Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 331 (2018) 72–93
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

Bayesian inference with Subset Simulation: Strategies and improvements

Wolfgang Betz[a,*], Iason Papaioannou[a], James L. Beck[b], Daniel Straub[a]

[a] *Engineering Risk Analysis Group, Technische Universität München, 80333 München, Germany*
[b] *Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, United States*

Received 10 May 2017; received in revised form 8 November 2017; accepted 13 November 2017
Available online 2 December 2017

Abstract

Bayesian Updating with Structural reliability methods (BUS) reinterprets the Bayesian updating problem as a structural reliability problem; i.e. a rare event estimation. The BUS approach can be considered an extension of rejection sampling, where a standard uniform random variable is added to the space of random variables. Each generated sample from this extended random variable space is accepted if the realization of the uniform random variable is smaller than the likelihood function scaled by a constant c. The constant c has to be selected such that $1/c$ is not smaller than the maximum of the likelihood function, which, however, is typically unknown a-priori. A c chosen too small will have negative impact on the efficiency of the BUS approach when combined with sampling-based reliability methods. For the combination of BUS with Subset Simulation, we propose an approach, termed aBUS, for *adaptive* BUS, that does not require c as input. The proposed algorithm requires only minimal modifications of standard BUS with Subset Simulation. We discuss why *aBUS* produces samples that follow the posterior distribution - even if $1/c$ is selected smaller than the maximum of the likelihood function. The performance of *aBUS* in terms of the computed evidence required for Bayesian model class selection and in terms of the produced posterior samples is assessed numerically for different example problems. The combination of BUS with Subset Simulation (and aBUS in particular) is well suited for problems with many uncertain parameters and for Bayesian updating of models where it is computationally demanding to evaluate the likelihood function.

Keywords: Bayesian updating; Bayesian model class selection; Subset Simulation; Structural reliability; MCMC; BUS

1. Introduction

Bayesian inference provides a consistent framework to reduce uncertainties in existing models with new information. Uncertainty is represented by a probability distribution over the model parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^M$. The information about $\boldsymbol{\theta}$ already acquired in the past is described by the *prior* distribution $p(\boldsymbol{\theta})$, which represents one's

* Corresponding author.
E-mail addresses: wolfgang.betz@tum.de (W. Betz), iason.papaioannou@tum.de (I. Papaioannou), jimbeck@caltech.edu (J.L. Beck), straub@tum.de (D. Straub).

https://doi.org/10.1016/j.cma.2017.11.021

initial belief on the parameters $\boldsymbol{\theta}$. Information/data $\mathbf{d}$ that becomes available in form of *measurements or observations* is embedded in the Bayesian analysis through the likelihood function $L(\boldsymbol{\theta}|\mathbf{d}) = p(\mathbf{d}|\boldsymbol{\theta})$, which comes from substituting $\mathbf{d}$ into a stochastic model that predicts what the data should be for given $\boldsymbol{\theta}$. The learning process in Bayesian inference is formalized through *Bayes' theorem* as:

$$p(\boldsymbol{\theta}|\mathbf{d}) = {c_{\mathrm{E}}}^{-1} \cdot L(\boldsymbol{\theta}|\mathbf{d}) \cdot p(\boldsymbol{\theta}) \tag{1}$$

where $p(\boldsymbol{\theta}|\mathbf{d})$ is the posterior distribution that represents the posterior state of knowledge about the uncertain parameter vector $\boldsymbol{\theta}$, and c_{E} is a normalizing scalar.

Except for some special cases, the posterior distribution cannot be derived analytically, and posterior samples are usually generated numerically. Markov chain Monte Carlo (MCMC) methods constitute a popular class of methods to sample from the posterior distribution [1,2]. One problem of MCMC methods is that after an initial burn-in phase, the samples may not yet have reached the stationary distribution of the Markov chain [3]. That is, finding an appropriate burn-in period in MCMC is often a non-trivial problem. Another issue is that standard MCMC algorithms usually cannot be applied efficiently for problems with many uncertain parameters. Some specialized MCMC algorithms [4–7] can cope with high dimensional problems, however, they require additional evaluations of the likelihood function or its gradient for each generated sample.

The constant c_{E} in Eq. (1) is a measure for the plausibility of the assumed model class [8]:

$$c_{\mathrm{E}} = \int_{\boldsymbol{\Theta}} L(\boldsymbol{\theta}|\mathbf{d}) \cdot p(\boldsymbol{\theta}) \,\mathrm{d}\boldsymbol{\theta} \tag{2}$$

c_{E} is referred to as the *evidence* [9]; alternatively it is also known as *marginal likelihood* or *integrated likelihood*. The evidence c_{E} is required for Bayesian model class selection and model averaging [9–11]. It is typically challenging to compute the evidence c_{E} because of the multi-dimensional integral in Eq. (2). If the system is globally identifiable [12,13], asymptotic approximations [9,14] can be applied. Otherwise, the evidence is usually estimated numerically. A review of techniques to compute c_{E} is given in Cheung and Beck [15]. Methods to generate posterior samples are not necessarily suitable to estimate the evidence, and vice versa.

A recently introduced framework for Bayesian updating, called BUS [16], converts the evaluation of posterior densities into an equivalent reliability problem. In structural reliability, probabilities of rare events are estimated [17–19]. By interpreting the Bayesian updating problem as a rare event estimation, existing structural reliability methods can be used to perform the Bayesian analysis. Moreover, an estimate for the evidence c_{E} is obtained as a by-product of BUS. The Subset Simulation (SuS) algorithm [20] is a structural reliability method that is well suited for BUS: (i) SuS can efficiently handle problems with many uncertain parameters; (ii) SuS can efficiently estimate very small probabilities that may arise within BUS. The use of SuS in BUS is referred to as *BUS–SuS* in the following.

A limitation of the original BUS is that prior to the analysis a constant c has to be selected [21]: On the one hand, c^{-1} should not be smaller than the maximum value that the likelihood function can take: $c^{-1} \geq L_{\max}$. On the other hand, selecting c^{-1} conservatively large decreases the efficiency of the method. Therefore, an appropriate choice of c is crucial. However, in many cases the maximum of the likelihood function is not known in advance. In some cases, probabilistic information on the optimal value of c can be derived as a function of the data size [16]. For the use of SuS within BUS, two strategies that avoid selecting c have been proposed recently. In [22] an adaptive strategy to learn the maximum of the likelihood function during the simulation is suggested. In [21] the equivalent structural reliability problem is redefined such that the stopping criterion of SuS depends on c^{-1}, but not the underlying limit-state function. In [23,24] an alternative strategy to BUS is presented, based on the concept of Approximate Bayesian Computation that also allows the use of SuS for Bayesian updating; it avoids the issue of selecting any constant like c but at the expense of getting only approximate posterior samples. The BUS approach is combined in [25] with an adaptive neural network surrogate model.

This contribution focuses on the application of SuS within BUS. We pick up and extend the idea originally proposed in [22], of learning the constant c^{-1} on the fly. The proposed algorithm is termed *aBUS* (as a substitute for adaptive BUS) and requires only minimal modifications of standard BUS with SuS. We also discuss why *aBUS* produces samples that follow the posterior distribution—even if c^{-1} is selected smaller than the maximum of the likelihood function.

The structure of the paper is as follows: In Section 2, we formally introduce the BUS approach. In Section 3, the combination of BUS and SuS is explained in-depth. The proposed algorithm *aBUS* that adaptively learns the value of c^{-1} is introduced in Section 4. In Section 5, illustrative applications are presented to demonstrate the efficiency of the proposed method numerically using different examples. Section 6 briefly summarizes the obtained findings.