MATEMATIKA, 2015, Volume 31, Number 1, 25–36

Bayesian Inference using Multiple-try Metropolis Hastings Scheme for the Efficiency of Estimating Gumbel Distribution Parameters

[1]**Nor Azrita Mohd Amin,** [2]**Mohd Bakri Adam and** [3]**Noor Akma Ibrahim**

[1]Institute of Engineering Mathematics
Universiti Malaysia Perlis, Kampus Pauh Putra, 02600 Arau, Perlis, Malaysia
[1,2,3]Institute of Mathematical Research
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
e-mail: [1]norazrita@unimap.edu.my, [2]bakri@science.upm.edu.my, [3]nakma@putra.upm.edu.my

Abstract This paper aims to explore the efficiency for estimating the parameters of Gumbel simulated data using Multiple-try Metropolis algorithm (MTM). Several goodness-of-fit tests are used to compare the performance of MTM and the former, Metropolis-Hastings algorithm (MH). Concerning for a fair comparison, this study uses the equivalent starting point, the similar number of iterations and also the same length of burn-in periods. The numerical studies show that the MTM method performs slightly better than MH method after 5000 iterations to meet the stationary distribution. More candidates in the proposals lead to a higher accuracy of MTM estimation.

Keywords Markov chain Monte Carlo; Multiple-try Metropolis algorithm; Metropolis-Hastings algorithm; goodness of fit; Gumbel distribution

2010 Mathematics Subject Classification 60J22 62F15 62G32 60G70

1 Introduction

The Multiple-try Metropolis algorithm (MTM) and the Metropolis-Hastings algorithm (MH) are developed based on the foundation of Markov chain Monte Carlo (MCMC) simulation technique. The idea of Bayesian MCMC arise when the target distribution, say $\pi(x)$ is complex such that it is difficult to sample from it directly. A series of samples generated from $\pi(x)$ will construct an aperiodic and irreducible Markov chain with state space E, and stationary to $\pi(x)$. The simulated values from the long enough chains can be treated as a dependent samples from the target distribution and used as a basis for summarizing the important features of $\pi(x)$ (Brooks [1]). A common feature of methods for simulating from $\pi(x)$ is their reliance on cumulative evolutions of small, albeit random changes. A fundamental construction that enables such "local-search" methods is a Markov transition function of the MH (Liu *et al.* [2]).

Bayesian Markov chain has gained its popularity in statistical analysis for the inferences of posterior distributions as well as to predict the future probability and used in decision making. MH algorithm (Metropolis *et al.* [3], Hastings [4]) and Gibbs sampling (Geman and Geman [5]) are very famous and most practical MCMC for simulation studies. MH used to obtain a sequence of random samples from a probability distribution for which direct sampling is difficult. This sequence can be used to approximate the required distribution. Chib and Greenberg [6] and Gamerman and Lopes [7] provide a comprehensive preliminary details as well as an intensive development and applications of MCMC specifically on MH techniques. MH is a fundamental algorithm for many Markov chain simulation approaches while the Gibbs sampling is a good alternative if the full conditional distributions for each

parameter are known. Gibbs sampler move aimed at updating the parameters of the current model sampling from the full conditional distribution. The Gibbs sampler is facing the difficulty to deal with the required conditional distributions. If the posterior doesn't look like any distribution or having no conjugacy, then MH is the suitable method to use to generate random samples from a target distribution $\pi(x)$ for which direct sampling is cumbersome.

MH is constantly preferable for handling complex computational problems. However, to speed up the convergence and reduce the "burn-in " period, several extensions have been proposed in the literatures. For instance, the MTM scheme introduced by Liu *et al.* [2] in which the next state of the Markov chain is selected from a set of independent samples drawn from a generic proposal density according to certain weights. The main advantage of MTM is that it can explore a larger portion of the sample space without a decrease in the acceptance rate. The MTM method consists of proposing at each step, a fixed number of moves and then selecting one of them with a certain probability to increase the exploration sample space and for better mixing. Martino and Read [8] explore the flexibility of the MTM and confirmed the detailed balance condition is satisfied for MTM. Recent work on MTM can be found in Bedard *et al.* [9], Casarin [10] and Pandolfi *et al.* [11] with some luminous additional information. Currently there is a wide variety of MCMC algorithms developed and practiced. But it is important to understand that each idea have its own distinct advantages and drawbacks between one another. MCMC methods offer a great statistical tool and have been explored in diverse area. The execution of these approaches requires deep understanding and skill and this paper hopefully give a little knowledge for those who work in the same field. This study works on a simulation study of Gumbel distribution using Inversion method with the MH and MTM approaches are used for fitting the model. The interesting issue to discuss is about how long chains should be considered to be run and how the number of proposal in the MTM scheme could improve the MH algorithm.

2 Methodology

In general, the Bayesian concept consists of transferring the initial belief about the parameter, θ represented by the prior distribution $f(\theta)$ into a posterior distribution $f(\theta|x)$ that consist of the additional information provided by the data x in the likelihood function $L(x|\theta)$. Then the posterior density for θ by Bayes' rule is obtained up to a proportionality constant by multiplying the prior density to the likelihood function,

$$f\left(\theta|x\right) \propto f\left(\theta\right) L\left(x|\theta\right).$$

Coles and Tawn [12] illustrate the usefulness of prior in data analysis. The proper elicitations of prior using expert information complement the data and lead to improve estimate of extreme data.

The explosion of interest in Bayesian methods over the last ten to twenty years has been the result of the convergence of modern computing power and efficient MCMC algorithms for sampling from the posterior distribution (Carlin and Louis [13]). The main Monte carlo procedure is about drawing the random samples from the target probability density function (pdf) while the MCMC generates a Markov chain such that its stationary distribution coincides with the target pdf (Martino and Read [8]). The advantage of MCMC to sample

directly from the posterior distribution and construct sample estimates could avoid the difficulties to evaluate the complex integrate function which is extremely important in Bayesian statistics, see Brooks [1] for details.

2.1 Metropolis Hastings algorithm

MH routine is capable to simulate a series from an arbitrary density as a basis for summarizing features of the equilibrium distribution which is a Bayesian posterior distribution for unknown parameter θ (Smith and Roberts [14]). The simple steps of MH algorithm can be summarized as follows.

(i) Set the initial values for (μ^0, σ^0).

(ii) Given that the chain is currently at (μ^j, σ^j) :
Draw a candidate value $\mu^{can} \sim N(\mu^j, \upsilon_\mu)$ and $\sigma^{can} \sim N(\sigma^j, \upsilon_\sigma)$ for some suitably chosen variance υ_μ, υ_σ and take

$$\mu^{(j+1)} = \begin{cases} \mu^{can} & \text{with probability } p, \\ \mu^j & \text{with probability } 1-p \end{cases}$$

where p is the acceptance probability,

$$p = \min\left\{1, \frac{\pi\left(\mu^{can}|\sigma^j\right)}{\pi\left(\mu^j|\sigma^j\right)}\right\}.$$

and $\pi(\mu|\sigma)$ is the conditional posterior distribution for μ. This is implemented by drawing $u \sim \text{Unif}(0,1)$ and taking $\theta^{(j+1)} = \theta^{(can)}$ if and only if $u < p$.

(iii) Iterate the updating procedure.

The variance of the candidate value, υ is typically chosen by trial and error and aiming the acceptance probability to be roughly around 30%. The time it takes for the chains to converge is varies depending on the starting point. Usually a certain number of the first draws are thrown or also known as the burn-in period. This is to make the draws closer to the stationary distribution and less dependent on the starting point. The local moves in the MH algorithm is causing low converging algorithms while increasing the searching region of each MH step characterized by the Markov transition function will decrease the MH ratio and therefore not an effective algorithm. MTM algorithm tackles this problem by proposing multiple trial points from the transition function for a larger searching region (Liu *et al.* [2]).

2.2 Multiple-try Metropolis algorithm

Basically, MH used to sample from a posterior distribution which is complicated to sample from directly. MH always works for any arbitrary distribution but it can be very slow. Therefore Liu *et al.* [2] came out with the new ideas for improving this algorithm. MTM modified the standard MH by replacing the single proposal y with a set of k independent and identically distributed (iid) proposals $y_1, \ldots, y_k$ from the proposal distribution, $q(y|x)$

in order to make larger step-size jumps, without lowering the acceptance rate. Based on Pandolfi *et al.* [15], suppose that $q(y|x) > 0$ if and only if $q(x|y) > 0$ and $\lambda(x|y)$ be a nonnegative symmetric function, define

$$w(x, y) = \pi(x)q(y|x)\lambda(x|y)$$

then the MTM algorithm can be simplified as the following:

(i) Draw k iid proposals $y_1, \ldots, y_k$ from $q(y|x)$, and compute $w_i = w(\theta_i, x)$ for $i = 1, \ldots, k$.

(ii) Select $y = y_j$ from $y_1, \ldots, y_k$ with probability proportional to w_i.

(iii) Draw $x_1, \ldots, x_{k-1}$ from $q(.|y)$ and set $x_k^* = x$ and compute $w_i^* = w(x_i^*, y)$.

(iv) Accept y with probability:

$$a = \min\left\{1, \frac{w_1 + \ldots + w_k}{w_1^* + \ldots + w_k^*}\right\}.$$

In order to attain the detailed balance condition, MTM used selection probabilities which are proportional to the product of the target, the proposal, and a λ function which has to be non-negative and symmetric function in x and y which can be chosen by the user. Liu *et al.* [2] listed some of the possible function for $\lambda(x|y)$. It can be proven that MTM satisfied the detailed balance condition and therefore produces a reversible Markov chain with $\pi(x)$ as the stationary distribution. It is sufficient to prove that $\pi(x_t)P(x_t, y) = \pi(y)P(y, x_t)$ where $P(x_t, y)$ is the transition probability of the Markov chain from state x_t to y. The proof is as follow:

$$\begin{aligned}
\pi(x_t)P(x_t, y) &= k\pi(x_t)T(x_t, y)p_y \int \ldots \int T(x_t, y_1)\ldots T(x_t, y_{k-1}) \min\left\{1, \frac{\pi(y)T(y, x_t)p_{x_t}}{\pi(x_t)T(x_t, y)p_y}\right\} \\
&\quad \times T(y, x_1^*)\ldots T(y, x_{k-1}^*)dy_1\ldots dy_{k-1}dx_1^*\ldots dx_{k-1}^* \\
&= k\int \ldots \int T(x_t, y_1)\ldots T(x_t, y_{k-1}) \min\{\pi(x_t)T(x_t, y)p_y, \pi(y)T(y, x_t)p_{x_t}\} \\
&\quad \times T(y, x_1^*)\ldots T(y, x_{k-1}^*)dy_1\ldots dy_{k-1}dx_1^*\ldots dx_{k-1}^* \\
&= k\pi(y)T(y, x_t)p_{x_t} \int \ldots \int T(x_t, y_1)\ldots T(x_t, y_{k-1}) \min\left\{1, \frac{\pi(x_t)T(x_t, y)p_y}{\pi(y)T(y, x_t)p_{x_t}}\right\} \\
&\quad \times T(y, x_1^*)\ldots T(y, x_{k-1}^*)dy_1\ldots dy_{k-1}dx_1^*\ldots dx_{k-1}^* \\
&= \pi(y)P(y, x_t).
\end{aligned}$$

MTM is efficiently expanding the proposal region to improve convergence performance by generating a larger number of candidates and therefore improving exploration of f near $x^{(t)}$ (Givens and Hoeting [16]). These techniques have witnessed a recent surge of interest because they lend themselves easily to parallel implementations. In this work, the R code is developed to implement the MH and MTM algorithms for the inferences of Gumbel distribution (R, [17]). For both algorithm, Normal distribution, $M(x^t, \sigma^2)$ is used as a proposal distribution centered on the current value, x^t. The variance σ^2 plays an important role for the method to converge to the stationary distribution although undoubtedly the progress can be exceedingly slow. Too large σ^2 will cause almost every steps of the algorithm will be rejected while if too small, almost every steps will be accepted.

2.3 Goodness-of-Fit Tests

The goodness-of-fit tests perform an essential role for determining the best fitted model of the observed data. This study proposes the goodness-of-fit tests for Gumbel distribution simulated data using Anderson-Darling (AD), Cramer-von Mises (CVM), Kolmogorov-Smirnov (KS) test and the Root Mean Square Error (RMSE). The smaller value of the goodness-of-fit test reflects the better fit of the data. Some brief definitions of those tests are listed as follow.

2.3.1 Anderson-Darling

Let $x_1 < x_2 < \ldots < x_n$ be order statistics for the observations with size n. The computational formula of the AD statistic is

$$\mathrm{AD} = -\sum_{i=1}^{n} \{\ln[G(x_i)] + \ln[1 - G(x_{n+1-i})]\} - n$$

where G(.) is the cdf of Gumbel distribution.

2.3.2 Kolmogorov-Smirnov

The KS statistic is based on the value of D given by

$$\mathrm{D} = \max_{i=1,\ldots,n} [\delta_i]$$

where

$$\delta_i = \max \left[\frac{i}{n} - G(x_i), G(x_i) - \frac{i-1}{n} \right].$$

2.3.3 Cramer-von Mises

Similar to the AD conditions, the computational formula of the CVM statistic is

$$\mathrm{CVM} = \frac{1}{12n} + \sum_{i=1}^{n} \left[G(x_i) - \frac{2i-1}{2n} \right]^2.$$

2.3.4 Root Mean Square Error

The theoretical formulation of RMSE is given by the following equation

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - X_i^*)^2}$$

where X_i and X_i^* are the sample data and the estimated values of ith observations given by Gumbel distribution. The quantile of Gumbel distribution is given by

$$x = \mu - \sigma \left(\log \left(- \log \left(U_{i,n} \right) \right) \right)$$

where U is the hypothetical distribution function (Abidin *et al.* [18]). There are many types of hypothetical distribution function. The common choice is $U_{i,n} = \frac{i-0.5}{n}$ and the rank $U_{i,n}$, of ith order statistics from sample size of n is uniformly distributed $U(0,1)$ (Balakrishnan *et al.* [19]).

3 Simulation Data

An essential characteristic of statistical modeling is the connections between proximity to the real world and might be magnificently managed with simulation methods. Gumbel distribution is often used to model the distribution of the maximum (or the minimum) level of a process which is practical and relevant for predicting the future extreme events such as flood, earthquake, air and water pollutions or other natural disaster. In extreme value theory, Gumbel distribution is a special case of the generalized extreme value distribution in which the shape paremeter, ξ is equal to zero. There are two parameters in Gumbel distribution, location, μ and scale, σ parameter. The distribution function for Gumbel distribution is given by equation (1).

$$G(z) = \exp\left\{-\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]\right\}, \quad -\infty < \mu < \infty, \quad \sigma > 0 \tag{1}$$

and differentiating equation (1) to have the density function as in equation (2).

$$g(z) = \frac{1}{\sigma}\exp\left\{-\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right] - \left(\frac{z-\mu}{\sigma}\right)\right\}. \tag{2}$$

There are several techniques for simulating data from a specified distribution. Very common approaches are using the inversion method and the rejection method. The basic inversion method works as the following steps.

(i) Let U have a uniform distribution on the interval from 0 to 1.

(ii) Generate random numbers from U, $u_1, \ldots, u_n$.

(iii) For each j from 1 to n, let $x_j = F^{-1}(u_j)$. For Gumbel distribution, let $u = G(z)$ and therefore

$$x = \mu - \sigma \log(-\log(u)).$$

Generate a sequence of *iid* random samples, x_j for $j = 1, \ldots, n$ from Gumbel distribution, $G(z)$. Figure 1 displays the histogram and the density plot of the simulated data with length 100. The generated series for a distribution will not give precisely the values of parameter of interest, but at least very close to it. A classical frequentist approach based on maximum likelihood estimation is used as a benchmark value for estimating the parameters before proceed to the MCMC approach. In addition, this benchmark values will be used as an initial values of the iterations for both MH and MTM as suggested in Brooks [1].

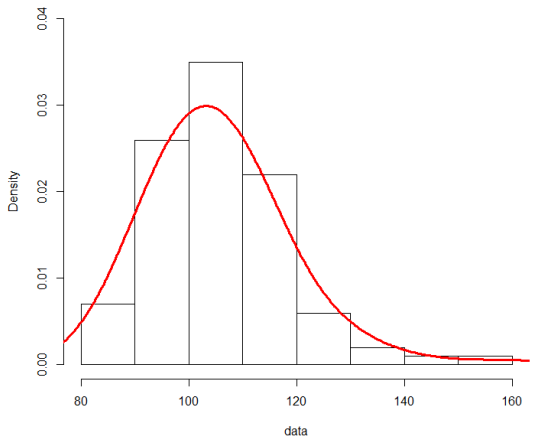

Figure 1: Histogram and Density Plot of Gumbel$(100, 10)$ Simulated Data

3.1 Inference for Gumbel Distribution

Many techniques have been proposed to estimate the parameters of extreme value models. These include graphical techniques based on versions of probability plots, moment-based techniques, maximum likelihood estimator and the Bayesian approach. Each technique has its pros and cons. This paper focuses on the Bayesian MCMC approaches based on MTM and MH algorithms for the efficiency of estimating the parameters. At present, literatures on Bayesian methodology in Malaysia is still on the early stage that make the expert information is not yet available to supply prior information on certain issues. Therefore non-informative prior is always preferable for the analysis. The combination of Gumbel and Rayleigh pair of priors for Gumbel parameters have been assumed for posterior analysis based on Rostami and Adam [20]. The likelihood function is given by

$$L(\theta|x) = \frac{1}{\sigma^n}\exp\left\{-\sum_{i=1}^{n}\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]-\sum_{i=1}^{n}\left(\frac{z-\mu}{\sigma}\right)\right\}.$$

Thus we have two parameter model with parameters μ and σ. Define the prior for both parameter, $\pi(\theta)$, then we have

$$\pi(\mu) \propto \exp\left\{-\left(\frac{\mu-\mu_0}{K_0}\right) - \exp\left(-\frac{\mu-\mu_0}{K_0}\right)\right\},$$
$$\pi(\sigma) \propto \sigma\exp\left\{-\frac{\sigma^2}{2\lambda_0^2}\right\}.$$

The posterior distribution of $\pi(\mu, \sigma|x)$ is given by

$$\pi(\mu, \sigma|x)$$
$$= \sigma^{1-n}\exp\left\{-\sum_{i=1}^{n}\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]-\sum_{i=1}^{n}\left(\frac{z-\mu}{\sigma}\right)-\left(\frac{\mu-\mu_0}{K_0}\right)-\exp\left(-\frac{\mu-\mu_0}{K_0}\right)-\frac{\sigma^2}{2\lambda_0^2}\right\}.$$

The conditional posterior distribution, $\pi(\theta|x)$ for the parameters are obtained by ignoring all the terms that does not involved parameter μ and σ respectively. Then

$$\pi(\mu|x) = \exp\left\{-\sum_{i=1}^{n}\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]-\sum_{i=1}^{n}\left(\frac{z-\mu}{\sigma}\right)-\left(\frac{\mu-\mu_0}{K_0}\right)-\exp\left(-\frac{\mu-\mu_0}{K_0}\right)\right\},$$

$$\pi(\sigma|x) = \sigma^{1-n}\exp\left\{-\sum_{i=1}^{n}\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]-\sum_{i=1}^{n}\left(\frac{z-\mu}{\sigma}\right)-\frac{\sigma^2}{2\lambda_0^2}\right\}.$$

Since it is difficult to recognize the pattern of the distribution, Gibbs sampler is discarded and Metropolis Hasting approach is employed for the posterior inference. Given the updating condition for parameter θ as below,

$$\theta^{(j+1)} = \begin{cases}\theta_{can} & \text{if } u < p,\\ \theta^{(j)} & \text{otherwise,}\end{cases}$$

in which p is the acceptance probability,

$$p = \min\left\{1, \frac{\pi\left(\theta^{can}|x\right)}{\pi\left(\theta|x\right)}\right\}.$$

It is more convenient to take the logarithm of p. Therefore $\log(p_1)$ and $\log(p_2)$ as in equation (3) and equation (4) represent the acceptance probability for parameter μ and σ with Z is given as $Z = (z-\mu)/\sigma$:

$$\log p_1 = \frac{n}{\sigma}\left(\mu^{can}-\mu\right)+\frac{(\mu-\mu^{can})}{\kappa_0}+\sum_{i=1}^{n}\left\{\exp\left(-Z\right)-\exp\left(-\frac{z-\mu^{can}}{\sigma}\right)\right\}$$
$$+\exp(-\frac{\mu-\mu^0}{\kappa_0})-\exp(-\frac{\mu^{can}-\mu_0}{\kappa_0}), \tag{3}$$

and

$$\log p_2 = (1-n)(\log\sigma^{can}-\log\sigma)+\frac{1}{2\lambda_0^2}(\sigma^2-(\sigma^{can})^2)+\sum_{i=1}^{n}\left[(z-\mu)\left(\frac{1}{\sigma}-\frac{1}{\sigma^{can}}\right)\right]$$
$$+\sum_{i=1}^{n}\left[\exp\left(-Z\right)-\exp\left(-\frac{z-\mu}{\sigma^{can}}\right)\right]. \tag{4}$$

The MH and MTM algorithms were performed with the initial values (μ_0, σ_0) are the maximum likelihood estimates. For MTM algorithm, we set $\lambda(x, y) = 1$ and the proposals were drawn from normal distribution. The goodness-of-fit tests are used to compare the estimated parameters.

4 Results and Discussions

In MCMC simulations, the initial values, number of iterations and the burn-in periods are the issues that should be considered. The burn-in periods is the initial 'warm-up' periods

that have to be discarded to avoid the bias caused by the chosen starting point. Therefore determining the duration of the necessary burn-in period is important since the rates of convergence of different algorithms on different target distributions may vary considerably (Brooks, [1]). The first 500 iterations have to be discarded since the plot roughly before 500^{th} iterations for MH algorithm are not converge to the stable values. For comparison with MTM, the same iterations and burn-in periods are taken. The statistics of 10 runs with 2000, 5000 and 10 000 iterations for each single run are estimated using both MTM and MH algorithms.

The estimated values for both methods are very close to each other and comparable with the reference parameters. The goodness-of-fit tests are applied to compare the posterior summaries for MTM and MH for 2000, 5000 and 10 000 iterations. From the goodness-of-fit values in Table 1, we can conclude that after 5000 and 10000 iterations, MTM gives better performance than MH. There are not much differences of the goodness-of-fit values between estimates of 5000 and 10000 iterations for both methods since they have met the stationary condition. Table 1 also shows that the goodness-of-fit for 2000 iterations does not provides the stability of estimations for both methods. Therefore the subsequent results will be based on 5000 iterations.

Table 1: Goodness-of-fit Tests for MH and MTM with Different Number of Iterations

	AD	KS	CVM	RMSE
2000 iterations, 500 burn-in periods				
MTM	0.2818	0.0577	**8.3776**	1.3740
MH	0.2818	**0.0575**	8.3780	**1.3715**
5000 iterations, 500 burn-in periods				
MTM	**0.2787**	**0.0570**	**8.3772**	**1.3644**
MH	0.2801	0.0573	8.3776	1.3677
10000 iterations, 500 burn-in periods				
MTM	**0.2791**	**0.0571**	**8.3772**	**1.3659**
MH	0.2808	0.0574	8.3777	1.3698

The summary of the posterior draws are stated as in Table 2. The trace plots for μ and σ using MTM and MH algorithms were displayed in Figure 2 and Figure 3 indicates that convergence states have been achieved. The number of proposal in MTM method is the major difference between MTM and MH algorithm. Results in Table 3 depicts the higher number of proposals provides a more accurate estimate, with longer computational time.

5 Conclusions

Basically, MH is used to sample from a posterior distribution that is difficult to sample from directly. MH is always works for any arbitrary distributions but it can be very slow. Hence, MTM come out with the new ideas for improving MH algorithm by demonstrating that by generating a set of proposals will speed up the convergence instead of only depend on one proposal. MTM is the extension of MH algorithm wherein the subsequent state of

the Markov chains is chooses from a set of proposals. The chosen value is based on the corresponding weight determined by the posterior distribution. It is crucial to decide on how long the MCMC simulations need to be run for any application problems since the generated values is commonly used as a foundation for the inferences. Longer iterations and more complex functions require more computation time and storage space.

Table 2: Parameter Estimation Using MTM and MH for 5,000 Iterations

	μ	σ
MTM	100.107(0.761) (98.776,101.715)	9.9642(0.7867) (8.5506,11.6211)
MH	100.100(0.725) (98.823,101.677)	9.9515(0.7711) (8.6126,11.6194)

Table 3: Parameter Estimation Using MTM with Different Number of Proposals

k	μ	σ	AD	KS	CVM	RMSE
3	100.1748	10.0238	0.2944	0.0599	8.3786	1.4124
5	100.1200	9.9646	0.2811	0.0576	8.3774	1.3722
7	100.1184	9.9544	0.2796	0.0573	8.3772	1.3679
10	100.1188	9.9506	**0.2791**	**0.0572**	**8.3771**	**1.3666**

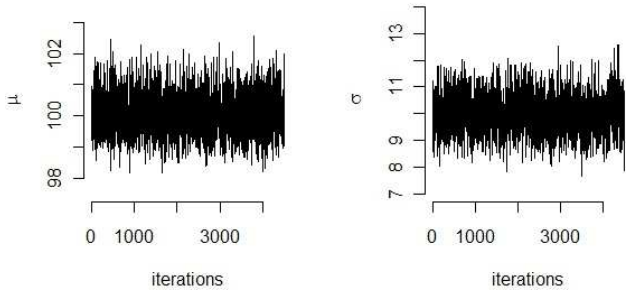

Figure 2: Trace Plot of MTM Run for Gumbel Fit

This study found out that the output for 5000 iterations are slightly equivalent with the output for 10 000 iterations after 500 burn-in periods. Therefore longer iterations are not required. To make sure the persistence, the average values of the output of 10 runs are taken for comparison. From the result, there are not much differences between the parameter estimation using MTM and MH. However, all goodness-of-fit tests performed suggest the better fit of MTM algorithm compared to MH. The trace plots visualize the more accurate plot using MTM than MH. For MTM analysis, the higher number of proposals k give better estimation but it need longer time for computations. MTM improves the convergence performance by increasing both the step size and the acceptance rate. This new method assured the efficient estimation scheme for modeling extreme data in term of

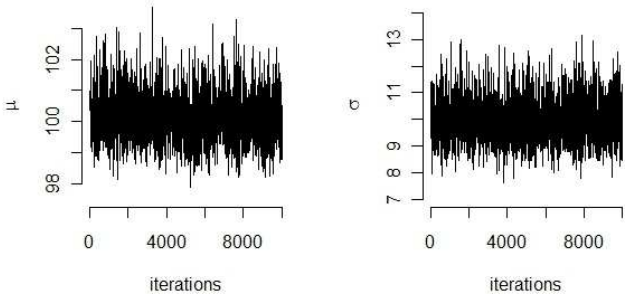

Figure 3: Trace Plot of MH Run for Gumbel Fit

the speed of convergence and small burn-in periods. The reason of MTM appears faster than MH is the additional computation in the iterations than MH does.

Acknowledgments

The author would like to thank the Department of Environment, Malaysia for the data used and the Ministry of Education Malaysia for the SLAI scholarship and research grant (ERGS/1/2013/STG06/UPM/02/3).

References

[1] Brooks, S. P. Markov chain monte carlo method and its application. *Journal of the Royal Statistical Society. Series D*. 1998. 47(1): 69–100.

[2] Liu, J. S., Liang, F. and Wong, W. H. The multiple-try method and local optimization in metropolis sampling. *Journal of the American Statistical Association*. 2000. 95(449): 121–134.

[3] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*. 1953. 21(6): 1087–1092.

[4] Hastings, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika*. 1970. 57(1): 97–109.

[5] Geman, S. and Geman, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*. 1984. 6(6): 721–741.

[6] Chib, S. and Greenberg, E. Understanding the metropolis-hastings algorithm. *The American Statistician*. 1995. 49(4): 327–335.

[7] Gamerman, D. and Lopes, H. F. *Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference*. 2nd Edition. United States: Chapman and Hall. 2006.

[8] Martino, L. and Read, J. On the flexibility of the design of multiple try metropolis schemes. *Computational Statistics*. 2013. 28(6): 2797–2823.

[9] Bedard, M., Douc, R. and Moulines, E. Scaling analysis of multiple-try mcmc methods. *Stochastic Processes and their Applications*. 2012. 122(3): 758–786.

[10] Casarin, R., Craiu, R. and Leisen, F. Interacting multiple try algorithms with different proposal distributions. *Statistical Computation*. 2013. 23: 185–200.

[11] Pandolfi, S., Bartolucci, F. and Friel, N. A generalized multiple-try version of the reversible jump algorithm. *Computational Statistics and Data Analysis*. 2014. 72: 298–314.

[12] Coles, S. G. and Tawn, J. A. A bayesian analysis of extreme rainfall data. *Applied Statistics*. 1996. 45: 463–478.

[13] Carlin, B. P. and Louis, T. A. *Bayesian Method for Data Analysis*. 3rd Edition. United States: Chapman and Hall. 2009.

[14] Smith, A. F. M. and Roberts, G. O. Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society. Series B*. 1993. 55(1): 3–23.

[15] Pandolfi, S., Bartolucci, F. and Friel, N. A generalization of the multiple-try metropolis algorithm for bayesian estimation and model selection. *Journal of Machine Learning Research*. 2010. 9: 581–588.

[16] Givens, G. H. and Hoeting, J. A. *Computational Statistics*. 2nd Edition. United States: John Wiley & Sons. 2012.

[17] Team, R. D. C. R: A language and environment for statistical computing. Technical Report. R Foundation for Statistical Computing. 2008.

[18] Abidin, N. Z., Adam, M. B. and Midi, H. The goodness-of-fit test for gumbel distribution: A comparative study. *MATEMATIKA*. 2012. 28(1): 35–48.

[19] Balakrishnan, N., Davies, K. F., Keating, J. P. and Mason, R. L. Correlation-type goodness of fit test for extreme value distribution based on simultaneous closeness. *Communications in Statistics-Simulation and Computation*. 2011. 40: 1074–1095.

[20] Rostami, M. and Adam, M. B. Analyses of prior selections for gumbel distribution. *MATEMATIKA*. 2013. 29(1): 95–107.