# Bayesian updating and model class selection with Subset Simulation

F. A. DiazDelaO[a,*], A. Garbuno-Inigo[a], S. K. Au[a], I. Yoshida[b]

[a]*Institute for Risk and Uncertainty, School of Engineering, University of Liverpool Brownlow Hill, Liverpool L69 3GH, United Kingdom.*
[b]*Department of Urban and Civil Engineering, Tokyo City University. 1-28-1 Tamazutsumi Setagaya-ku, Tokyo 158-8557, Japan.*

---

## Abstract

Identifying the parameters of a model and rating competitive models based on measured data has been among the most important and 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 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 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 Subset Simulation. The formulation, called BUS (Bayesian Updating with Structural reliability methods), is based the standard rejection principle. Its theoretical correctness and efficiency requires the prudent choice of a multiplier, which has remained an open question. This paper presents a fundamental study of the multiplier and investigates its bias effect when it is not properly chosen. A revised formulation of BUS is proposed, which fundamentally resolves the problem such that Subset Simulation can be implemented without knowing the multiplier a priori. An automatic stopping condition is also provided. Examples are presented to illustrate the theory and applications.

*Keywords:* Bayesian inference, BUS, Subset Simulation, Markov Chain Monte Carlo, model

updating

---

## 1. 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, Richard, 1961, Jaynes and Bretthorst, 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 and Katafygiotis, 1998, Beck, 2010], which resolved a number of philosophically challenging issues of the inverse problem, such has the treatment of multiple sets of parameters giving the same model fit to the data, an issue known as *identifiability*.

Let $\Theta \in \mathbb{R}^n$ be a set of parameters of a model $\mathcal{M}$, based on which a probabilistic prediction of the data $\mathcal{D}$ can be formulated through the likelihood function $P(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M})$. Clearly, the probability distribution of $\Theta$ depends on the available information. Based only on knowledge in the context of $\mathcal{M}$, the distribution is described by the prior distribution $P(\boldsymbol{\theta}|\mathcal{M})$. 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

$$P(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}) = P(\mathcal{D}|\mathcal{M})^{-1}\, P(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M})\, P(\boldsymbol{\theta}|\mathcal{M}), \tag{1.1}$$

where

*Corresponding author

$$P(\mathcal{D}|\mathcal{M}) = \int_{\boldsymbol{\Theta}} P(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M})\, P(\boldsymbol{\theta}|\mathcal{M})\, d\boldsymbol{\theta}, \tag{1.2}$$

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

$$E[r(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M})] = \int r(\boldsymbol{\theta})\, P(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M})\, d\boldsymbol{\theta}. \tag{1.3}$$

As far as the posterior distribution of $\boldsymbol{\theta}$ for a given model $\mathcal{M}$ is concerned, the constant in Eq. (1.2) 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 $P(\mathcal{M})P(\mathcal{D}|\mathcal{M})$ [Carlin and Chib, 1995, Chen et al., 2012, Beck and Yuen, 2004]. In that context, $P(\mathcal{D}|\mathcal{M})$ 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 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 and 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.

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 and Beck, 1998, Katafygiotis and 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, 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 and Au, 2002, Cheung and Beck, 2010, Ching and 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 and Beck, 2003, Schuëller et al., 2004, Katafygiotis and 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 and Beck, 2001, Au and 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 and Hsieh, 2007, Song and Kang, 2009, Taflanidis and 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 and Au, 2010, Qi 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 SuS. 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 and Beck, 2003, Au, 2004, Au and 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 and 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. It offers no guarantee on correctness, however, due to the incomplete nature of finite sampling information which seems inevitable. The problem with the choice of the multiplier remains open.

This work is motivated by the choice of the multiplier and more fundamentally its mathematical and philosophical role in the BUS formulation. A rigorous mathematical study is carried out to provide fundamental understanding of the multiplier, which leads to a revised BUS 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 SuS 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.

## 2. 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.

### *2.1. 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 $Y = h(\boldsymbol{\theta})$, be a scalar response quantity of interest that depends on the set of uncertain parameters $\boldsymbol{\theta}$ distributed as the parameter probability density function (PDF) $q(\boldsymbol{\theta})$. The function $h(\cdot)$ represents the relationship between the uncertain input parameters and the output response. The parameter PDF $q(\cdot)$ 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 $b$:

$$P(Y > b) = \int q(\boldsymbol{\theta})\, I(\boldsymbol{\theta} \in F)\, d\boldsymbol{\theta}, \tag{2.1}$$

where

$$F = \{Y > b\} = \{\boldsymbol{\theta} \in \mathbb{R}^n : h(\boldsymbol{\theta}) > b\}, \tag{2.2}$$

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 $r(\boldsymbol{\theta})$ (say) conditional on the failure event, *i.e.*

$$E[r(\boldsymbol{\theta})|F] = \int r(\boldsymbol{\theta})\, q(\boldsymbol{\theta}|F)\, d\boldsymbol{\theta}, \tag{2.3}$$

where

$$q(\boldsymbol{\theta}|F) = P_F^{-1}\, q(\boldsymbol{\theta})\, I(\boldsymbol{\theta} \in F), \tag{2.4}$$

is the PDF of conditional on failure.

When the relationship between $Y$ and $\boldsymbol{\theta}$, *i.e.* the function $h(\cdot)$, is complicated, analytical or conventional numerical integration is not feasible for computing $P(Y > b)$ or $E[r(\boldsymbol{\theta})|F]$ and thus advanced computational methods are required for their efficient determination. SuS offers an efficient solution by generating a sequence of sample populations of $\boldsymbol{\theta}$ conditional on increasingly rare failure events $\{Y > b_i\}$, where $\{b_i : 1, 2, \ldots\}$ 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, 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.

### 2.2. Subset Simulation procedure

A typical SuS algorithm is presented as follows [Au and Beck, 2001, Au and 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 target failure probability.

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

$$p_k^{(0)} = \frac{N-k}{N}, \qquad k = 1, \ldots, N. \tag{2.5}$$

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.*

$$b_1 = b_{N(1-p_0)}^{(0)}. \tag{2.6}$$

By construction, the $p_0 N$ samples of $\boldsymbol{\theta}$ corresponding to $\{b_{N(1-p_0)+j}^{(0)} : j = 1, \ldots, p_0 N\}$ 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 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 $\{b_1^{(1)} : k = 1, \ldots, N\}$. The value $b_k^{(1)}$ gives the estimate of $b$ corresponding to exceedance probability $p_k^{(1)} = P(Y > b)$ where

$$p_k^{(1)} = p_0 \frac{N-k}{N}, \qquad k = 1, \ldots, N. \tag{2.7}$$

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.*

$$b_2 = b_{N(1-p_0)}^{(1)}. \tag{2.8}$$

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 $i$ $(i = 1, \ldots, N)$, in the ordered list of sample values of $Y$ denoted by $\{b_k^{(i)} : k = 1, \ldots, N\}$, the value $b_k^{(i)}$ gives the estimate of $b$ corresponding to exceedance probability $p_k^{(i)} = P(Y > b)$ where

$$p_k^{(i)} = p_0^i \frac{N-k}{N}, \qquad k = 1, \ldots, N. \tag{2.9}$$

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 and Beck, 2001, Schuëller 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, *i.e.* Papadopoulos et al. [2012], Zuev and Katafygiotis [2011], Bourinet et al. [2011]. See also the review in Section 5.9 of Au and 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].

## 3. BUS formulation

In this section we briefly review the BUS formulation [Straub and Papaioannou, 2014, 2016] 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 $p(\boldsymbol{\theta})$, $\mathcal{L}(\boldsymbol{\theta})$ to denote the likelihood function $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$, $P_{\mathcal{D}}$ to denote the normalizing constant $P(\mathcal{D}|\mathcal{M})$, and $p_{\mathcal{D}}(\boldsymbol{\theta})$ to denote the posterior PDF. 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 $p_{\mathcal{D}}(\boldsymbol{\theta})$ (rewritten from (1.1)):

$$p_{\mathcal{D}}(\boldsymbol{\theta}) = P_{\mathcal{D}}^{-1}\, q(\boldsymbol{\theta})\, \mathcal{L}(\boldsymbol{\theta}). \tag{3.1}$$

### *3.1. 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:

$$c\,\mathcal{L}(\boldsymbol{\theta}) \le 1. \tag{3.2}$$

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 $p_{\mathcal{D}}(\boldsymbol{\theta}) \propto q(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta})$ in (3.1) 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 $U < c\mathcal{L}(\boldsymbol{\theta})$, return $\boldsymbol{\theta}$ as the sample. Otherwise go back to Step 1.

It can be shown [Straub and Papaioannou, 2014] that the sample $\boldsymbol{\theta}$ returned from the above algorithm is distributed as $p_{\mathcal{D}}(\boldsymbol{\theta})$, that is by marginalising the auxiliary component $u$ as

$$p_{\boldsymbol{\theta}'}(\boldsymbol{\theta}) = \int_0^1 p_{\boldsymbol{\theta}',u}(\boldsymbol{\theta}, u)\, du \propto p_{\mathcal{D}}(\boldsymbol{\theta}). \tag{3.3}$$

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 sample drawn from the prior PDF $q(\boldsymbol{\theta})$ often has a low likelihood value $\mathcal{L}(\boldsymbol{\theta})$ when the data is informative about the uncertain parameters, leading to significant change from the prior to the posterior PDF.

### *3.2. 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 $(\boldsymbol{\theta}, u)$ having the joint PDF $q(\boldsymbol{\theta})\, I(0 \leq u \leq 1)$, where the *failure event* is defined as

$$F = \{U < c\,\mathcal{L}(\boldsymbol{\theta})\}. \tag{3.4}$$

Suppose that by some means (*e.g.* SuS) we can obtain a *failure sample* distributed as $q(\boldsymbol{\theta})\, I(0 \le u \le 1)$ and conditional on the failure event $F$. The PDF of the failure sample, denoted by $(\boldsymbol{\theta}', U')$, is given by

$$p_{\boldsymbol{\theta}',U'}(\boldsymbol{\theta}, u) = P_F^{-1}\, q(\boldsymbol{\theta})\, I(0 \le u \le 1)\, I(u < c\mathcal{L}(\boldsymbol{\theta})), \tag{3.5}$$

where

$$P_F = \int\int q(\boldsymbol{\theta})\, I(0 \le u \le 1)\, I(u < c\mathcal{L}(\boldsymbol{\theta}))\, du\, d\boldsymbol{\theta}, \tag{3.6}$$

is the *failure probability* of the reliability problem.

In the above formulation, the driving response variable can be defined as

$$Y = c\,\mathcal{L}(\boldsymbol{\theta}) - U, \tag{3.7}$$

so that the failure event corresponds to

$$F = \{Y > 0\}. \tag{3.8}$$

Populations of failure samples conditional on the intermediate failure events $F_i = \{Y > b_i\}$ for adaptively increasing $b_i$ $(i = 1, 2, \ldots)$ 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.

Note that in the original formulation the driving response variable was in fact defined $Y = U - c\,\mathcal{L}(\boldsymbol{\theta})$. The presentation in (3.7) is adopted so that it is consistent with the conventional SuS literature, where the intermediate threshold levels increase rather than decrease as the simulation level ascends.

## 4. Likelihood multiplier

One issue of concern in the BUS formulation is the choice of the multiplier $c$ satisfying the inequality in (3.2), which is not always trivial. Some suggestions were given, by inspecting the mathematical structure of the likelihood function [Straub and 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 rigorously investigate more fundamentally the role of the multiplier and its effect on the results if it is not properly chosen. The investigation leads to a reformulation of BUS, 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 $Y$ in (3.7). Clearly, the multiplier affects the distribution of the driving variable as well as the generated samples. Recall that only those samples conditional on $Y = c\,\mathcal{L}(\boldsymbol{\theta}) - U > 0$ 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.

From the inequality in (3.2), the choice of the multiplier is governed by the region in the parameter space of $\boldsymbol{\theta}$ where the value of $\mathcal{L}(\boldsymbol{\theta})$ is large. The largest admissible value of $c$ is given by

$$c_{\max} = \left[\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})\right]^{-1} . \qquad (4.1)$$

This result is well-known in the rejection sampling 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 $\mathcal{L}(\boldsymbol{\theta})$ and derive inequalities to propose a choice of $c$ that guarantees $c\mathcal{L}(\boldsymbol{\theta}) \leq 1$. In such cases, it is computationally beneficial to use that value. However, in general it is difficult by numerical means to have a choice of $c$ that guarantees the inequality.

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 (3.2) was used in establishing the third equality in (3.3). Suppose this inequality is violated, say, within some region $B$:

$$B = \{\boldsymbol{\theta} \in \mathbb{R}^n : c\mathcal{L}(\boldsymbol{\theta}) > 1\}. \tag{4.2}$$

Then for any $\boldsymbol{\theta} \in B$, $I(u < c\,\mathcal{L}(\boldsymbol{\theta})) = 1$ for $u \in (0,1)$ and so (3.3) implies

$$p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = P_F^{-1}\, q(\boldsymbol{\theta}) \int_0^1 I(u < c\mathcal{L}(\boldsymbol{\theta}))du = P_F^{-1} q(\boldsymbol{\theta}). \tag{4.3}$$

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 $p_{\mathcal{D}}(\boldsymbol{\theta})$ as in (3.3):

$$p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = P_F^{-1}\, q(\boldsymbol{\theta})\, c\mathcal{L}(\boldsymbol{\theta}) \propto p_{\mathcal{D}}(\boldsymbol{\theta}). \tag{4.4}$$

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 $q(\boldsymbol{\theta})$ is taken to be constant and so $p_{\mathcal{D}}(\boldsymbol{\theta}) \propto c\mathcal{L}(\boldsymbol{\theta})$. 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 $q(\boldsymbol{\theta})$.

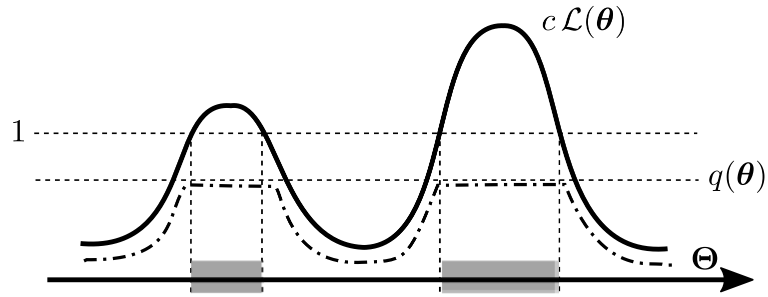


**Figure 1:** *Truncation of distribution in rejection algorithm. Center line - resulting distribution (short of the constant $P_F^{-1}$); shaded interval - truncation region $B$ where $c\mathcal{L}(\boldsymbol{\theta}) > 1$.*

As long as the multiplier satisfies the inequality in (3.2), 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.

## 5. 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 (3.4) can be rewritten as

$$F = \left\{ \ln \left[ \frac{\mathcal{L}(\boldsymbol{\Theta})}{U} \right] > -\ln c \right\}. \tag{5.1}$$

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

$$Y = \ln \left[ \frac{\mathcal{L}(\boldsymbol{\Theta})}{U} \right], \tag{5.2}$$

and the target failure event can now be written as

$$F = \{Y > b\}, \tag{5.3}$$

where

$$b = -\ln c. \tag{5.4}$$

The base of the logarithm is arbitrary but we choose to use natural logarithm here to facilitate the 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 (3.2), the distribution of the samples conditional on the failure event $F = \{U < c\mathcal{L}(\boldsymbol{\theta})\}$ 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 $Y = c\mathcal{L}(\boldsymbol{\theta}) - U$ in (5.4) 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 (5.4) into (3.2) and rearranging, the inequality constraint in terms of $b$ is given by, for all $\boldsymbol{\theta}$,

$$b > \ln \mathcal{L}(\boldsymbol{\theta}). \tag{5.5}$$

From (4.1), the maximum admissible value of $c$ is $c_{\max} = [\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})]^{-1}$. Correspondingly the minimum value of $b$ beyond which the distribution of samples will settle at the posterior PDF is

$$b_{\min} = -\ln c_{\max} = \ln\left[\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})\right]. \tag{5.6}$$

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 behaviour for $b > b_{\min}$ .

The logarithm in the above formulation is introduced for analytical and computational reasons, so that the driving variable is a well-defined random variable. In particular

$$Y = \ln\left[\frac{\mathcal{L}(\boldsymbol{\theta})}{U}\right] = \ln \mathcal{L}(\boldsymbol{\theta}) + \ln(U^{-1}). \tag{5.7}$$

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

$$\begin{aligned} E[Y] &= E[\ln \mathcal{L}(\boldsymbol{\theta}) + \ln U^{-1}] \\ &= E[\ln \mathcal{L}(\boldsymbol{\theta})] + 1, \end{aligned} \tag{5.8}$$

$$\begin{aligned} E[Y^2] &= E\{[\ln \mathcal{L}(\boldsymbol{\theta}) + \ln U^{-1}]^2\} \\ &= E\{[\ln \mathcal{L}(\boldsymbol{\theta})]^2\} + 2E[\ln \mathcal{L}(\boldsymbol{\theta})]E[\ln U^{-1}] + E\{[\ln U^{-1}]^2\} \\ &= E\{[\ln \mathcal{L}(\boldsymbol{\theta})]^2\} + 2E[\ln \mathcal{L}(\boldsymbol{\theta})] + 2, \end{aligned} \tag{5.9}$$

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

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.

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

1. *The distribution of $\boldsymbol{\theta}$ conditional on $\{Y > b\}$ is $p_{\mathcal{D}}(\boldsymbol{\theta}) = P_{\mathcal{D}}^{-1}\, q(\boldsymbol{\theta})\, \mathcal{L}(\boldsymbol{\theta})$ where $P_{\mathcal{D}} = \int q(z)\, \mathcal{L}(z)\, dz$ is a normalizing constant;*

2. *$P_{\mathcal{D}} = e^b\, P(Y > b)$.*

*Proof.* In order to prove the first part of the above theorem, first note that events $\{Y > b\}$ and $\{c\mathcal{L}(\boldsymbol{\theta}) > U\}$ are equivalent. Integrating out the uniform random variable from the PDF of the failure sample given by equation (3.5) gives:

$$\begin{aligned} p_{\boldsymbol{\theta}'}(\boldsymbol{\theta}) &= \int_0^1 p_{\boldsymbol{\theta}',U'}(\boldsymbol{\theta},u)\,du \\ &= p_F^{-1}\,q(\boldsymbol{\theta})\int_0^1 I(0\le u\le 1)\,I(u<c\,\mathcal{L}(\boldsymbol{\theta}))\,du \\ &= p_F^{-1}\,q(\boldsymbol{\theta})\,c\mathcal{L}(\boldsymbol{\theta}) \\ &\propto p_{\mathcal{D}}(\boldsymbol{\theta}). \end{aligned} \tag{5.10}$$

The result will be valid for any $c < [\max_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})]^{-1}$, or equivalently for any $b > \ln[\max_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})]$. For the second part of the theorem, since $Y = \ln[\mathcal{L}(\boldsymbol{\theta})/U]$ and $(\boldsymbol{\theta}, U)$ has a joint PDF $q(\boldsymbol{\theta})I(0 < u < 1)$, $P(Y > b)$ is given by

$$\begin{aligned} P(Y>b) &= \int\int q(\boldsymbol{\theta})\,I(0<u<1)\,I\left(\ln\left[\frac{\mathcal{L}(\boldsymbol{\theta})}{u}\right]>b\right)\,du\,d\boldsymbol{\theta} \\ &= \int q(\boldsymbol{\theta})\int_0^1 I(u<e^{-b}\mathcal{L}(\boldsymbol{\theta}))\,du\,d\boldsymbol{\theta} \\ &= e^{-b}\int q(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta})\,d\boldsymbol{\theta}, \end{aligned} \tag{5.11}$$

since $\int_0^1 I(u < e^{-b}\mathcal{L}(\boldsymbol{\theta}))\,du = e^{-b}\mathcal{L}(\boldsymbol{\theta})$ when $e^{-b}\mathcal{L}(\boldsymbol{\theta}) < 1$ for all $\boldsymbol{\theta}$ ($b$ is admissible). Observe, from the definition of the posterior (1.1), that $P_{\mathcal{D}}$ is simply the last integral in (5.10). Thus,

$$P_{\mathcal{D}} = e^b P(Y > b) \qquad\qquad b > b_{\min}. \tag{5.12}$$

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

■

## 6. Bayesian model class selection

In addition to providing the posterior distribution and estimating the updated expectation in (1.3), the posterior samples can be used for estimating the normalizing constant $P_{\mathcal{D}}$ in (1.2). 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 $p_{\mathcal{D}}(\boldsymbol{\theta})$. Consider the failure probability $P(Y > b)$, which can be estimated using the samples in SuS.

Note that equation (5.12) can be rewritten as

$$P(Y > b) = e^{-b} P_{\mathcal{D}} \qquad\qquad b > b_{\min}. \tag{6.1}$$

Since $P_{\mathcal{D}}$ 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 (5.2).

## 7. Characteristic trends

As shown in the last section, when $b > b_{\min}$ the failure probability $P(Y > b)$ is theoretically related to the evidence $P_{\mathcal{D}}$ through (5.12). 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 posterior samples. We argue that the variation of $P(Y > b)$ with $b$ takes on different characteristics on two different regimes of $b$. This can be used to tell whether the threshold value of a particular simulation level has already passed $b_{\min}$ in a SuS run, thereby suggesting a 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 $P_{\mathcal{D}}$ at $b > b_{\min}$. When $b > b_{\min}$, we know from (6.1) that $P(Y > b) = P_{\mathcal{D}} e^{-b}$ 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:

$$V(b) = b + \ln P(Y > b). \tag{7.1}$$

This function can be used for computing the log-evidence $\ln P_{\mathcal{D}}$ as it can be readily seen that

$$V(b) = \ln P_{\mathcal{D}} \qquad\qquad b > b_{\min}. \tag{7.2}$$

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 $\ln P_{\mathcal{D}}$ after $b > b_{\min}$. The characteristic behavior of $\ln P(Y > b)$ and $V(b)$ are depicted in Figure 2.

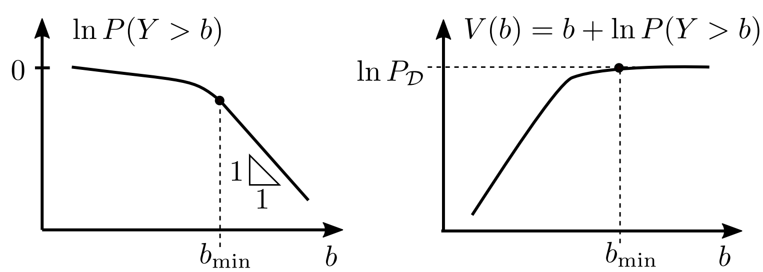


**Figure 2:** *Characteristic trends of* $\ln P(Y > b)$ *and* $V(b)$.

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, once the trainsition in the slope of $\ln P(Y > b)$ and $V(b)$ is complete. On this basis, we present an automatic stopping condition that is enforced once the algorithm detects that the transition has occurred.

## 8. Automatic Stopping Strategy

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.

The primary target of the Bayesian updating problem is to generate posterior samples of $\boldsymbol{\Theta}$ distributed as the posterior PDF $p_{\mathcal{D}}(\boldsymbol{\theta}) \propto q(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta})$, where $q(\boldsymbol{\theta})$ is the prior distribution assumed to be chosen from a standard class of distributions (e.g., Gaussian, exponential); and $\mathcal{L}(\boldsymbol{\theta})$ is the likelihood function for a given set of data. As reviewed in Section 2, 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 $Y = \ln[\mathcal{L}(\boldsymbol{\theta})/U]$, where $\boldsymbol{\theta}$ is distributed as $q(\boldsymbol{\theta})$ and $U$ is uniformly distributed on [0,1]; with $\boldsymbol{\theta}$ and $U$ independent. The conditional samples are collected from the level whose threshold level is determined to be greater than $b_{\min}$.

### *8.1. Stopping criterion*

From the discussion in Section 7 and the definition of SuS, it is clear that the intermediate failure levels will continue to increase as the algorithm progresses. For a given level $k$ where $b_k$ is an admissible value for the failure event, the samples generated will eventually be distributed as desired. The following theorem establishes theoretical guarantees that such failure level can be achieved in a finite number of iterations, given some regularity assumptions. Moreover, it provides a stopping criterion to terminate the algorithm and prevent the generation of unnecessary SuS levels.

**Theorem 2.** *Let the Bayesian inference problem be defined by an upper-bounded likelihood function $\mathcal{L}(\boldsymbol{\theta})$, a prior density $q(\boldsymbol{\theta})$ and associated posterior $p(\boldsymbol{\theta}|\mathcal{D})$. The marginal distribution of $\boldsymbol{\theta}$ conditional on the intermediate failure levels, denoted by $p(\boldsymbol{\theta}|F_k)$, converges to the posterior. Moreover, there exist constants $e^{-b_k}$ and a monotone decreasing sequence $a_k$, such that*

$$\lim_{k\to\infty} a_k = 0. \tag{8.1}$$

*where $a_k$ is the prior probability of the set $B_k = \{\boldsymbol{\theta} : e^{-b_k}\mathcal{L}(\boldsymbol{\theta}) > 1\}$.*

*Proof.* In Theorem 1, it was proved that as long as the $j$-th failure level satisfies $b_j > b_{\min}$, any sample generated will be distributed according to the target posterior distribution. The level $b_j$ is said to be a terminal level since any value of $b_{j+1}$ is, by definition, $b_{j+1} > b_j$. Hence, the samples will be distributed as desired for any terminal level.

To prove the theorem, let us characterise a non-terminal level $k$ such that $b_k < b_{\min}$. For the optimal threshold level $b_{\min}$, the inequality

$$u < e^{-b_{\min}}\mathcal{L}(\boldsymbol{\theta}) < 1, \tag{8.2}$$

is guaranteed for any value of a failure sample $(\boldsymbol{\theta}, u)$ being distributed jointly as equation (3.5). In contrast, a non-terminal level satisfies $e^{-b_{\min}}\mathcal{L}(\boldsymbol{\theta}) < e^{-b_k}\mathcal{L}(\boldsymbol{\theta})$ and it is not possible to determine an analogous right-hand side of inequality (8.2). Let the inadmissible set be defined as $B_k = \{\boldsymbol{\theta} : e^{-b_k}\mathcal{L}(\boldsymbol{\theta}) > 1\}$. It follows that the marginal distribution of the target variable is given by

$$p(\boldsymbol{\theta}|F_k) \propto \begin{cases} q(\boldsymbol{\theta}) & \text{if}\, \boldsymbol{\theta} \in B_k \\ \\ e^{-b_k}\, q(\boldsymbol{\theta})\, \mathcal{L}(\boldsymbol{\theta}) & \text{if}\, \boldsymbol{\theta} \in B_k^{\mathsf{c}}. \end{cases} \tag{8.3}$$

Note that for all samples in the inadmissible set $B_k$, the marginal is proportional to the prior distribution, whilst for the samples in the admissible set $B_k^{\mathsf{c}}$ the target density is proportional to the posterior distribution. Marginalising in order to compute the normalising constant results in

$$\begin{aligned} P_{F_k} &= \int_{\boldsymbol{\Theta}} \left[ q(\boldsymbol{\theta})\, I(\boldsymbol{\theta} \in B_k) + e^{-b_k}\, q(\boldsymbol{\theta})\, \mathcal{L}(\boldsymbol{\theta})\, I(\boldsymbol{\theta} \in B_k^{\mathsf{c}}) \right] d\boldsymbol{\theta} \\ &= \int_{B_k} q(\boldsymbol{\theta})\, d\boldsymbol{\theta} + e^{-b_k} \int_{B_k^{\mathsf{c}}} q(\boldsymbol{\theta})\, \mathcal{L}(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\ &= P_{\boldsymbol{\theta}}(B_k) + e^{-b_k}\, P_{\mathcal{D}}\, P_{\boldsymbol{\theta}|\mathcal{D}}(B_k^{\mathsf{c}}), \end{aligned} \tag{8.4}$$

where $P_{\boldsymbol{\theta}}(B_k)$ denotes the probability of event $B_k$ under the prior distribution and $P_{\boldsymbol{\theta}|\mathcal{D}}(B_k^{\mathsf{c}})$ denotes the probability of event $B_k^{\mathsf{c}}$ under the posterior distribution. Note that equation (8.4) is consistent with the case where $b_k$ is a terminal level. If that is the case, the pair $(\boldsymbol{\theta}, u)$ satisfies $u < e^{-b_k}\mathcal{L}(\boldsymbol{\theta})$ by the definition of the driving variable $Y$ and thus $B_k = \varnothing$. Let us rewrite the inadmissible set as

$$B_k = \{\boldsymbol{\theta} : \mathcal{L}(\boldsymbol{\theta}) > e^{b_k}\}. \tag{8.5}$$

Given an increasing sequence of failure levels, it can be seen that the sequence of inadmissible sets is monotone decreasing, namely

$$B_k \supset B_{k+1} \supset \ldots \supset \varnothing. \tag{8.6}$$

This fact is depicted in Figure 3.

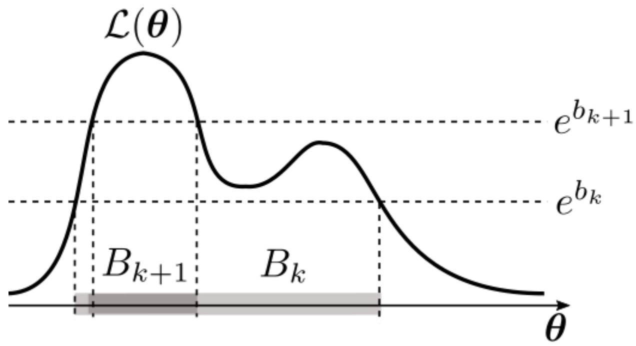


**Figure 3:** *Increasing failure levels and likelihood.*

Additionally, since the prior distribution is a probability measure, it satisfies the monotonicity property, namely $P(B_{k+1}) \leq P(B_k)$ for all $k$. Let us define the sequence $a_k$ as the prior probability of the inadmissible sets, *i.e.* $a_k = P_{\boldsymbol{\theta}}(B_k)$. As a consequence of the monotonicity property, it follows that $a_k$ is a monotone decreasing sequence of values converging to zero from above, denoted by

$$a_k \searrow 0. \tag{8.7}$$

Moreover, since the sets $B_k$ are monotone decreasing, then the sequence of complements is increasing, that is

$$B_k^{\mathsf{c}} \subset B_{k+1}^{\mathsf{c}} \subset \ldots \subset \Theta. \tag{8.8}$$

Let $m_k$ denote the posterior probability of the set $B_k^{\mathsf{c}}$. Analogous to $a_k$, the sequence $m_k$ is monotone increasing converging to 1 from below. This is denoted by

$$m_k \nearrow 1. \tag{8.9}$$

Expressions (8.7) and (8.9) allow to establish that for a sufficiently large value of $k$

$$p_{F_k} = e^{-b_k}\, P_{\mathcal{D}}, \tag{8.10}$$

is satisfied and the result is established.

■

The preceding theorem allows us to propose a stopping criterion for the BUS algorithm with driving variable $Y = \log[\mathcal{L}(\boldsymbol{\theta})/u]$ using SuS . The value of $a_k$ can be made arbitrarily small by means of the failure level $b_k$, which is learnt automatically during the algorithm. The computation of $a_k$ is challenging, since it involves a multiple integral. Note that the prior probability can be written as

$$a_k = P_{\boldsymbol{\theta}}(B_k) = P_{\boldsymbol{\theta}}(\mathcal{L}(\boldsymbol{\theta}) > e^{b_k}) \tag{8.11}$$

which is in itself a reliability problem, where the likelihood $\mathcal{L}(\boldsymbol{\theta})$ takes the role of a performance function and $e^{b_k}$ is a reliability threshold. Since the prior distributions are chosen from a standard catalogue of density functions and the probability is assumed to be small, such integral can also be computed by means of SuS. In this setting, computing equation (8.11) can be regarded as performing an *inner level* SuS. The sampling of the expanded variables $(\boldsymbol{\theta}, u)$ from the failure levels in equation (3.5), is regarded as *outer level* SuS.

### *8.2. Posterior statistical estimation*

The posterior samples $\{\boldsymbol{\theta}_k^{(m)} : k = 1, \ldots, N\}$ 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 (1.3) is estimated by simple averaging:

$$E[r(\boldsymbol{\theta})|\mathcal{D},\mathcal{M})] \approx \frac{1}{N}\sum_{k=1}^{N} r(\boldsymbol{\theta}_k^{(m)}). \tag{8.12}$$

On the other hand, based on (5.12), the evidence can be estimated by

$$P(\mathcal{D}|\mathcal{M}) = P_{\mathcal{D}} \approx \hat{P}_{\mathcal{D}} = e^{b_m} p_0^m. \tag{8.13}$$

Taking logarithm, the log-evidence is estimated by

$$\ln P(\mathcal{D}|\mathcal{M}) = \ln P_{\mathcal{D}} \approx \ln \hat{P}_{\mathcal{D}} = b_m + m \ln p_0. \tag{8.14}$$

*8.3. 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 $p_{\mathcal{D}}(\boldsymbol{\theta})$. 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 necessary. The stopping criterion based on the inner-outer procedure discussed above guards against this scenario.

For the evidence estimate in (8.13), 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 c.o.v.$(\ln P) \approx \text{std}(\ln P) \approx \text{std}(b_m)$, 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 $\hat{P}_b$ (say) for $P(Y > b)$ for fixed $b$ is available, rather than the c.o.v. of the $b$ quantile value $b_m$ for fixed exceedance probability. It can be reasoned, however, that the c.o.v. of $\hat{P}_{\mathcal{D}}$ (where $b_m$ is random) can be approximated by the c.o.v. of $e^b\hat{P}_b$ for fixed $b$ (then taking $b = b_m$ obtained in a simulation run). The latter is equal to the c.o.v. of $\hat{P}_b$, for which standard estimation formula is available [Au and Beck, 2001, Au and Wang, 2014].

### *8.4. Comparison with original BUS formulation*

Table 1 provides a comparison between the original 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.

| | BUS | Proposed |
|---|---|---|
| Driving variable | $Y = c\mathcal{L}(\boldsymbol{\theta}) - U$ for any $c < [\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})]^{-1}$ | $Y = \ln[\mathcal{L}(\boldsymbol{\theta})/U]$ |
| Target failure event | $F = \{Y > 0\}$ | $F = \{Y > b\}$ for any $b > \ln[\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})]$ |
| Evidence calculation | $P_{\mathcal{D}} = cP(Y > 0)$ | $P_{\mathcal{D}} = e^b P(Y > b)$ for any $b > \ln[\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})]$ |
| Stopping criterion | When threshold value of simulation level is equal to zero. | After inner-outer SuS procedure automatically determines that the threshold $b_{min}$ has been crossed by driving the sequence $a_k \searrow 0$. |

**Table 1:** *Comparison of original BUS and proposed reformulation. Note that the original definition of the driving variable in BUS is $Y = U - c\mathcal{L}(\boldsymbol{\theta})$. For consistency with SuS literature, it has been reexpresed as shown here.*

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 $Y = c\mathcal{L}(\boldsymbol{\theta}) - U$. For example, samples conditional on $Y > 0.1$ cannot be directly used. Since the threshold values $b_1, b_2, \ldots$ 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 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 $b > b_{\min}$ ($Y = \ln[\mathcal{L}(\boldsymbol{\theta})/U]$) can be taken as a posterior sample. The value of $b_{\min}$ is unknown but whether $b > b_{\min}$ can be determined from the inner-outer procedure discussed in Section 8.

## 9. 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.

*9.1. Example 1. 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 is to identify the interstory stiffnesses which allow the structural response to be subsequently updated. The first and second story masses are given by $16.5 \times 10^3$ kg and $16.1 \times 10^3$ kg respectively. Let $\boldsymbol{\theta} = [\theta_1, \theta_2]$ be the stiffness parameters to be identified. The interstory stiffnesses are thus parameterized as $k_1 = \theta_1 \overline{k}_1$ and $k_2 = \theta_2 \overline{k}_2$, where the nominal values for the stiffnesses are given by $k_1 = k_2 = 29.7 \times 10^6$ N/m. The joint prior distribution $q(\cdot)$ for $\theta_1$ and $\theta_2$ 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 $\mathcal{D} = \{\tilde{f}_1, \tilde{f}_2\}$ be the modal data used for the model updating, where 3.13 Hz and 9.83 Hz are the identified natural frequencies. The posterior PDF is formulated following Vanik et al. [2000] as

$$p_{\mathcal{D}} \propto \exp[-J(\boldsymbol{\theta})/2\epsilon^2]\, q(\boldsymbol{\theta}), \tag{9.1}$$

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

$$J(\boldsymbol{\theta}) = \sum_{j=1}^{2} \lambda_j^2\, [f_j^2(\boldsymbol{\theta})/\tilde{f}_j^2 - 1]. \tag{9.2}$$

Here, $\lambda_1$ and $\lambda_2$ 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.

Figure 4 shows the Markov chain samples for $\boldsymbol{\theta} = [\theta_1, \theta_2]$ at six consecutive simulation levels. The results are shown in 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.

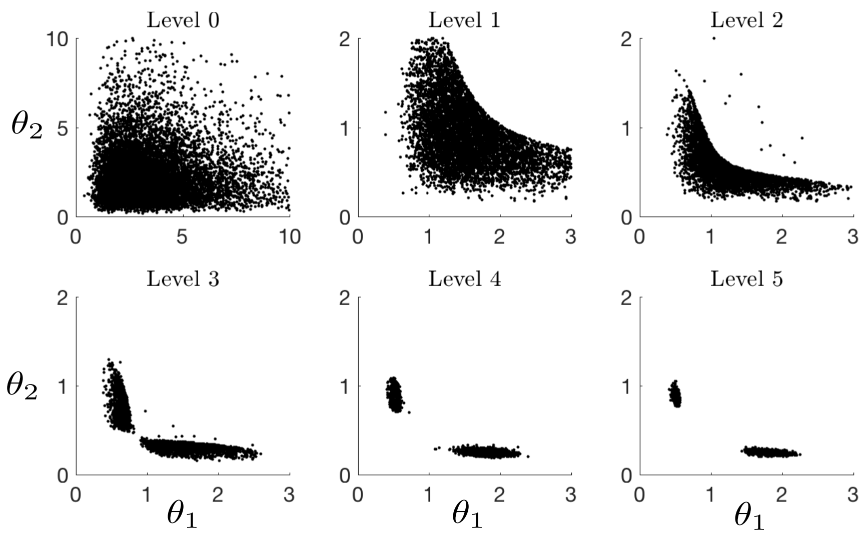


**Figure 4:** *Markov chain samples in the Lognormal space for the stiffness parameters $\theta = [\theta_1, \theta_2]$ from Level 0 (prior distribution) to Level 5.*

Figure 5 shows the marginal histograms for $\theta_1$ and $\theta_2$ corresponding to those samples in Figure 4. For comparison, the solid lines show the target marginal posterior distributions obtained by numerically integrating the expression for the posterior PDF, which is still feasible for this two-dimensional example. It is apparent that the distribution of the samples has settled either in Level 4 or Level 5. In reality, the exact target PDF is not available and so alternative means must be employed to determine whether the distribution of the samples has settled at the target one. Within the context of the current methodology, this is done through the proposed automatic stopping strategy and confirmed by the plots of the log-failure probability and log-evidence versus the threshold level.

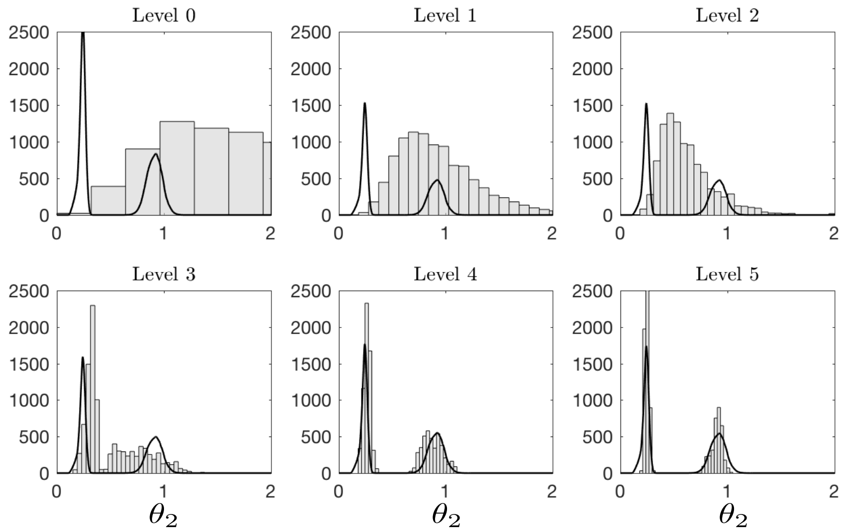


**Figure 5:** *Posterior marginal PDF for $\theta_2$ at different simulation levels. The target marginal posteriors were obtained numerically and are shown for comparison.*

Figure 6 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 (see Figure 2), 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 $b > 0$, the figure shows the boundaries of each level computed via SuS. Additionally, the log-evidence was computed following (7.1) and is shown in Figure 6. As with the log-CCDF, the theoretical prediction of the characteristic trend is also verified for this case, whereby the curve flattens when the transition is complete. Table 2 shows the evolution of the threshold (columns 2 and 3). The transition is complete after Level 4, where the probability of inadmissibility $a_k$ converges to zero (as defined in Section 8). For a tolerance of $a_k = 10^{-8}$, the fourth column in Table 2 shows that the posterior samples should be collected from Level 5. This corresponds with the clearly bimodal distributions in figures 4 and 5. It is guaranteed that the samples in the subsequent Sus levels would all be distributed according to the target posterior PDF. However, for statistical estimation their quality deteriorates as the simulation level ascends because their correlation tends to increase. Thus, the algorithm stops in Level 5.

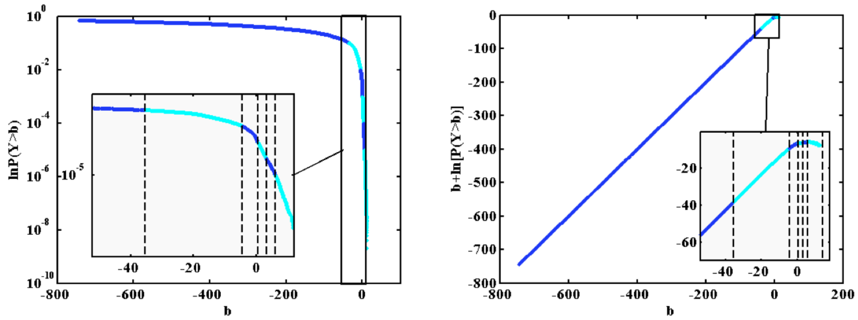


**Figure 6:** *Log-CCDF computed through SuS (left plot) for the identifiable case. The curve slowly transitions into a straight line with negative unit slope. Correspondingly, the log-evidence (right plot) flattens as the threshold exceeds $b_{\min}$. The dotted lines show the thresholds for different simulation levels.*

| Level | $b_k$ | $c_k$ | $a_k$ |
|---|---|---|---|
| 0 | | | |
| 1 | -4.291e+02 | 2.325e+186 | 5.3300e-01 |
| 2 | -6.237e+01 | 1.221e+27 | 1.3800e-01 |
| 3 | -9.331e+00 | 1.128e+04 | 2.8700e-02 |
| 4 | 2.203e+00 | 1.105e-01 | 4.0400e-03 |
| 5 | 5.780e+00 | 3.088e-03 | 0.0000e+00 |

**Table 2:** *Evolution of the threshold and the probability of inadmissibility.*

### *9.2. 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 $m_1 = \theta_3\overline{m}_1$ and $m_2 = \theta_4\overline{m}_2$, where the nominal values for the are given by $m_1 = 16.5 \times 10^3$ kg and $m_2 = 16.1 \times 10^3$ kg. Thus, for this case, $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3, \theta_4]$ where the marginal prior distributions for $\theta_1$ and $\theta_2$ are the same Lognormals as in the previous example. The prior marginal distributions for $\theta_3$ and $\theta_4$ 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 7 shows the Markov chain samples for $\boldsymbol{\theta}$ ($\theta_1$ versus $\theta_2$ for visualization purposes) at simulation levels 0 through 5. Again, the updated distribution results in a bimodal posterior PDF.

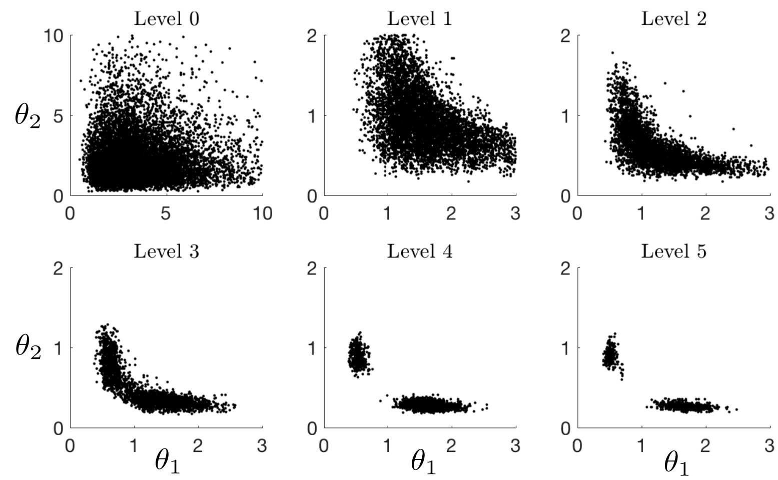


**Figure 7:** *Markov chain samples in the Lognormal space for the stiffness parameters $\theta_1$ and $\theta_2$ of the unidentifiable case at simulation levels 0 (prior distribution) to level 5.*

Analogously, Figure 8 shows the samples for $\theta_3$ and $\theta_4$ in the Lognormal space. There is no noticeable pattern in the distribution of the masses, consistent with the findings in Beck and Au [2002]. The characteristics of this example are very similar to the ones displayed by the locally identifiable case. The automatic stopping condition is also reached when $a_k \leq 10^{-8}$, for which the posterior samples are also collected in Level 5. We omit the characteristic trend plots and corresponding table for brevity.

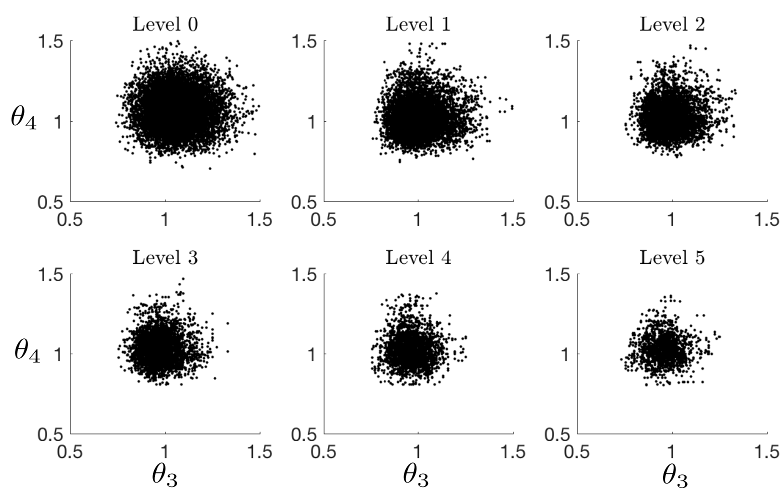


**Figure 8:** *Markov chain samples in the Lognormal space for the mass parameters $\theta_3$ and $\theta_4$ of the unidentifiable example at simulation levels 0 to level 5.*

### 9.3. Example 3. Model Class Selection

Following the two preceding examples, we can estimate the log-evidence corresponding to each model according to equation (8.14). Figure 9 shows the ratio of the evidence for the identifiable case to the evidence of the locally unidentifiable 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 no reason to prefer the unidentifiable model over the more parsimonious one.

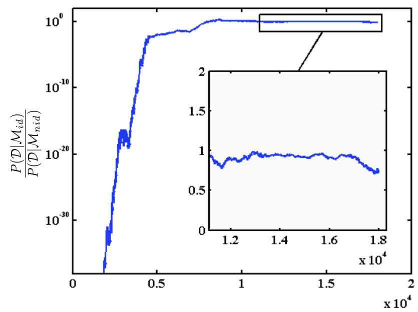


**Figure 9:** *Ratio of evidence of the identifiable model to the evidence of the locally unidentifiable model. Since this ratio converges to 1, there is no preference of either model over each other, given the available data.*

## 10. Conclusions

We have presented a fundamental analysis of BUS, a recently proposed framework that establishes an analogy between the Bayesian updating problem and the engineering 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, it was shown that the samples generated at different levels of SuS can be used directly as posterior samples as long as their threshold is greater than the minimum admissible value and the probability of inadmissibility is zero. We have proposed an inner-outer SuS procedure that provides an automatic stopping condition for the algorithm. The theoretical predictions of our study have been verified by applying our proposed strategy to illustrative examples.

## 11. Acknowledgements

This work was performed while the third 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. The second author gratefully acknowledges the Consejo Nacional de Ciencia y Tecnología (CONACYT) for the award of a scholarship from the Mexican government.

## References

S. Au. Probabilistic failure analysis by importance sampling Markov chain simulation. *Journal of Engineering Mechanics*, 130(3):303–311, 2004.

S. Au and Y. Wang. *Engineering Risk Assessment with Subset Simulation*. Wiley, 2014.

S. K. Au. Reliability-based design sensitivity by efficient simulation. *Computers & Structures*, 83 (14):1048–1061, 2005.

S.-K. Au and J. L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277, Oct. 2001.

S.-K. Au and J. L. Beck. Important sampling in high dimensions. *Structural Safety*, 25(2):139–163, Apr. 2003.

S. K. Au, Z. Cao, and Y. Wang. Implementing advanced Monte Carlo simulation under spreadsheet environment. *Structural Safety*, 32(5):281–292, 2010.

J. L. Beck. Bayesian system identification based on probability logic. *Structural Control and Health Monitoring*, 17(7):825–847, 2010.

J. L. Beck and S.-K. Au. Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation. *Journal of Engineering Mechanics*, 128(4):380–391, 2002.

J. L. Beck and L. S. Katafygiotis. Updating models and their uncertainties. I: Bayesian statistical framework. *Journal of Engineering Mechanics*, 124(4):455–461, 1998.

J. L. Beck and K.-V. Yuen. Model selection using response measurements: Bayesian probabilistic approach. *Journal of Engineering Mechanics*, 130(2):192–203, 2004.

W. Betz, I. Papaioannou, and D. Straub. Adaptive variant of the BUS approach to Bayesian updating. 9th International Conference on Structural Dynamics (EURODYN), Porto, Portugal, 2014.

J.-M. Bourinet, F. Deheeger, and M. Lemaire. Assessing small failure probabilities by combined subset simulation and support vector machines. *Structural Safety*, 33(6):343–353, 2011.

B. P. Carlin and S. Chib. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 473–484, 1995.

M.-H. Chen, Q.-M. Shao, and J. G. Ibrahim. *Monte Carlo methods in Bayesian computation.* Springer Science & Business Media, 2012.

S. H. Cheung and J. L. Beck. Calculation of posterior probabilities for Bayesian model class assessment and averaging from posterior samples based on dynamic system data. *Computer-Aided Civil and Infrastructure Engineering*, 25(5):304–321, Feb. 2010.

M. Chiachio, J. L. Beck, J. Chiachio, and G. Rus. Approximate Bayesian computation by subset simulation. *SIAM J. Sci. Comput.*, 36(3):1339–1358, 2014.

J. Ching and Y.-C. Chen. Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection and model averaging. *Journal of Engineering Mechanics*, 133(7): 816–832, 2007.

J. Ching and Y. Hsieh. Approximate reliability-based optimization using a three-step approach based on subset simulation. *Journal of Engineering Mechanics*, 133(4):481–493, 2007.

D. Ewins. *Modal testing*. Research Studies Press, Baldock, 2000.

G. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer, New York Berlin Heidelberg, 1996.

H. Haario, E. Saksman, and J. Tamminen. Componentwise adaptation for high dimensional MCMC. *Computational Statistics*, 20(2):265–273, 2005.

W. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 82:711–732, 1970.

D. E. Hudson. Dynamic tests of full-scale structures. *Journal of Engineering Mechanics Division*, 103(6):1141–1157, 1977.

E. Jaynes and G. Bretthorst. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

L. Katafygiotis and K. Zuev. Geometric insight into the challenges of solving high-dimensional reliability problems. *Probabilistic Engineering Mechanics*, 23(2-3):208–218, 2008.

L. S. Katafygiotis and J. L. Beck. Updating models and their uncertainties II: Model identifiability. *Journal of Engineering Mechanics, ASCE*, 124(4):463–467, 1998.

L. S. Katafygiotis and H.-F. Lam. Tangential-projection algorithm for manifold representation in unidentifiable model updating problems. *Earthquake engineering & structural dynamics*, 31(4): 791–812, 2002.

H.-S. Li and S.-K. Au. Design optimization using Subset Simulation algorithm. *Structural Safety*, 32 (6):384–392, 2010.

D. Malakoff. Bayes offers a 'new' way to make sense of numbers. *Science*, 286(5444):1460–1464, 1999.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Physics*, 21(6):1087–1092, 1953.

C. Papadimitriou, J. L. Beck, and L. S. Katafygiotis. Updating robust reliability using structural test data. *Probabilistic Engineering Mechanics*, 16(2):103–113, 2001.

V. Papadopoulos, D. G. Giovanis, N. D. Lagaros, and M. Papadrakakis. Accelerated subset simulation with neural networks for reliability analysis. *Computer Methods in Applied Mechanics and Engineering*, 223:70–80, 2012.

W. Qi, Z. Lu, and C. Zhou. New topology optimization method for wing leading-edge ribs. *Journal of Aircraft*, 48(5):1741–1748, sep 2011.

C. T. Richard. The algebra of probable inference, Johns Hopkins Unversity Press, 1961.

C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Science & Business Media, 2004.

G. Schuëller, H. Pradlwarter, and P. Koutsourelakis. A critical appraisal of reliability estimation procedures for high dimensions. *Probabilistic Engineering Mechanics*, 19(4):463–474, 2004.

J. Song and W.-H. Kang. System reliability and sensitivity under statistical dependence by matrix-based system reliability method. *Structural Safety*, 31(2):148 – 156, 2009.

D. Straub. Reliability updating with equality information. *Probabilistic Engineering Mechanics*, 26 (2):254–258, 2011.

D. Straub and I. Papaioannou. Bayesian Updating with Structural Reliability Methods. *Journal of Engineering Mechanics, Trans. ASCE*, 141(3), 2014.

D. Straub, I. Papaioannou, and W. Betz. Bayesian analysis of rare events. *Journal of Computational Physics*, 314:538 – 556, 2016.

A. A. Taflanidis and J. L. Beck. Life-cycle cost optimal design of passive dissipative devices. *Structural Safety*, 31(6):508 – 522, 2009. Optimization under Uncertainty with Emphasis on Structural Applications Special Issue: Optimization.

M. Vanik, J. Beck, and S.-K. Au. Bayesian probabilistic approach to structural health monitoring. *Journal of Engineering Mechanics*, 126:738–745, 2000.

Y. Wang, Z. Cao, and S.-K. Au. Practical reliability analysis of slope stability by advanced Monte Carlo simulations in a spreadsheet. *Canadian Geotechnical Journal*, 48(1):162–172, 2010.

K. M. Zuev and L. S. Katafygiotis. Modified Metropolis-Hastings algorithm with delayed rejection. *Probabilistic Engineering Mechanics*, 26(3):405–412, 2011.