Proc. of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Jinsong Huang, D.V. Griffiths, Shui-Hua Jiang, Anna Giacomini and Richard Kelly

doi: 10.3850/978-981-18-5182-7_07-010-cd

ISGSR 2022

Bayesian Gaussian Mixture Model Learning with Subset Simulation

Guanting Zeng[1], Iason Papaioannou[1] and Zijun Cao[2]

[1]Engineering Risk Analysis Group, Technische Universität München, Arcisstr. 21, 80290 München, Germany.
E-mail: {guanting.zeng,iason.papaioannou}@tum.de
[2]MOE Key Laboratory of High-Speed Railway Engineering, Institute of Smart City and Intelligent Transportation, Southwest Jiaotong University, Chengdu 611756, P. R. China; E-mail: zijuncao@swjtu.edu.cn

Abstract: The Gaussian mixture model (GMM) provides a convenient and flexible means for probabilistic modelling of geotechnical parameters. We employ the Bayesian approach for fitting GMMs to data, which enables quantifying the uncertainty of GMM parameter estimates and selecting the number of components in the mixture through comparing the marginal likelihood of the data (aka evidence) for each model. To solve the Bayesian updating problem and estimate the marginal likelihood, we develop a variant of the adaptive BUS (Bayesian Updating with Structural reliability methods) algorithm that effectively explores all modes of the posterior distribution. The latter is achieved through the implementation of an adaptive Markov chain Monte Carlo sampler within subset simulation-based BUS, termed elliptical slice sampler. We demonstrate the effectiveness of the proposed method with a univariate geotechnical dataset taken from the ISSMGE TC304 database.

Keywords: Gaussian mixture model; Bayesian learning; model selection; aBUS; elliptical slice sampling.

1 Introduction

Probabilistic modeling of soil properties requires the choice of a modeling approach to construct their probability density function. Often geotechnical datasets reveal a multimodal behavior, which cannot be accurately captured by standard parametric model families, e.g., the widely used Gaussian or lognormal distribution models. The Gaussian mixture model (GMM), defined as the weighted sum of Gaussian distributions, is a flexible distribution model that can be used to model arbitrary distributions. Fitting of GMM based on data is often performed through maximum likelihood estimation (MLE), e.g., using the expectation-maximization algorithm. Alternatively, the Bayesian approach can be applied, which generates point estimates as well as the full posterior distribution of the GMM parameters. The Bayesian approach enables quantifying the uncertainty in the parameter estimates and avoids overfitting, which is inherent to MLE approaches. Moreover, it provides a consistent means for model selection, i.e., selection of the number of components in the mixture, through comparing estimates of the marginal likelihood of the data for each model.

In this work, we develop a method for Bayesian learning of GMM that is based on adaptive BUS (Bayesian Updating with Structural reliability methods), originally developed in (Betz et al. 2018). The method leverages subset simulation (SuS) to explore the multimodal posterior distribution of the GMM parameters and estimate the marginal likelihood. To enable exploration of all modes of the posterior distribution, we implement an adaptive Markov chain Monte Carlo (MCMC) sampler within SuS, termed elliptical slice sampler (Murray et al. 2010). We compare the proposed method to the standard adaptive BUS approach for GMM learning with a univariate geotechnical dataset taken from the ISSMGE TC304 database.

2 Bayesian learning of GMM

The probability density function (PDF) of the GMM can be written as a weighted linear combination of K Gaussian components (McLachlan and Peel 2000)

$$p(x|\boldsymbol{\omega}) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \sigma_k), \tag{1}$$

where $N(x|\mu_k, \sigma_k)$ is the Gaussian PDF of the k-th component with mean value μ_k and standard deviation σ_k; $\pi_k \in [0,1]$ is the weighting coefficient of the k-th component and $\sum_{k=1}^{K} \pi_k = 1$.

Let $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$ denote a set of independent observations of the geotechnical property of interest. Learning the parameters of the GMM model of Eq. (1) requires the selection of the number of components K and the parameters of the corresponding model M_K, $\boldsymbol{\omega}_K = \{\boldsymbol{\pi}_K, \boldsymbol{\mu}_K, \boldsymbol{\sigma}_K\}$ with $\boldsymbol{\pi}_K = \{\pi_1, \dots, \pi_K\}$, $\boldsymbol{\mu}_K = \{\mu_1, \dots, \mu_K\}$ and $\boldsymbol{\sigma}_K = \{\sigma_1, \dots, \sigma_K\}$. The posterior distribution of the parameters $\boldsymbol{\omega}_K$ given the data $\mathbf{X}$ and model M_K is given by Bayes' rule as

$$p(\boldsymbol{\omega}_K|\mathbf{X}, M_K) = \frac{p(\mathbf{X}|\boldsymbol{\omega}_K, M_K)p(\boldsymbol{\omega}_K|M_K)}{P(\mathbf{X}|M_K)}, \tag{2}$$

where $p(\boldsymbol{\omega}_K|M_K)$ is the prior PDF of $\boldsymbol{\omega}_K$ given M_K, $p(\mathbf{X}|\boldsymbol{\omega}_K, M_K)$ is the likelihood function describing the data $\mathbf{X}$ and given by

$$p(\mathbf{X}|\boldsymbol{\omega}_K, M_K) = \prod_{i=1}^{N} p(x_i|\boldsymbol{\omega}_K, M_K), \tag{3}$$

and $P(\mathbf{X}|M_K)$ is the marginal likelihood (aka model evidence). $P(\mathbf{X}|M_K)$ expresses the likelihood of model M_K given the data. The probability of model M_K given the data $\mathbf{X}$ can be obtained through Bayes' rule as

$$P(M_K|\mathbf{X}) = \frac{P(\mathbf{X}|M_K)\,P(M_K)}{P(\mathbf{X})}, \tag{4}$$

where $P(M_K)$ is the prior probability and $P(\mathbf{X}) = \sum_{i=1}^{K_{\max}} P(\mathbf{X}|M_i)\,P(M_i)$. The quantities $P(M_K|\mathbf{X})$ are used to guide the selection of the most appropriate model given a set of candidate models as the one that maximizes $P(M_K|\mathbf{X})$ with respect to $K = 1, \ldots, K_{\max}$ (Fruhwirth-Schnatter et al. 2019). In the absence of prior knowledge on the probability of each model, a reasonable choice of the prior probabilities is $P(M_K) = 1/K_{\max}$, in which case maximizing $P(M_K|\mathbf{X})$ becomes equivalent to maximizing the model evidence $P(\mathbf{X}|M_i)$.

Learning GMM is a fundamentally unidentifiable problem due to the so-called label switching issue (Fruhwirth-Schnatter et al. 2019, Deng et al. 2022). This unidentifiability is due to the fact that the same set of mixture components with a different permutation of the labels can be combined to give the same mixture distribution. That is, there are $K!$ equivalent ways of arranging the components to obtain the same mixture PDF. This implies that the likelihood function of Eq. (3) and, hence, the posterior distribution $p(\boldsymbol{\omega}_K|\mathbf{X}, M_K)$, have $K!$ modes. This poses challenges in sampling from $p(\boldsymbol{\omega}_K|\mathbf{X}, M_K)$ and estimating the marginal likelihood $P(\mathbf{X}|M_K)$. Deng et al. (2022) applied a random Gibbs sampling approach combined with bridge sampling to effectively populate the multimodal posterior distribution and estimate the marginal likelihood. In this paper, we extend the BUS approach to effectively treat this problem.

3 ESS-based adaptive BUS

The BUS approach, originally proposed by Straub and Papaioannou (2015), is based on an interpretation of the rejection sampling algorithm for Bayesian updating as a structural reliability problem. Consider the Bayesian updating problem of Eq. (2). BUS extends the space of uncertain variables through introducing an additional standard uniform random variable v, and defines the limit-state function

$$h(\boldsymbol{\omega}_K, v) = v - c\,p(\mathbf{X}|\boldsymbol{\omega}_K, M_K), \tag{5}$$

where c is a constant that satisfies $c\,p(\mathbf{X}|\boldsymbol{\omega}_K, M_K) \leq 1$ for all $\boldsymbol{\omega}_K$ in the support of $p(\boldsymbol{\omega}_K|M_K)$. It can be shown that the posterior distribution can be retrieved by censoring the prior distribution $p(\boldsymbol{\omega}_K|M_K)$ on the domain $\{h(\boldsymbol{\omega}_K, v) \leq 0\}$ (Straub and Papaioannou 2015). Therefore, samples from the posterior distribution $p(\boldsymbol{\omega}_K|\mathbf{X}, M_K)$ can be obtained through the failure samples of the structural reliability problem for evaluating the probability $p_a = P(h(\boldsymbol{\omega}_K, v) \leq 0)$. The marginal likelihood can be estimated based on an estimate $\hat{p}_a$ of p_a as

$$\hat{P}(\mathbf{X}|M_K) = \frac{\hat{p}_a}{c}\,. \tag{6}$$

Straub and Papaioannou (2015) applied SuS (Au and Beck 2001) to efficiently solve the structural reliability problem in BUS. SuS defines a sequence of intermediate nested failure events and generates samples conditional on these events through application of MCMC algorithms (Papaioannou et al. 2015). Betz et al. (2018) proposed an adaptive version of SuS-based BUS, which adaptively estimates the constant c in Eq. (5) as the reciprocal of the maximum over the likelihood function values for all samples.

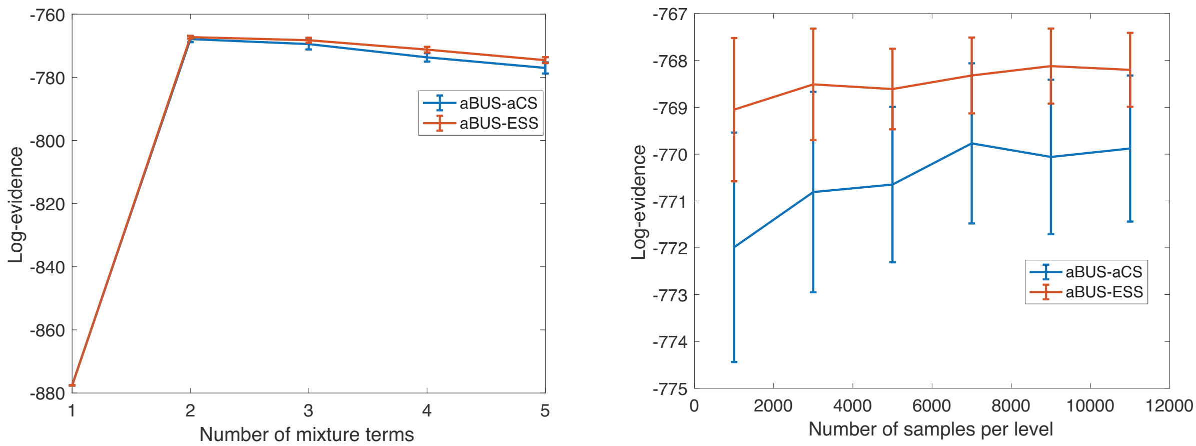
Betz et al. (2018) employed an adaptive conditional sampling (aCS) algorithm proposed in (Papaioannou et al. 2015) for the MCMC step in SuS-based BUS. The aCS algorithm efficiently samples high-dimensional conditional target distributions through expressing the target in terms of an underlying standard Gaussian distribution. aCS is a random walk algorithm and, hence, is not able to switch between modes when sampling multimodal target distributions. To enable effective sampling from the multimodal posterior distribution of Eq. (2), we implement an adaptive MCMC algorithm termed elliptical slice sampler (ESS) within SuS-based BUS. The basic idea of ESS, originally developed by Murray et al. (2010), is to generate each candidate sample from an ellipse on a two-dimensional plane that passes from the current sample. Samples are generated iteratively through a shrinking rule until a candidate sample falls in the target intermediate failure domain. More details on the ESS algorithm can be found in (Murray et al. 2010). Through generating candidates on an ellipse, ESS is more likely to explore different areas and, hence, travel between the different modes of the target distribution as compared to the aCS algorithm.

4 Numerical example

The sensitivity (S_t) data from the F-CLAY/7/216 dataset in TC304dB is used to illustrate the proposed method. We calculate the evidence in Eq. (2) for $K = 1, \dots, 5$ and compare the results obtained with the adaptive BUS with aCS (aBUS-aCS) and ESS (aBUS-ESS). For the prior $p(\boldsymbol{\omega}_K | M_K)$ of each considered model, we choose a weakly informative prior, whose parametric form is discussed in (Deng et al. 2022). To evaluate the performance of the two algorithms, we perform 100 independent runs. For both samplers the intermediate probability for SuS is chosen as 0.1.

Figure 1(a) shows the statistics of the evidence estimates obtained with the two samplers for different number of mixture terms K. For ease of readability, the standard deviations of the log-evidence estimates are given in Table 1. We see that the standard deviation of the estimates obtained with both samplers increases with increase of the number of components. This is due to the fact that, as K increases, the number of modes of the posterior distribution increases factorially. The standard deviation of aBUS-ESS is significantly smaller than that of aBUS-aCS, which is due to the ability of ESS to switch between modes during the MCMC sampling process. Figure 1(b) plots the statistics of the estimates for $K = 3$ and increasing number of samples per level. The standard deviation of the log-evidence estimates decreases with increase of the number of samples per level for both methods. For low number of samples, the methods appear to underestimate the log-evidence, which is likely related to the strongly skewed distribution of the SuS estimate of the probability in Eq. (6). This underestimation is significantly stronger for aBUS-aCS than for aBUS-ESS. Additionally, it is found that the log-evidence from aBUS-aCS is smaller than that from aBUS-ESS, which is again attributed to the underestimation of the probability in Eq. (6).

We note that the ESS sampler increases considerably the number of likelihood evaluations as compared to the aCS sampler, as it requires several likelihood evaluations for each generated sample until the sample is accepted. However, for the specific application of Bayesian mixture model learning the likelihood function of Eq. (3) can be computed efficiently, hence, this additional cost is affordable.



(a) Log-evidence for different number of mixture terms with 10^4 samples per level

(b) Log-evidence for $K = 3$ mixture terms and different number of samples per level

Figure 1. Comparison of the statistics (mean and ± one standard deviation) of the log-evidence estimates using aBUS-aCS and aBUS-ESS.

Table 1 Standard deviation of the log-evidence estimates for different number of mixture terms obtained using the two samplers with 10^4 samples per level.

Sampler	$K = 1$	$K = 2$	$K = 3$	$K = 4$	$K = 5$
aBUS-aCS	0.06	0.97	1.70	1.34	1.71
aBUS-ESS	0.04	0.41	0.79	0.89	0.94

5 Conclusions

This paper presents an approach for Bayesian learning of Gaussian mixture models (GMM) that implements the elliptical slice sampler (ESS) within the adaptive BUS approach. ESS is able to switch between modes during the MCMC process in BUS and is therefore suitable for sampling the multimodal posterior of the GMM parameters. The ability of the algorithm to estimate the log-evidence for different number of terms in the mixture is tested with a geotechnical dataset. The results demonstrate that the proposed method results in significantly lower standard deviation of the log-evidence estimates compared to the original adaptive BUS algorithm.

Acknowledgments

The authors would like to thank the members of the TC304 Committee on Engineering Practice of Risk Assessment & Management of the International Society of Soil Mechanics and Geotechnical Engineering for developing the database 304dB used in this study and making it available for scientific inquiry.

References

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.

Betz, W., Papaioannou, I., Beck, J. L., and Straub, D. (2018). Bayesian inference with subset simulation: strategies and improvements. *Computer Methods in Applied Mechanics and Engineering*, 331, 72-93.

Deng, Q. X., He, J., Cao, Z. J., Papaioannou, I., Li, D. Q., and Phoon, K. K. (2022). Bayesian learning of Gaussian mixture model for calculating debris flow exceedance probability. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 16(1), 154-177.

Fruhwirth-Schnatter, S., Celeux, G., and Robert, C. P. (2019). *Handbook of Mixture Analysis*. CRC press, Boca Raton.

McLachlan, G., and Peel, D. (2000). *Finite Mixture Models*, Willey, New York.

Murray, I., Adams, R., and MacKay, D. (2010). Elliptical slice sampling. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (pp. 541-548).* JMLR Workshop and Conference Proceedings.

Papaioannou, I., Betz, W., Zwirglmaier, K., and Straub, D. (2015). MCMC algorithms for subset simulation. *Probabilistic Engineering Mechanics*, 41, 89-103.

Straub, D., and Papaioannou, I. (2015). Bayesian updating with structural reliability methods. *Journal of Engineering Mechanics*, 141(3).