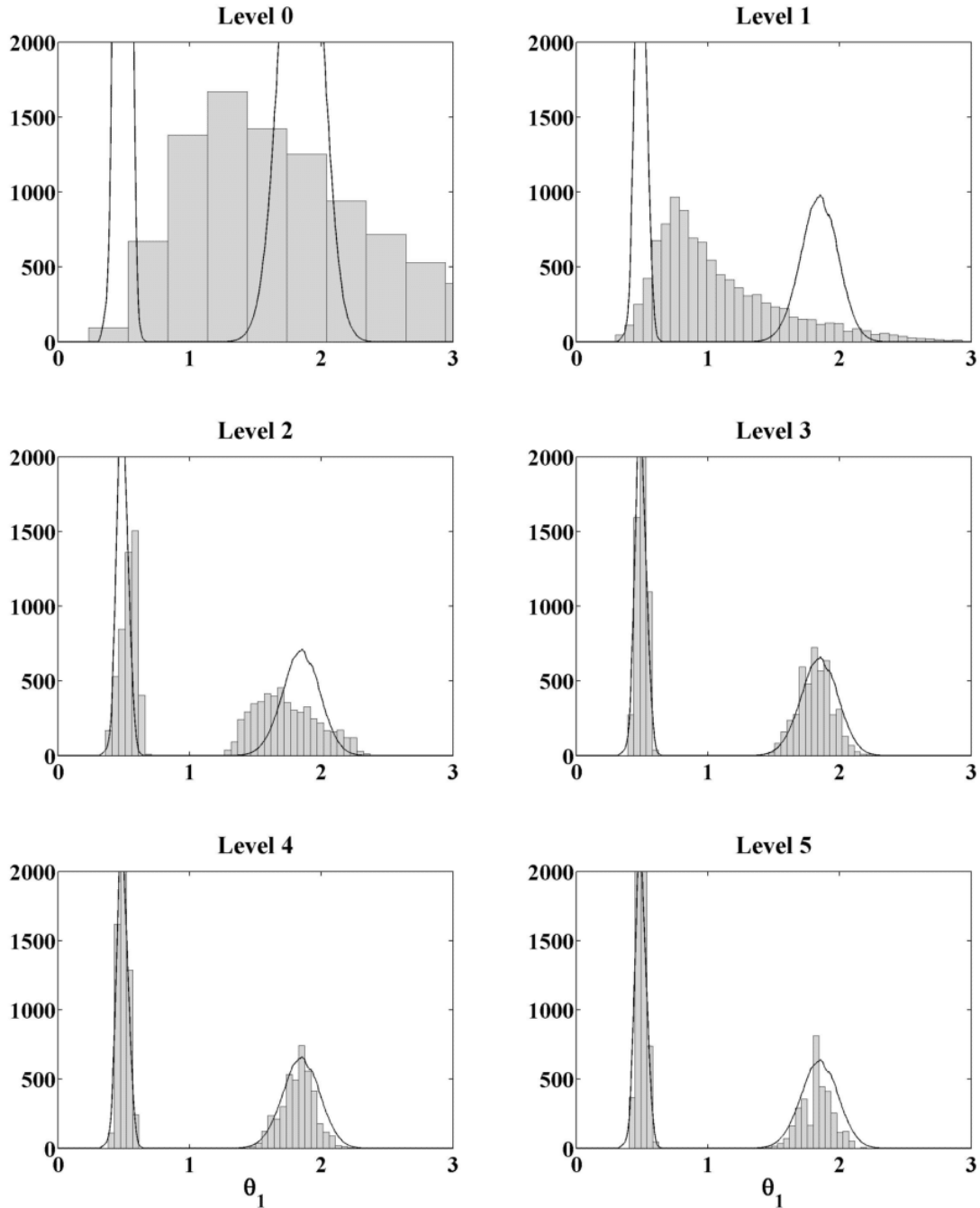


Figure 5. Posterior marginal PDF for θ_1 at different simulation levels. The target marginal posteriors were obtained numerically and are shown for comparison.

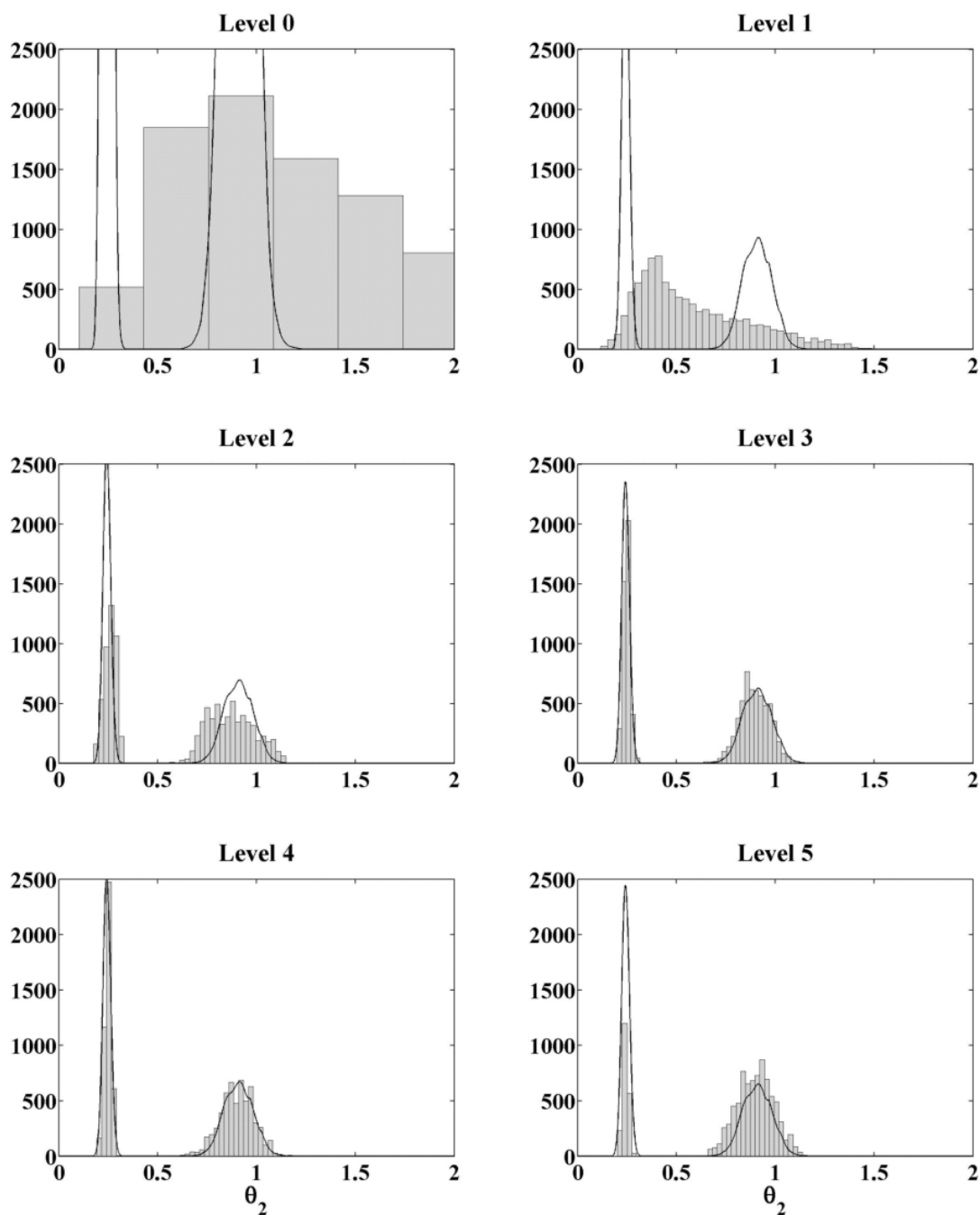


Figure 6. Posterior marginal PDF for θ_2 at different simulation levels. The target marginal posteriors were obtained numerically and are shown for comparison.

Figure 7(a) plots the estimate of the log-CCDF of Y , i.e., $\ln P(Y > b)$ versus b . The general shape of the resulting simulated curve coincides with the characteristic trend predicted by the theory, that is, there is a transition from a slowly decreasing function to a line with slope equal to -1. When zooming into the region where $b \approx 0$, the figure shows the boundaries of each level computed via SuS. The transition appears to complete somewhere after Level 3. This corresponds to the theoretical value of $c=1$, found in the original BUS formulation. Additionally, the log-evidence was computed following (43) and is shown in Figure 7(b). As with the log-CCDF, the theoretical prediction of the characteristic trend is also verified for this case. The curve flattens after Level 3. It should be noted that the samples after Level 3 theoretically all have the same distribution equal to the target posterior PDF. However, for statistical estimation their quality deteriorates as the simulation level ascends because their correlation (due to mechanism of SuS) tends to increase. According to the theory, b_{\min} should lie in the neighborhood of the threshold for Level 3. As pointed out before, the flattening of the characteristic trend of the log-evidence applies only to the theoretical (exact) quantities. In our case, the simulated curve exhibits random deviation, i.e., statistical error. Figure 5 suggests that the samples at Level 3 should be taken as the posterior samples for Bayesian updating purposes. Those samples at the higher levels can still be used but they have lower quality due to higher correlations. In fact, in the actual implementation, one can stop after it is ascertained that the threshold level of the last simulation level has passed b_{\min} , e.g., Level 4 in the present case.

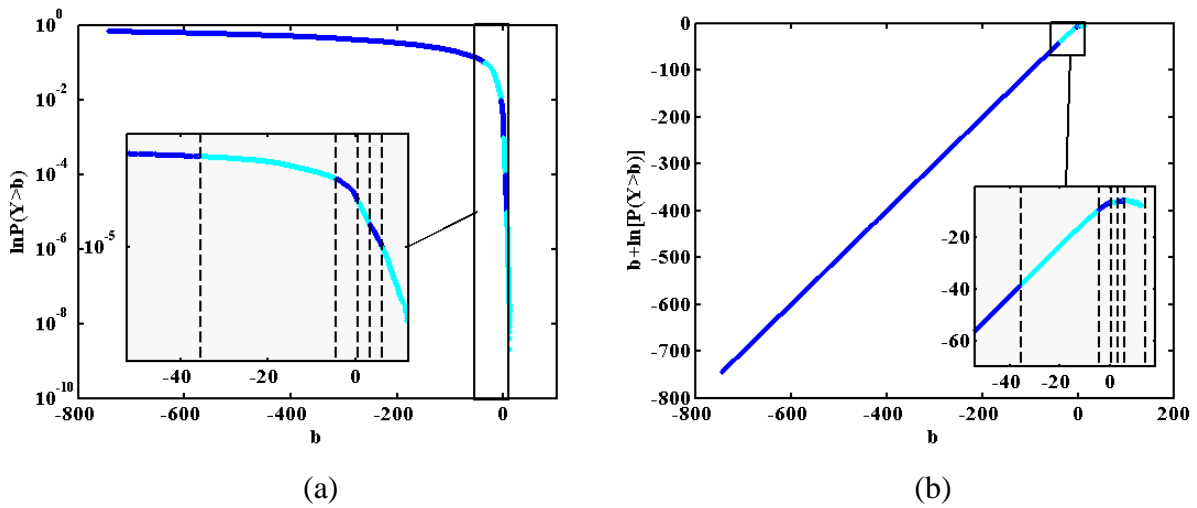


Figure 7. Log-CCDF computed through SuS (a) for the identifiable case. The curve slowly transitions into a straight line with negative unit slope. Correspondingly, the log-evidence (b)

flattens as the threshold exceeds b_{\min} . The dotted lines show the thresholds for different simulation levels.

Example 2. Two-DOF shear frame: unidentifiable case

The exercise was repeated for the case where the story masses are also unknown and need to be updated. The problem is characterized as unidentifiable, since there are an infinite number of combinations of parameter values that can explain the measured modal frequencies. In addition to the stiffnesses, the masses are parameterized as $m_1 = \theta_3 \bar{m}_1$ and $m_2 = \theta_4 \bar{m}_2$, where the nominal values for the are given by $\bar{m}_1 = 16.5 \times 10^3$ kg and $\bar{m}_2 = 16.1 \times 10^3$ kg. Thus, for this case, $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3, \theta_4]$ where the marginal prior distributions for θ_1 and θ_2 are the same Lognormals as in the previous example. The prior marginal distributions for θ_3 and θ_4 are both assumed to be Lognormals with most probable values equal to 0.95 and standard deviation of 0.1. The joint prior PDF is therefore taken as the product of the four Lognormals. Figure 8 shows the Markov chain samples for $\boldsymbol{\theta}$ (θ_1 versus θ_2 for visualization purposes) at simulation levels 0 through 5. Again, the updated distribution results in a bimodal posterior PDF. For the sake of brevity, the results are only displayed in the Lognormal space.

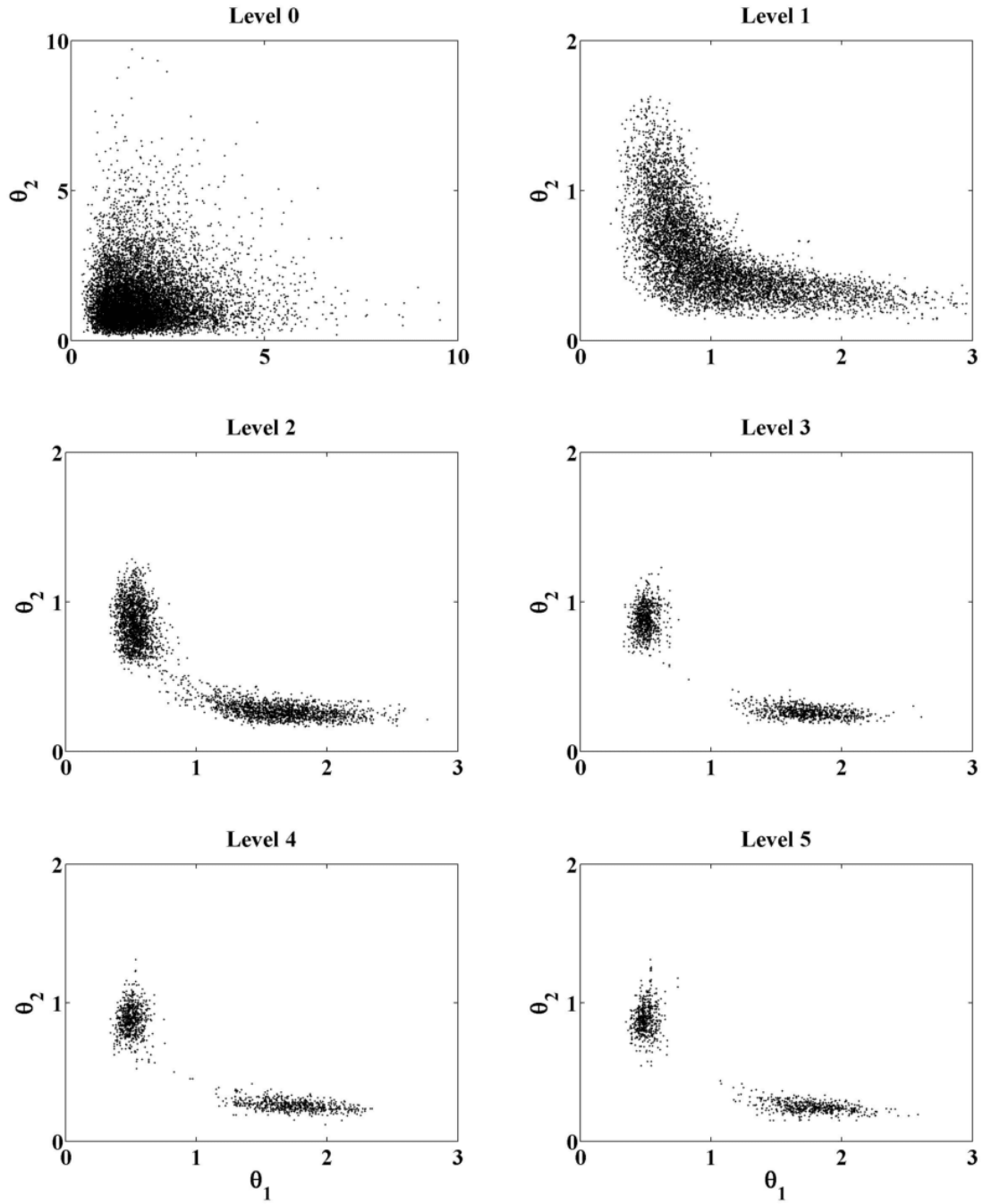


Figure 8. Markov chain samples in the standard Gaussian space for the stiffness parameters θ_1 and θ_2 of the unidentifiable case at simulation levels 0 (prior distribution) to level 5.

Analogously, Figure 9 shows the samples for θ_3 and θ_4 in the Lognormal space. There is no noticeable pattern in the distribution of the masses, consistent with the findings in Beck and Au (2002).

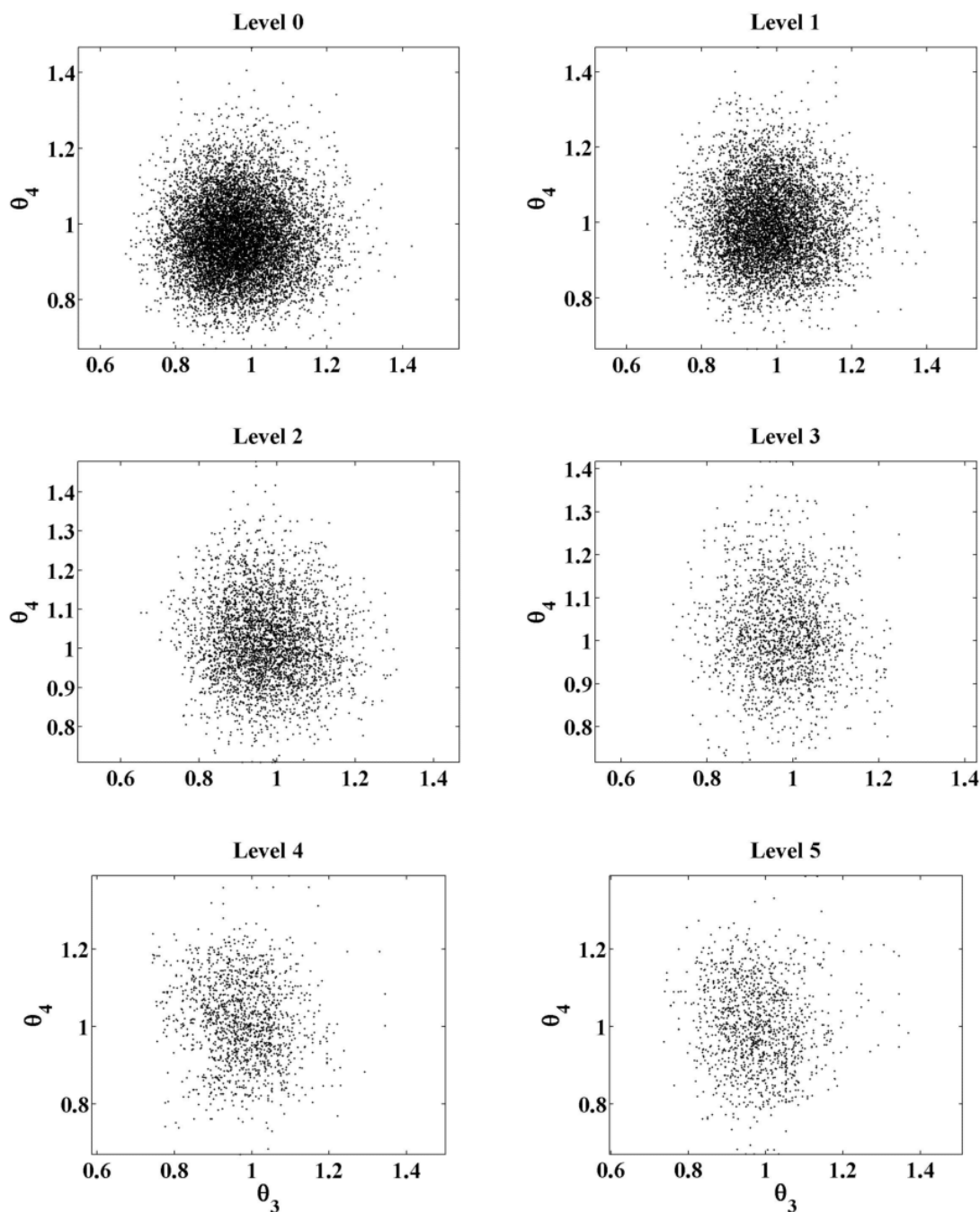


Figure 9. Markov chain samples in the Lognormal space for the mass parameters θ_3 and θ_4 of the unidentifiable example at simulation levels 0 to level 5.

The log-CCDF for the unidentifiable case shows similar characteristics to the ones displayed by the locally identifiable case. The transition into a straight line also appears to be complete in Level 3 and thus the optimal b should lie in this neighborhood. This is confirmed by the flattening log-evidence curve. The results are displayed in Figure 10.

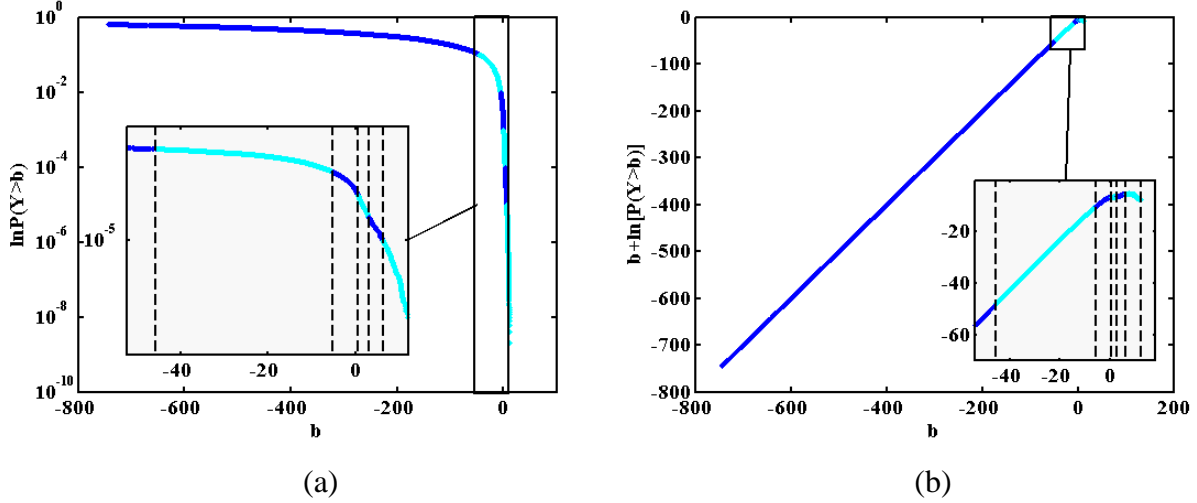


Figure 10. Log-CCDF computed through SuS (a) for the unidentifiable case. The curve slowly transitions into a straight line with negative unit slope. Correspondingly, the log-evidence (b) flattens as the threshold passes b_{\min} . The dotted lines show the thresholds for different simulation levels.

Example 3. Model Class Selection

Following the two preceding examples, we can compute the log-evidence corresponding to each model according to equation (44). Figure 12 shows the ratio of the evidence for the unidentifiable case to the evidence of the locally identifiable case. Discounting the random deviation due to simulation error, the ratio of evidence seems to converge to 1, which suggests that, given the available data, there is not enough evidence to prefer the unidentifiable model over the more parsimonious one.

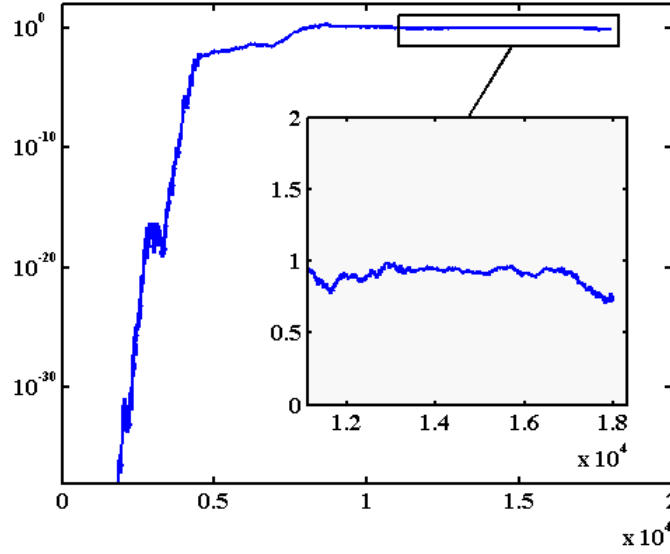


Figure 11. Ratio of evidence of the unidentifiable model to the evidence of the locally identifiable model.

Conclusions

We have presented a fundamental analysis of BUS, a recently proposed framework that establishes an analogy between the Bayesian updating problem and the reliability problem. This work was motivated by the question of choosing the correct likelihood multiplier and it has led to an improved formulation which resolves this question. By redefining the target failure event, we have expressed the driving variable in the equivalent reliability problem using the likelihood function alone, without the multiplier. This redefinition provides the key advantage over the original BUS, since our implementation no longer requires a predetermined value for the multiplier in order to start the SuS runs. This immediately eliminates the need to perform additional runs in case an inadmissible or inefficient value for the multiplier is chosen. Moreover, the samples generated at different levels of SuS can be used directly as posterior samples as long as their thresholds are greater than the minimum admissible value. We have also performed a rigorous analysis that predicts the properties of the log-evidence function, which has provided a stopping criterion for the proposed formulation. The lowest level to reach the target posterior distribution is determined by the ratio of the highest to the average value of likelihood function, which reveals a fundamental characteristic reflecting the difficulty of Bayesian

updating problems. The theoretical predictions of our study have been verified by applying our proposed strategy to two illustrative examples.

Acknowledgements

This work was performed while the first author was on short-term fellowship funded by the Japan Society for the Promotion of Science (JSPS) and hosted at the Tokyo City University, whose supports are gratefully acknowledged.

References

- Au, S. K. (2004). "Probabilistic failure analysis by importance sampling Markov chain simulation". *Journal of Engineering Mechanics*, 130(3):303-311.
- Au, S. K. (2005). "Reliability-based design sensitivity by efficient simulation". *Computers & Structures*, 83(14):1048-1061.
- Au, S. K., and Beck, J. L. (2001). "Estimation of small failure probabilities in high dimensions by subset simulation". *Probabilistic Engineering Mechanics*, 16(4):263-277.
- Au, S. K., and Beck, J. L. (2003a). "Importance sampling in high dimensions". *Structural Safety*, 25(2):139-163.
- Au, S. K., and Beck, J. L. (2003b). "Subset simulation and its applications to seismic risk based on dynamic analysis". *Journal of Engineering Mechanics*, 129(8):901-917.
- Au, S. K., and Wang, Y. (2014). *Engineering risk assessment with Subset Simulation*. John Wiley & Sons, Singapore.
- Au, S. K., Cao, Z. J., and Wang, Y. (2010). "Implementing Advanced Monte Carlo under spreadsheet environment". *Structural Safety*, 32(5):281-292.
- Betz, W., Papaioannou, I., and Straub, D. (2014). "Adaptive variant of the BUS approach to Bayesian updating". In *Proceedings of 9th European Conference on Structural Dynamics*, 30 June - 2 July 2014, Porto, Portugal.
- Beck, J. L. (2010). "Bayesian system identification based on probability logic". *Structural Control and Health Monitoring*, 17(7):825-847.

- Beck, J. L., and Au S. K. (2002). “Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation”. *Journal of Engineering Mechanics*, 128(4): 380-391.
- Beck, J. L., and Katafygiotis, L. S. (1998). “Updating models and their uncertainties. I: Bayesian statistical framework.” *Journal of Engineering Mechanics*, 124(4):455–461.
- Beck, J. L, and Yuen, K. V. (2004). “Model selection using response measurements: Bayesian probabilistic approach”, *Journal of Engineering Mechanics*, 130(2):192-203.
- Bourinet, J.M., Deheeger, F., and Lemaire, M. (2011). “Assessing small failure probabilities by combined subset simulation and Support Vector Machines”. *Structural Safety*, **33**, 343–353.
- Carlin, B.P., and Chib, S. (1995). “Bayesian model choice via Markov chain Monte Carlo methods”. *Journal of the Royal Statistical Society, B*, 57, 473-84.
- Chen, M.H., Shao, Q.M., and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag, New York.
- Cheung, S. H., and Beck, J. L. (2009). “Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters”. *Journal of Engineering Mechanics*, 135(4):243-255.
- Chiachio, M., Beck, J. L., Chiachio, J., and Rus, G. (2014). “Approximate Bayesian Computation by Subset Simulation”. *Journal of Scientific Computing*, SIAM. In print.
- Ching, J., and Chen, Y. C. (2007). “Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging”. *Journal of Engineering Mechanics*, 133(7):816-832.
- Ching, J., and Hsieh, Y. H. (2007). “Approximate reliability-based optimization using a three-step approach based on subset simulation”. *Journal of Engineering Mechanics*, **133**(4), 481-493.
- Cox, R. T. (1961). *The algebra of probable inference*, Johns Hopkins, Baltimore.
- Ewins, D. (2000). *Modal testing*, 2nd Edition, Research Studies Press, Baldock.
- Fishman, G. S. (1996). *Monte Carlo: concepts, algorithms and applications*. Springer-Verlag, New York.
- Haario, H., Saksman, E., and Tamminen, J. (2005). “Componentwise adaptation for high dimensional MCMC”. *Computational Statistics*, 20:265-273.

- Hastings, W.K. (1970). “Monte Carlo sampling methods using Markov chains and their applications”. *Biometrika*, **57**, 97–109.
- Hudson, D. E. (1977). “Dynamic tests of full-scale structures.” *Journal of Engineering Mechanics Division*, 103(6):1141–1157.
- Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press, UK.
- Katafygiotis, L. S., and Beck, J. L. (1998). “Updating models and their uncertainties. II: Model identifiability”. *Journal of Engineering Mechanics*, 124(4): 463-467.
- Katafygiotis, L. S., and Lam, H. F. (2002). “Tangential-projection algorithm for manifold representation in unidentifiable model updating problems”, *Earthquake Engineering and Structural Dynamics*, 31(4):791-812.
- Li, H. S., Au, S. K. (2010). “Design optimization using Subset Simulation algorithm”. *Structural Safety*, 32(6):384-392.
- Malakoff, D. (1999). “Bayes offers a ‘new’ way to make sense of numbers”. *Science*, 286:460-1464.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., et al. (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1091.
- Papadimitriou, C., Beck, J. L., Katafygiotis, L. S. (2001). “Updating robust reliability using structural test data”. *Probabilistic Engineering Mechanics*, 16(2):103-113.
- Papadopoulos, V., Giovanis, D. G., Lagaros, N. D., and Papadrakakis, M. (2012) Accelerated subset simulation with neural networks for reliability analysis, *Computer Methods in Applied Mechanics and Engineering*, **223–224**, 70–80.
- Robert, C., and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York.
- Schueller, G. I., Pradlwarter, H. J., and Koutsourelakis, P. S. (2004). “A critical appraisal of reliability estimation procedures for high dimensions”. *Probabilistic Engineering Mechanics*, 19:463-474.
- Song, S. F., Lu, Z. Z., and Qiao, H. W. (2009). “Subset simulation for structural reliability sensitivity analysis”, *Reliability Engineering and System Safety*, **94**, 658–665.
- Straub, D. (2011). “Reliability updating with equality information”. *Probabilistic Engineering Mechanics*, 26:254–258.
- Straub, D., and Papaioannou, I. (2014). “Bayesian Updating with Structural Reliability Methods”. *Journal of Engineering Mechanics*. Accepted for publication.

- Taflanidis, A. A., and Beck, J. L. (2009). "Stochastic Subset Optimization for reliability optimization and sensitivity analysis in system design". *Computers and Structures*, **87**(5-6), 318-331.
- Vanik, M. W., Beck, J. L., and Au, S. K. (2000). "Bayesian probabilistic approach to structural health monitoring". *J. Eng. Mech.*, 126(7), 738-745.
- Wang, Y., Cao, Z., and Au, S. K. (2010). "Practical reliability analysis of slope stability by advanced Monte Carlo simulations in a spreadsheet". *Canadian Geotechnical Journal* 48(1):162-172.
- Wang, Q., Lu, Z. Z., and Zhou, C. (2011). "New topology optimization method for wing leading-edge ribs", *Journal of aircraft*, **48**(5), 1741-1748.
- Zuev, K. M., and Katafygiotis, L. S. (2011). "Modified Metropolis–Hastings algorithm with delayed rejection". *Probabilistic Engineering Mechanics*, **26**, 405–412.