A Review of Stochastic Sampling Methods for Bayesian Inference Problems

Adolphus Lye

Institute for Risk and Uncertainty, University of Liverpool, United Kingdom.
E-mail: adolphus.lye@liverpool.ac.uk

Alice Cicirello

Dynamics, Vibration and Uncertainty (DVU) Lab, University of Oxford, United Kingdom.
E-mail: alice.cicirello@eng.ox.ac.uk

Edoardo Patelli

Institute for Risk and Uncertainty, University of Liverpool, United Kingdom.
E-mail: epatelli@liverpool.ac.uk

This study was done with the aim to analyze and evaluate the strengths and limitations of the Markov Chain Monte-Carlo (MCMC), Transitional Markov Chain Monte-Carlo (TMCMC), and Sequential Monte-Carlo (SMC) sampling methods in the context of solving engineering design problems. For each of these methods discussed in this paper, a case example will also be presented in the form of simply toy-model problems to demonstrate its use and effectiveness in estimating parameters under uncertainty and comparing it with determined results. For the MCMC case example, a simple harmonic oscillator will be looked into to estimate the value of the spring constant, k. For the TMCMC case example, the problem will be extended into a coupled oscillator problem and the goal would be to estimate the values of two spring constants to which there is imprecise knowledge: κ and κ_{12}. Finally, for the SMC case example, a simple harmonic oscillator will be analyzed once again as a static linear system to estimate the spring constant, k. As such, this conference paper is also targeted at readers who are new to these methods and to provide succinct information in facilitating the understanding of the three sampling approaches.

Keywords: Bayesian Inference, Random Sampling, Estimation Methods, Markov Chain Monte-Carlo, Transitional Markov Chain Monte-Carlo, Sequential Monte-Carlo.

1. Introduction

In engineering design problems, one of the problems faced is the absence or lack of available physical data given that it wold be too costly to obtain new information. Under such circumstances, these new information would have to be obtained via machine-learning techniques and computer simulation of the physical system. From there, the data that is obtained from both physical and computer simulation experiments would be integrated to perform the necessary mathematical modelling or model-updating of the physical system for various purposes such as the estimation of an unknown parameter. One such framework under which such integration and update of information can be done would be the Bayesian inference method.

The Bayesian inference method originates from the Bayes' Rule in statistics which is defined as such according to Beech et al. (1959):

$$P(\boldsymbol{\theta}|D) = \frac{P(D|\boldsymbol{\theta}) \cdot P(\boldsymbol{\theta})}{P(D)} \tag{1}$$

In Eq. (1), $\boldsymbol{\theta}$ represents the vector of parameters of interest, D represents the observed data, $P(\boldsymbol{\theta})$ represents the prior probability distribution of $\boldsymbol{\theta}$, $P(D|\boldsymbol{\theta})$ represents the likelihood probability distribution, $P(\boldsymbol{\theta}|D)$ represents the posterior (or target) probability distribution, and $P(D)$ is the evidence probability distribution.

The prior distribution, $P(\boldsymbol{\theta})$, can be described as a distribution which reflects what is known about the unknown parameters, $\boldsymbol{\theta}$, before any data is collected. For instance, a Uniform distribution is chosen as the prior distribution when the only bounds of values of $\boldsymbol{\theta}$ is known. The likelihood distribution, $P(D|\boldsymbol{\theta})$, reflects the distribution of the measured quantity, D, which depends on the value of $\boldsymbol{\theta}$. This is usually known through data collection and it illustrates the degree to which the parameters D and $\boldsymbol{\theta}$ agree. The posterior distribution, $P(\boldsymbol{\theta}|D)$, represents the updated distribution of $\boldsymbol{\theta}$ after taking into account the values of D. And finally, the evidence distribution, $P(D)$ can be simply described as the normalization constant to ensure the posterior probability sums to one.

In most cases, the constant $P(D)$ may be dif-

Proceedings of the 29th European Safety and Reliability Conference.
Edited by Michael Beer and Enrico Zio

ISBN: 981-973-0000-00-0 :: doi: 10.3850/981-973-0000-00-0_output

ficult to compute analytically and is usually not accounted for in solving Bayesian inference problems. As such, Eq. (1) can be simplified to the following form:

$$P(\boldsymbol{\theta}|D) \propto P(D|\boldsymbol{\theta}) \cdot P(\boldsymbol{\theta}) \tag{2}$$

This conference paper seeks to provide a review of three different sampling methods adopted in addressing Bayesian inference problems, namely Markov Chain Monte-Carlo (MCMC) sampling, Transitional Markov Chain Monte-Carlo (TMCMC) sampling, and Sequential Monte-Carlo (SMC) sampling. For each of these methods, a brief description and explanation of its concepts will be provided. This will then be followed by an illustration in the form of a simple case example to show how they can be applied to solve an engineering Bayesian inference problem. The results will be compared to that obtained analytically highlight its effectiveness. After which, an overall evaluation of the sampling method will be discussed before finally concluding the paper.

2. Markov chain monte-carlo sampling

MCMC is a computerized sampling method which aims to sample from a probability distribution as highlighted by Gamerman and Lopes (2006). Its advantage comes as a result of being able to identify the characteristics of an unknown distribution from which it samples without the need to know all of the mathematical properties of that distribution. As such, this makes MCMC a more favoured approach in extracting information on a distribution of interest, thereby making such technique useful in solving Bayesian inference problems according to Ravenzwaaij et al. (2018).

There are two underlying key concepts behind the workings of MCMC sampling: Monte-Carlo and Markov chain.

Monte-Carlo, as stated by Robert and Casella (2013), involves the practice of taking repeated samples in order to estimate a property of a distribution of interest. As such, it provides an alternative method to solve deterministic problems via randomness. Markov chain provides a sequential manner in which the samples will be drawn from the target distribution. Through this process, the random sample which is drawn at present will be used as a basis on which the next random sample will be drawn. This eventually forms a chain which contains a special property on its own in which each new random sample depends only on the previous random sample and no earlier. Such property is referred to, by Serfozo (2014), as the "Markov" property.

A straight-forward MCMC approach would be through the use of the Metropolis algorithm which provides a selection criteria of the samples chosen to reflect the posterior distribution. It involves the use of a symmetric proposal distribution, such as the Normal distribution, to choose and select a new sample based on the target distribution. A summary of the workings of the Metropolis algorithm is as follows:

- Step 1: Starting from a random sample, θ_1, the symmetric proposal distribution would then select the next random sample, θ_2. For example, if a Normal distribution is used as the proposal distribution, θ_2 will be chosen randomly from that distribution with mean θ_1 and a defined value of standard deviation assigned by the user.
- Step 2: Upon choosing θ_2, the probability of θ_2 with respect to the target distribution, $P(\theta_2|D)$, is compared to that for θ_1, $P(\theta_1|D)$. This is done by taking the ratio between these two quantities giving rise to α. In essence:

$$\alpha = \frac{P(\theta_2|D)}{P(\theta_1|D)} \tag{3}$$

- Step 3: A random number, r, is drawn from a Uniform distribution ranging between 0 and 1. If the value of α is greater than r, θ_2 will be accepted as the new sample and the process repeats from Step 1. Otherwise, θ_2 will be rejected and the process repeats from Step 1 using θ_1 again.

Full details can be found in the reference by Smith and Roberts (1993) and the approach is shown through an illustrative example in the next section.

2.1. *MCMC case example: Simple harmonic oscillator*

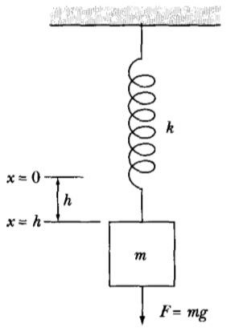

Fig. 1. Illustration of the set-up for the simple harmonic oscillator. Image taken from Thorton and Marion (2004)

A simple harmonic oscillator system illustrated in Figure 1 is analyzed with a determined value of the spring constant, $k = 0.6\ Nm^{-1}$, and mass, $m = 0.5\ kg$. In this problem, the parameter k is

not known precisely and objective of this analysis is to quantify and determine the precision of k.

The prior distribution for k is set to follow a Uniform distribution between $0.01\ Nm^{-1}$ and $3.00\ Nm^{-1}$. The likelihood distribution of the measured quantity, which in this case would be the frequency of the oscillation, $\omega(k)$, which is defined as:

$$\omega(k) = \sqrt{\frac{k}{m}} \tag{4}$$

The measured frequency is set to follow a Normal distribution with mean $\omega(k)$, which depends on the choice of values of k, and a standard deviation of $0.05\ Hz$, which is usually known from taking repeated measurements.

According to the Law of Large Number, it states that the sample mean will converge to the expected value with increased sampling. As such, the number of samples to be taken from this MCMC problem, N, will be set at 10000. The results will be presented in the form of a histogram as shown in Figure 2.

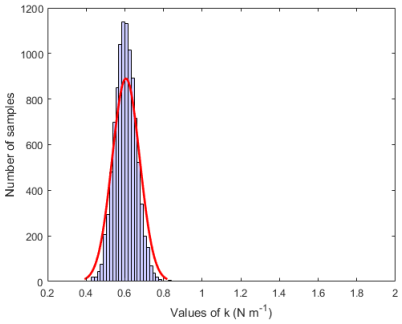

Fig. 2. Histogram illustrating the results from MCMC obtained for $N = 10000$.

Table 1. Results of the distribution function fitting of the histogram.

Distribution	Parameter(s) (Nm^{-1})
Normal	Mean: 0.604188 Stdev: 0.0702912

From Table 1, it can be seen that the histogram appears to follow a Normal distribution with a mean value of $0.604188\ Nm^{-1}$ which is a 0.698% discrepancy from the determined value of k. This indicates a high degree of accuracy associated with the estimated results obtained from MCMC. In addition to this, the value of standard deviation associated with the Normal distribution fitting is at $0.0702912\ Nm^{-1}$ which indicates that the degree of imprecision of the results is at 11.6%.

2.2. *Limitation(s) of MCMC*

As seen from the case example, the MCMC method is an effective tool to draw samples from target distributions which are not well-defined and had to be approximated using a prior and likelihood function. However, complications will arise when sampling directly from a target distribution with a distinct, sharp peak. This is due to the sample acceptance criteria within the MCMC algorithm which causes the samples to crowd around the peak making it ineffective in obtaining random samples from across the distribution. Such problem is made worse if there are multiple of such peaks which leads to a problem whereby the random samples would gather around one of those peaks and miss the other regions. Furthermore, Ching and Chen (2007) added that it will also become inefficient and computationally expensive in solving problems in which more than one parameters are to be estimated. To address this issue, the method of TMCMC will be used. Details will be provided in the next section.

3. Transitional markov chain monte-carlo sampling

TMCMC is a method devised by Ching and Chen (2007) and is inspired from the Adaptive Metropolis-Hastings (AMH) algorithm proposed by Beck and Au (2002). It adopts the concept of obtaining random samples from a series of relatively simpler "transitional" probability distributions instead of obtaining them directly from complex distributions.

According to Ching and Chen (2007), one universal way in devising such intermediate probability distribution stems from the proportionality relation in Eq. (2) and is as shown below:

$$P_j(\boldsymbol{\theta}|D) \propto P(D|\boldsymbol{\theta})^{\beta_j} \cdot P(\boldsymbol{\theta}) \tag{5}$$

In Eq. (5), j is the stage number taking integers values from 0 to m, m denotes the iteration number in updating the transition probability distribution, and β_j takes values such that $\beta_0 = 0 < \beta_1 < \ldots < \beta_{m-1} < \beta_m = 1$. This condition allows for the transition probability distribution to the take the form starting from that of prior and eventually converge to that of the posterior at the end of all the iterations.

The workings of the TMCMC algorithm can be summarized as such:

- Step 1: Random samples are obtained via

MCMC from the prior distribution (Iteration $j = 0$).
- Step 2: Add a small increment to β_j such that the transition from P_j to P_{j+1} is gradual and smooth.
- Step 3: At each iteration j, use the random samples drawn from P_j as the basis for the Metropolis algorithm to draw samples from P_{j+1} in the next iteration $j + 1$.
- Step 4: Repeat Steps (2) and (3) until iteration $j = m$ when $\beta_j = 1$.

To provide a better understanding of its applications, an illustrative example is provided in section 3.1 to demonstrate the use of TMCMC sampling method.

3.1. *TMCMC case example: Coupled oscillator*

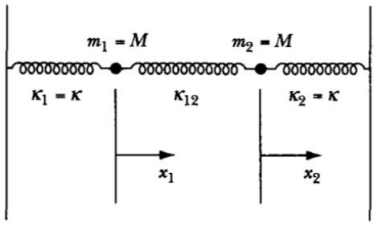

Fig. 3. Illustration of the set-up for the coupled oscillator. Image taken from Thorton and Marion (2004)

For this case example, the problem is extended from a simple harmonic oscillator to a coupled oscillator as shown in Figure 3. The set-up consists of two bobs with identical mass, $M = 0.5\ kg$, and three springs with spring constant κ_1, κ_{12}, and κ_2. κ_1 is set to be equal to κ_2 with a determined value of $\kappa' = 0.6\ Nm^{-1}$ while κ_{12} has a determined value of $1.0\ Nm^{-1}$. Here, the method of TMCMC will be used to quantify the uncertainty of κ' and κ_{12}.

The prior distributions for both κ' and κ_{12} are set to follow a Uniform distribution between $0.01\ Nm^{-1}$ and $3.00\ Nm^{-1}$. As in the case of the simple harmonic oscillator, the measured quantity would be the oscillation frequency. However, this time, there are two values of frequencies, ω_1 and ω_2, which are to be determined. These two values of frequencies correspond to the two distinct vibration modes associated with the coupled oscillator: in-phase and out-of-phase. The likelihood distribution of both ω_1 and ω_2 follow a Normal distribution with means $\omega_1(\kappa', \kappa_{12})$ and $\omega_2(\kappa', \kappa_{12})$ respectively and with standard deviations of $0.05\ Hz$. The mathematical expression of ω_1 and ω_2 as a function of κ' and κ_{12} are as follows:

$$\omega_1(\kappa', \kappa_{12}) = \sqrt{\frac{\kappa' + 2\kappa_{12}}{m}} \tag{6}$$

$$\omega_2(\kappa', \kappa_{12}) = \sqrt{\frac{\kappa'}{m}} \tag{7}$$

The number of samples to be obtained for this TMCMC case example, N, would be set at 10000. The results will be illustrated in the form of a contour profile of the scatter plots as seen in Figures 4 to 8.

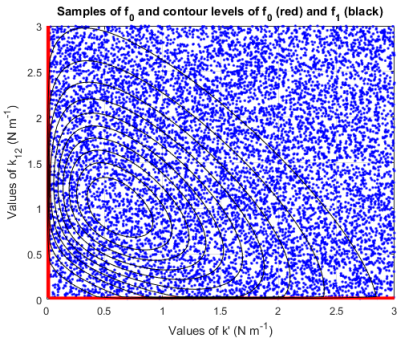

Fig. 4. Contour profile of the scatter plot at iteration $j = 0$ obtained for $N = 10000$.

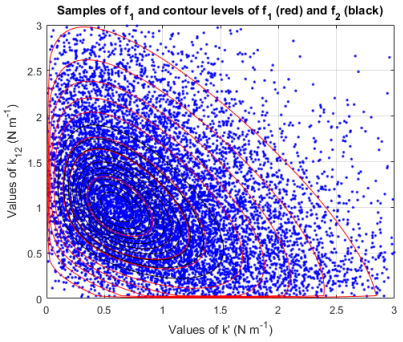

Fig. 5. Contour profile of the scatter plot at iteration $j = 1$ obtained for $N = 10000$.

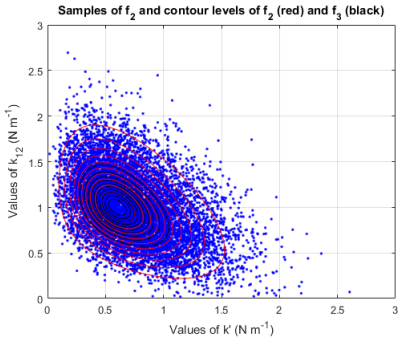

Fig. 6. Contour profile of the scatter plot at iteration $j = 2$ obtained for $N = 10000$.

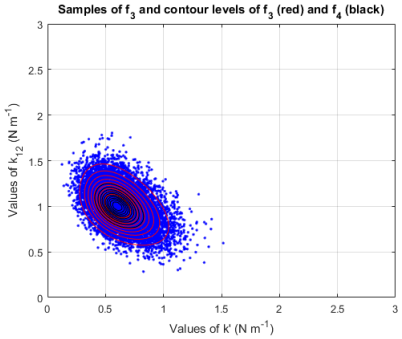

Fig. 7. Contour profile of the scatter plot at iteration $j = 3$ obtained for $N = 10000$.

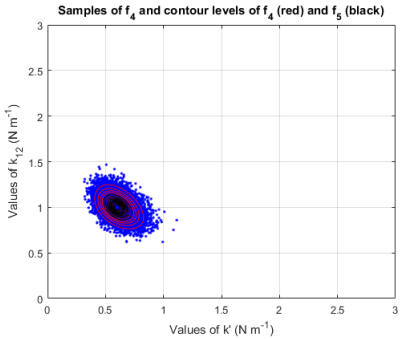

Fig. 8. Contour profile of the scatter plot at final iteration $j = 4$ obtained for $N = 10000$.

Note that in the results above, the red contours reflect the contour profile of the scatter plots in the current iteration while the black contours reflect the contour profile of the scatter plots in the next iteration. As seen in Figure 4, the scatter plot follows a Uniform distribution along the values of the parameters κ' and κ_{12} whereas in Figure 8, the scatter plot appears to possess a contour profile belonging to a Normal distribution with a distinct peak. In addition, the scattered points and the contour lines converge towards a specific coordinate which indicates the estimated values of κ' and κ_{12}.

Based on Figure 8, the estimated values of κ' and κ_{12} are summarized in the table below:

Table 2. Results of the estimated and determined values of κ' and κ_{12}.

Parameter	Estimated value (Nm^{-1})	Determined value (Nm^{-1})
κ'	0.6061	0.6
κ_{12}	1.0100	1.0

Based on solutions obtained analytically, the theoretical values of κ' and κ_{12} are as shown in Table 2. Compared to the estimated values obtained via the TMCMC method, the percentage discrepancy is at 0.678% and 1.00% for κ' and κ_{12} respectively which indicates a high degree of accuracy associated with the results.

3.2. *Limitation(s) of TMCMC*

The case example highlights a key advantage of the TMCMC method which is that it can be employed for multi-dimensional problems whereby more than one parameter is to be studied. However, as the dimension of the problem becomes higher, more data is to be analyzed resulting in the increase of computational cost, thereby reducing the efficiency of the TMCMC method. Such shortcoming also applies to the MCMC method. To address this, small subsets of 'highly-informative' data will be chosen selectively while the rest are discarded. This, however, may potentially exclude just as 'highly-informative' data as highlighted by Green and Maskell (2018). To tackle these problems, Green and Maskell (2018) propose the use of the SMC sampling method to solve Bayesian inference problems involving big data.

4. Sequential monte-carlo (SMC) sampling

SMC, according to Moral et al. (2006), is a sampling method consisting of a set of Monte-carlo algorithms and is generally employed to solve Bayesian inference problems involving large data sets and to sample from a time-evolving posterior in a sequential manner.

The workings of SMC is made up of two phases: Importance Sampling (IS) and Resampling.

IS is a technique which recognizes the impact of some input random samples over the others in the estimation of a parameter and thus puts more

weight on these data as mentioned by Rubinstein and Kroese (2016). This leads to the increase sampling of these relatively important samples resulting in the increased precision of the estimator. The mathematical description is as such. Suppose we want to determine the expected value of a function, $f(\boldsymbol{\theta})$, from an un-normalized target distribution, $T^*(\boldsymbol{\theta})$, it follows that:

$$E[f(\boldsymbol{\theta})] = \frac{\int f(\boldsymbol{\theta})T^*(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int T^*(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

$$E[f(\boldsymbol{\theta})] = \frac{\int f(\boldsymbol{\theta})q(\boldsymbol{\theta})\omega(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int q(\boldsymbol{\theta})\omega(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{8}$$

In the above equation, $\omega(\boldsymbol{\theta}) = \frac{T^*(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}$ and it represents the 'importance weight' of samples $\boldsymbol{\theta}$. $q(\boldsymbol{\theta})$ represents the proposal distribution. As such, Equation (8) can be re-expressed as such:

$$E[f(\boldsymbol{\theta})] \approx \sum_{i=1}^{N} f(\theta^i)\hat{\omega}^i \tag{9}$$

In Equation (9), $\hat{\omega}^i = \frac{\omega^i}{\sum_j \omega^j}$ and it represents the normalized importance weights for $i = 1, ..., N$ where N is the total number of samples.

Re-sampling, according to Green and Maskell (2018), is the idea of performing repeated sampling with replacement from the original sample. This makes up for the degeneracy problem which arises during IS whereby too few samples have significant weights. The mathematical description of Re-sampling is as such. We first define $f(\theta_j) = \delta(\theta_j - \boldsymbol{\theta})$, for which δ denotes a Dirac delta function. Using the definition of expected values, with $T(\boldsymbol{\theta})$ denoting the normalized target distribution, we obtain the following expression:

$$E[f(\theta^j)] = \int \delta(\theta^j - \boldsymbol{\theta})T(\boldsymbol{\theta})d\boldsymbol{\theta} = T(\theta^j) \tag{10}$$

$$\therefore\ E[f(\theta^j)] \approx \sum_{i=1}^{N} \delta(\theta^j - \theta^i)\hat{\omega}^i = \hat{\omega}^j \tag{11}$$

Equations (10) and (11) imply that if we have set of N samples for which the respective weight of each sample is known, $\{\theta^1, \hat{\omega}^1\}, ..., \{\theta^N, \hat{\omega}^N\}$, while a new set of samples, $\{\bar{\theta}^1, ..., \bar{\theta}^N\}$ is being chosen for which $P(\bar{\theta} = \theta^i) = \hat{\omega}^i$, then $\{\bar{\theta}^1, ..., \bar{\theta}^N\}$ will become the approximate samples obtained from $T(\boldsymbol{\theta})$.

To determine the condition under which Re-sampling is to be executed, Kong et al. (1994) first introduced the term 'effective sample size', N_{eff}, which is defined as follows:

$$N_{eff} = \frac{1}{\sum_i (\hat{\omega}^i)^2} \tag{12}$$

This serves as an indicator such that if this value of N_{eff} falls below a certain threshold value, N_t, Re-sampling will be carried out. In the paper by Green and Maskell (2018), this value of N_t is set at $\frac{N}{2}$.

The workings of a SMC algorithm can be summarized as such:

- Step 1: In the initialization step, iteration $k = 1$, samples $\{\theta_k^1, ..., \theta_k^N\}$ are obtained from the proposal distribution, $p(\theta_k)$ with initial weights $\omega(\theta_k^i) = \frac{T^*(\theta_k^i)}{q(\theta_k^i)}$ where $i = 1, ..., N$.
- Step 2: In the IS step, iteration $k = 2$, sample $\{\theta_k^1, ..., \theta_k^N\}$ from the new proposal distribution $p(\theta_k|\theta_{k-1})$ based on the previous sample sets and their respective weights. For the new samples obtained, their respective importance weights will be evaluated and normalized
- Step 3: Within the same iteration, if $N_{eff} < \frac{N}{2}$, the Re-sampling step is executed after the IS step. In general, for Re-sampling, the samples obtained from the previous iteration are sampled from the sample set obtained from the current iteration according to their respective importance weights.
- Step 4: For the subsequent new iterations, $k = 3, 4, ...$, return to Step 2 and the process repeats itself.

A simple application of the SMC sampling method would be in solving an unknown parameter within a static linear system. An example would be the simple harmonic oscillator which would be analyzed once again to estimate the unknown spring constant parameter.

4.1. *SMC case example: Simple harmonic oscillator*

With reference to the set-up as seen in Figure 1, the spring obeys the Hooke's Law which can be expressed in the form of a linear equation (Hooke (1678)):

$$F = k \cdot x \tag{13}$$

In the above equation, F represents the magnitude of the restoring force on the spring due to the weight of the mass, k represents the spring constant , and x represents the length of displacement by the spring due to the applied force.

For this case example, the measured quantity Y consists of 20 independent and equally spaced

artificial measurements between (and including) 1.0 N and 10.0 N. This measurement is also accompanied by external "noise", ϵ, which will be assumed to follow a Normal distribution with mean 0.0 N and a standard deviation, σ_ϵ, of 0.8 N. As such, the t^{th} measurement of Y for a given t^{th} measurement of F, where $t = 1, 2, ..., 20$, can be expressed as follows:

$$Y_t = k \cdot x_t + \epsilon \tag{14}$$

As such, the likelihood function, $P(Y_{1:20}|k)$, whereby $Y_{1:20}$ represents the sum of 20 independent measurements of Y, can be expressed in the form of a Normal distribution:

$$P(Y_{1:20}|k) \propto exp(-\frac{1}{2\sigma_\epsilon^2}\sum_{t'=1}^{20}(Y_{t'} - k \cdot x_{t'})^2) \tag{15}$$

On the other hand, the prior function for k, $P(k)$, would be taken to follow a Normal distribution with mean, μ_0, of 3.0 Nm^{-1} and standard deviation, σ_0, of 1.0 Nm^{-1}:

$$P(k) \propto exp(-\frac{1}{2\sigma_0}(k - \mu_0)^2) \tag{16}$$

Using Equation (2), the posterior, $P(k|Y_{1:20})$, can be expressed as follows:

$$P(k|Y_{1:20}) \propto exp(-\frac{1}{2\sigma^2}(k - \mu)^2) \tag{17}$$

whereby

$$\sigma^2 = \frac{1}{\sigma_\epsilon^{-2}\sum_{t'=1}^{20} x_{t'}^2 + \sigma_0^{-2}} \tag{18}$$

$$\mu = \sigma^2(\frac{1}{\sigma_\epsilon^2}\sum_{t'=1}^{20} x_{t'} \cdot y_{t'} + \frac{\mu_0}{\sigma_0^2}) \tag{19}$$

For this case example, the number of samples, N, to be obtained would be set at 10000 with 500 iterations to be performed. The results of the SMC sampling method is as shown in Figure 9.

Based on the graphical plot in Figure 9, the determined value of mean k is at 0.982 Nm^{-1}, while the determined values of the upper and lower bounds are 1.01 Nm^{-1} and 0.953 Nm^{-1} respectively.

The estimated value of k from the SMC sampling method shows little deviation from the determined values which indicates a high degree of precision associated with the results. In addition, the one sigma value, σ, is at 0.0291 Nm^{-1} which indicates a percentage error of 2.96% associated with the estimated value of k. This implies a high degree of precision of the estimated value of k which is expected given that a large sample size of $N = 10000$ is used.

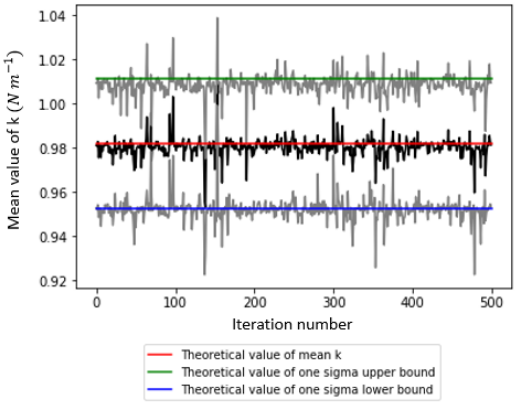

Fig. 9. Result of the estimated value of k obtained over 500 iterations for $N = 10000$.

4.2. *Limitation(s) of SMC*

A distinct advantage in the SMC sampling method lies in its ability to perform parallel computations as it is able to compute the state for all N samples in the system within one iteration. In contrast, the MCMC and TMCMC methods perform only series computation as they only compute the state of one sample per iteration. One key limitation of the SMC sampling method, however, lies in its assumption that measurements are taken independently of one another in the sense that the "noise" associated with one data is not carried forward in the next successive data as stated by Green and Maskell (2018). In reality, such assumption may not be true and for such problems, the SMC sampling method may not be able to efficiently address it.

5. Conclusion

In this conference paper, a review and discussion of MCMC, TMCMC and SMC sampling methods have been provided in detail. This aims to provide the readers, especially those who are new to these concepts, a concise introduction so as to gain a better understanding of each of these sampling methods.

In addition, case examples in the form of toy-problems have been analyzed which seeks to illustrate their applications in quantifying the uncertainty of the parameters of interest within the context of simple engineering problems. Such applications can be further extended to estimating parameters within a distribution function that is used to model the failure probability of an indus-

trial component so as to perform reliability analysis under imprecise or incomplete information.

Finally, the paper also evaluates and highlights the key strengths and limitations of each of these sampling methods so as to provide and better understanding as to which of these methods would be most effective under the different types of problems that needs to be addressed and the nature of their set-up. The MCMC sampling method would be useful in one-dimensional problems whereby only one parameter needs to be estimated while TMCMC would be useful in situations whereby the target distribution exhibits one or multiple sharp peaks or in higher-dimension problems when two or more parameters need to be estimated. The SMC sampling method would be preferred in cases when there is a large data set which needs to be handled and when the target distribution is updated with each new information.

Acknowledgement

This research would not have been possible without the guidance and inputs from my supervisors Dr Edoardo Patelli and Dr Alice Cicirello who are also co-authors of this paper. In addition, we would also like to thank Dr Matteo Broggi from Leibniz Universität Hannover and Dr Peter Green from University of Liverpool who provided their insights on the topics shared in this paper, without which, the writing of this paper would not have been made possible as well. Finally, we would also like to thank the Institute for Risk and Uncertainty for the facilities which allowed for this research to be carried out smoothly as well as the Singapore Nuclear Research and Safety Initiatives for the sponsorship for my PhD study.

References

Beck, J. L. and S. K. Au (2002). Bayesian updating of structural models and reliability using markov chain monte carlo simulation. *Journal of Engineering Mechanics 128*, 380–391. doi: 10.1061/(ASCE)0733-9399(2002)128:4(380).

Beech, D. G., M. G. Kendall, and A. Stuart (1959). The advanced theory of statistics. volume 1, distribution theory. *Applied Statistics 8*, 61. doi: 10.2307/2985818.

Ching, J. Y. and Y. C. Chen (2007). Transitional markov chain monte carlo method for bayesian model updating, model class selection, and model averaging. *Journal of Engineering Mechanics 133*. doi: 10.1061/(ASCE)0733-9399(2007)133:7(816).

Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference*. Boca Raton: Chapman Hall.

Green, P. and S. Maskell (2018). Estimating the parameters of dynamical systems from big data using sequential monte carlo samplers. *Mechanical Systems and Signal Processing 93*, 379–396. doi: 10.1016/j.ymssp.2016.12.023.

Hooke, R. (1678). *Lectures de Potentia Restitutiva, Or of Spring Explaining the Power of Springing Bodies*. John Martyn.

Kong, A., J. S. Liu, and W. H. Wong (1994). Sequential imputations and bayesian missing data problems. *Journal of the American Statistical Association 89*, 278–288. doi: 10.1080/01621459.1994.10476469.

Moral, P. D., A. Doucet, and A. Jasra (2006). Sequential monte carlo samplers. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 68*, 411–436.

Ravenzwaaij, D. V., P. Cassey, and S. D. Brown (2018). A simple introduction to markov chain monte-carlo sampling. *Psychonomic Bulletin Review 25*, 143–154. doi: 10.3758/s13423-016-1015-8.

Robert, C. P. and G. Casella (2013). *Monte Carlo Statistical Methods* (2 ed.). Springer Science & Business Media. ISBN: 978-1475741452.

Rubinstein, R. Y. and D. P. Kroese (2016). *Simulation and the Monte Carlo Method* (3 ed.). John Wiley & Sons. ISBN: 978-1118632161.

Serfozo, R. (2014). *Basics of Applied Stochastic Processes*. Berlin: Springer Berlin. ISBN: 978-3540893318.

Smith, A. F. M. and G. O. Roberts (1993). Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological) 55*, 3–23.

Thorton, S. T. and J. B. Marion (2004). *Classical Dynamics of Particles and Systems* (5 ed.). Brooks/Cole. ISBN: 978-0534408961.