

Bayesian Multiple Change-point Estimation in Normal with EMC¹⁾

Jaehee Kim²⁾ and Sooyoung Cheon³⁾

Abstract

In this paper, we estimate multiple change-points when the data follow the normal distributions in the Bayesian way. Evolutionary Monte Carlo (EMC) algorithm is applied into general Bayesian model with variable-dimension parameters and shows its usefulness and efficiency as a promising tool especially for computational issues. The method is applied to the humidity data of Seoul and the final model is determined based on BIC.

1. Introduction

The problem to identify the change-points is of increasing interest in various fields since the change-points should be detected and the further process should be applied. Also finding the number of change-points is another big issue as a multiple change-point problem. This is one of the challenging statistical problems where the dimension of the object of inference is not fixed.

Recently there has been a more interest in the statistical analysis of change-point detection and estimation. It is mainly because change-point problems can be occurred in many disciplines such as economics, finance, medicine, psychology, geology, meteorology, environmental studies and etc. and even in daily lives.

In almost all classic statistical inference is based upon the assumption that there exists a fixed probabilistic mechanism of data generation. Unlike classic statistical inference, the Bayesian change analysis has interesting aspect to incorporate the previous information reflected with the appropriate priors. The existence of more than one data generation process is the most important characteristic of complex system.

1) This research is supported by Korea Research Foundation R04-2004-000-10138-0.

2) Associate Professor, Department of Statistics, Duksung Women's University, Seoul 132-714, Korea.

Correspondence : jaehee@duksung.ac.kr

3) Ph.D. Candidate, Department of Statistics, Texas A&M University, College Station, TX77843, USA.

When the hypotheses of statistical homogeneity hold true, that is, there exists only one mechanism of data generation, the law of large numbers are applied to make an inference. However if there exists change in the data generation, the probabilistic law should be applied to differently. In this case all data obtained should be sorted in subsamples generated by different probabilistic mechanisms. After this classification the correct inferences can be made.

It is important to detect possible changes of data generation process and the appropriate statistical analysis of such data must begin with testing and decisions about possible change. Chen and Gupta (2000) dealt with the parametric change-point problem in several distributions.

As a Bayesian change-point approach Chernoff and Zacks (1964) considered a Markov model for the normal observations. Yao (1984) gave the same probability model to a change in a Markov model and a product partition model using an approximation for the posterior means. Barry and Hartigan (1993) modeled the process by supposing that there is an underlying sequence of parameters partitioned into contiguous blocks of equal parameter values and the beginning of each block is said to be a change-point. In their product partition models, the probability of any partition is proportional to a product of prior cohesions and given the blocks the parameters in different blocks have independent prior distribution. For their Bayesian approach, the Markov sampling technique is used to calculation the posterior probabilities. With more possible partitions, the model space becomes complex with multiple modes, For the computational algorithm of posterior probabilities of this complicated form, recent computational methods are developed and those can improve the calculation of the posterior probabilities.

Green (1995) proposed a new framework for the construction of reversible Markov chain samplers that jump between parameter subspaces of differing dimensionality. Its flexibility has wide applicability including multiple change-point problems. Liang and Wong (1999) applied Evolutionary Monte Carlo (EMC) algorithm which has active features of genetic algorithm (Holland (1975)) in the framework of Markov Chain Monte Carlo (MCMC) to the normal change-point problem.

In this paper, the multiple change-point problems are dealt when the observations are from the parametric distributions. A latent vector is introduced to partition the observations and indicate the change-point positions, and the other parameters are integrated out with an appropriate choice of prior distributions as a Bayesian approach. Especially the EMC algorithm is applied to find the multiple change-points when the number of change-points is unknown. The numerical results including applied to real data show that EMC is a promising approach for both simulation and optimization.

This paper is organized as follows. In section 2 a multiple change-point model is defined with prior distributions for the parameters. Section 3 describes an algorithm-evolutionary Monte Carlo (EMC) and its usefulness as a Monte Carlo sampler. Section 4 presents some numerical results including simulation and real data analysis. Finally section 5 concludes the paper with a discussion of change-point problem and EMC algorithm.

2. Multiple Change-point Model with Normal Errors

A sequence of observations undergoes sudden changes at unknown time points. We can model the process by supposing that there is an underlying sequence of parameters partitioned into contiguous blocks of equal parameter values and the beginning of each block is said to be a change-point. Observations are then assumed to be independent in different blocks given the sequence of parameters.

Let $Z = (z_1, z_2, \dots, z_n)$ denote the independent observation sequence ordered in time. There exists a partition on the set $1, 2, \dots, n$ into blocks such that the sequence follows the same distribution within blocks. That is, the change-points divide the partitions. Introduce a binary vector $x = (x_1, x_2, \dots, x_n)$ with $x_{c_1} = x_{c_2} = \dots = x_{c_k} = 1$ and being 0 elsewhere and $0 = c_0 = c_1 < \dots < c_{k+1} = n$. There are k change-points in the model and k is unknown. The multiple change-point model can be written as follows:

$$z_i \sim f_r(\cdot | \theta_r), \quad c_{r-1} < i < c_r \tag{2.1}$$

for $r = 1, 2, \dots, k+1$ and f_r depends on the parameters $\theta_r \in \Theta$. The parameters change at $c_1 + 1, c_2 + 1, \dots, c_k + 1$. Each c_1, c_2, \dots, c_k are called the change-point.

The observations are assumed to be independently drawn from a Normal distribution $N(\mu, \sigma^2)$ with unknown μ and σ^2 . After the change-point, the normal distribution may shift in both mean and variance.

In a Bayesian analysis, it is necessary to give probability distributions to both the change-points and the parameters. An important practical consideration is that some prior constraint is necessary on the number of change-points and on the amount of variability between blocks. These are parameters about which the data cannot be relied on to be conclusive. The prior distributions are given as follows.

Consider f_r is a normal density parameterized by $\theta_r = (\mu_r, \sigma_r^2)$. Let $\mathbf{x}^{(k)}$ denote a configuration of \mathbf{x} with k change-points. Let $\eta^k = (\mathbf{x}^{(k)}, \mu_1, \sigma_1^2, \dots, \mu_{k+1}, \sigma_{k+1}^2)$ and A_k be the space of models with k change-points, $\mathbf{x}^{(k)} \in A_k$ and $\chi = \cup_{k=0}^n A_k$.

The likelihood function of Z is

$$L(Z|\eta^{(k)}) = \prod_{j=c_n+1}^{c_1} f_1(z_j|\mu_1, \sigma_1^2) \cdots \prod_{j=c_k+1}^{c_{k+1}} f_1(z_j|\mu_{k+1}, \sigma_{k+1}^2).$$

Therefore the log-likelihood function of model $\eta^{(k)}$ is

$$\log L(Z|\eta^{(k)}) = - \sum_{i=1}^{k+1} \left\{ \frac{c_i - c_{i-1}}{2} \log \sigma_i^2 + \frac{1}{2\sigma_i^2} \sum_{j=c_{i-1}+1}^{c_i} (z_j - \mu_i)^2 \right\} \quad (2.2)$$

For a Bayesian analysis, consider the prior distribution for $\eta^{(k)}$ as

$$\pi(\mathbf{x}^{(k)}) = \frac{\lambda^k (n-1-k)!}{\sum_{j=0}^{n-1} \frac{\lambda^j}{j!}}, \quad k = 0, 1, \dots, n-1$$

This prior gives that A_k has a truncated up to $(n-1)$ Poisson distribution with parameter λ and each models in A_k is equally likely. Put an improper prior on μ 's and an inverse-gamma $IG(\gamma, \delta)$ on σ_i^2 's and the priors are independent. Then the log-prior density is

$$\log \pi(\eta^{(k)}) = a_k - \sum_{i=1}^{k+1} \left\{ (\gamma-1) \log \sigma_i^2 + \frac{\delta}{\sigma_i^2} \right\} \quad (2.3)$$

where $a_k = (k+1)(\gamma \log \delta - \log \Gamma(\gamma)) + \log(n-1-k)! + k \log \lambda$. Here γ, δ and λ are fixed hyperparameters. The log posterior of $\eta^{(k)}$ (up to an additive constant) can be obtained by adding (2.2) and (2.3). Integrating out $\mu_1, \sigma_1^2, \dots, \mu_{k+1}, \sigma_{k+1}^2$ from the full posterior distribution and taking a logarithm, we obtain

$$\begin{aligned} \log \pi(\mathbf{x}^{(k)} | Z) = & a_k + \frac{k+1}{2} \log 2\pi \\ & - \sum_{i=1}^{k+1} \left\{ \frac{1}{2} \log(c_i - c_{i-1}) - \log \Gamma\left(\frac{c_i - c_{i-1} - 1}{2} + \gamma\right) \right. \\ & \left. + \left(\frac{c_i - c_{i-1} - 1}{2} + \gamma\right) \log \left[\delta + \frac{1}{2} \sum_{j=c_{i-1}+1}^{c_i} z_j^2 - \frac{\left(\sum_{j=c_{i-1}+1}^{c_i} z_j\right)^2}{2(c_i - c_{i-1})} \right] \right\}. \end{aligned}$$

We can sample from this marginal posterior $\pi(\mathbf{x}^{(k)} | Z)$ by EMC technique and estimate the change-points which have greatest posterior probabilities. In addition to identifying the change-points, we can estimate the probabilities $p(A_k | Z)$ for $k_{\min} \leq k \leq k_{\max}$ where k_{\min} and k_{\max} specify the range of the number of change-points of interest from Liang and Liu (2005).

In a typical application with multiple parameter subspaces of different dimensionality, it will be necessary to devise different types of move between the subspaces. Tierney (1994) calls a hybrid sampler, by random choice between available moves at each transition, in order to traverse freely across the combined parameter space. As usual with Hastings algorithms, the probability of acceptance

is left undefined in this situation. Green (1995) derives an expression of the probability of acceptance which achieves the stated aim of attaining detailed balance within each move type and proposes the reversible jump Markov chain Monte Carlo (RJMCMC). Green (1995) applies RJMCMC to the multiple change-point problem considering the variability of the number of change-points. In this paper, we would like to use the EMC algorithm as a computational tool. In next section, the EMC algorithm is briefly reviewed.

3. Evolutionary Monte Carlo Algorithm

Suppose that we are working with the following Boltzmann distribution,

$$p(x) = \frac{1}{Z} \exp(-H(x)/\tau), \quad x \in \chi, \quad (3.1)$$

where Z is the normalizing constant, $\tau = (\tau_1, \dots, \tau_N)$ is the temperature, and χ is a sample space. Without loss of generality, we assume that χ is compact. In Bayesian statistics, $x = (x_1, \dots, x_N)$ is often a vector of parameters while the fitness function $H(x)$ is the negative of the log-posterior of x . This paper is focused on $H(x)$. In EMC, a different temperature τ_i is attached to each individual x_i , and the temperatures form a ladder with the ordering $\tau_1 > \dots > \tau_N$. EMC has many attractive features of simulated annealing and genetic algorithm into a framework of MCMC. It works well with simulating a population of Markov Chains in parallel where a different temperature is attached to each chain. The population is updated by mutation (Metropolis update in one single chain), crossover (partial states swapping between different chains), and exchange operators (full state swapping between different chains). Thus, EMC is the useful Monte Carlo sampler. Refer to Liang (2000) for more details on these operators. Based on these operators, EMC is summarized as follows.

Given an initial population $x = \{x_1, \dots, x_N\}$. and a temperature ladder $\tau = \{\tau_1, \dots, \tau_N\}$, one iteration of the Markov chain consists of two steps.

1. Apply the mutation or the crossover operator to the population with probabilities q_m and $1 - q_m$, respectively (q_m is the mutation rate).

2. Try to exchange x_i with x_j for N pairs (i, j) with i being sampled uniformly on $\{1, \dots, N\}$ and $j \pm 1$ with probability $w(x_j|x_i)$, where $w(x_{i+1}|x_i) = w(x_{i-1}|x_i) = 0.5$ and $w(x_2|x_1) = w(x_{N-1}|x_N) = 1$.

In the mutation operation, each point of the population is mutated independently. In the crossover operation, about 40% of points are chosen to mate. Note that the parental points are chosen in an iterative way, i.e., each time two

parental points are chosen from the current population which has been updated by the last crossover operation. This operation repeats for $[N/5]$ (the integer part of $N/5$) times. The algorithm has three free parameters, namely N , τ and q_m . The mutation rate q_m can be chosen to achieve a trade-off between the exploration and convergence of the algorithm (Spears (1992)) i.e., q_m is usually set to a large value for a small population to provide the system more opportunities to explore the sample space, and a small value for a large population to force the system to converge quickly, respectively. For example, we fixed $q_m = 0.2$ in all simulations of this paper. The temperature ladder τ can be set such that

$$\text{Var}(H(x_i))\delta^2 = O(1) \quad (3.2)$$

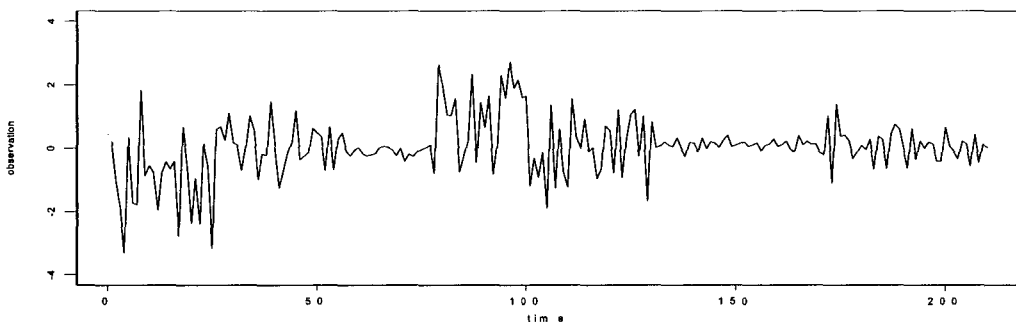
where $\delta = 1/\tau_{i+1} - 1/\tau_i$ i.e., this condition on δ is also equivalent to requiring that the distributions on temperature level τ_i and τ_{i+1} have a considerable overlap. Refer to simulated tempering (Marinari and Parisi (1992)) and parallel tempering (Geyer (1991), Hukushima and Nemoto (1996)).

4. Numerical Examples

4.1. Change-point Estimation with the Simulated Normal Data

In this simulation, we generated 200 independent observations z_1, \dots, z_{26} from $N(-0.5, 1.5)$, z_{27}, \dots, z_{56} from $N(0.25, 0.75)$, z_{57}, \dots, z_{76} from $N(-0.1, 0.16)$, z_{77}, \dots, z_{100} from $N(0.5, 1.5)$, z_{101}, \dots, z_{130} from $N(-0.25, 0.75)$; z_{131}, \dots, z_{170} from $N(0.1, 0.16)$, z_{171}, \dots, z_{200} from $N(-0.01, 0.5)$.

There are 6 change-points in the observation sequence. The time plot is shown in <Figure 1>.



<Figure 1> A Time Series Plot of the Observation Data

We assume that there are no more than 99 change-points in the observation sequence, and the change-points only occur after even observations, i.e., $d=99$, and $c_i \in \{2,4,\dots,198\}$. In the simulation, we set $\lambda=6$, $\gamma=0.05$ and $\delta=0.05$, which corresponds to a vague prior on σ_i^{-2} . The population size was 20 and the temperature levels were equally spaced between 10 and 1. A three-point mutation and an adaptive crossover were used with $p_m=0.2$, $p_0=0.01$, $p_1=0.02$ and $p_2=0.04$. EMC was run for 7000 iterations (the first 1000 iterations were discarded for the burn-in process). The overall acceptance probabilities of mutations, crossover and exchange operations were 0.58, 0.18 and 0.70, respectively. <Table 1> lists the 10 models with the largest log-posterior values found using EMC sampler. Note that the true change-point model was ranked as 1 in log-posterior values among all sampled models. Other results of the run are shown in <Figure 2>.

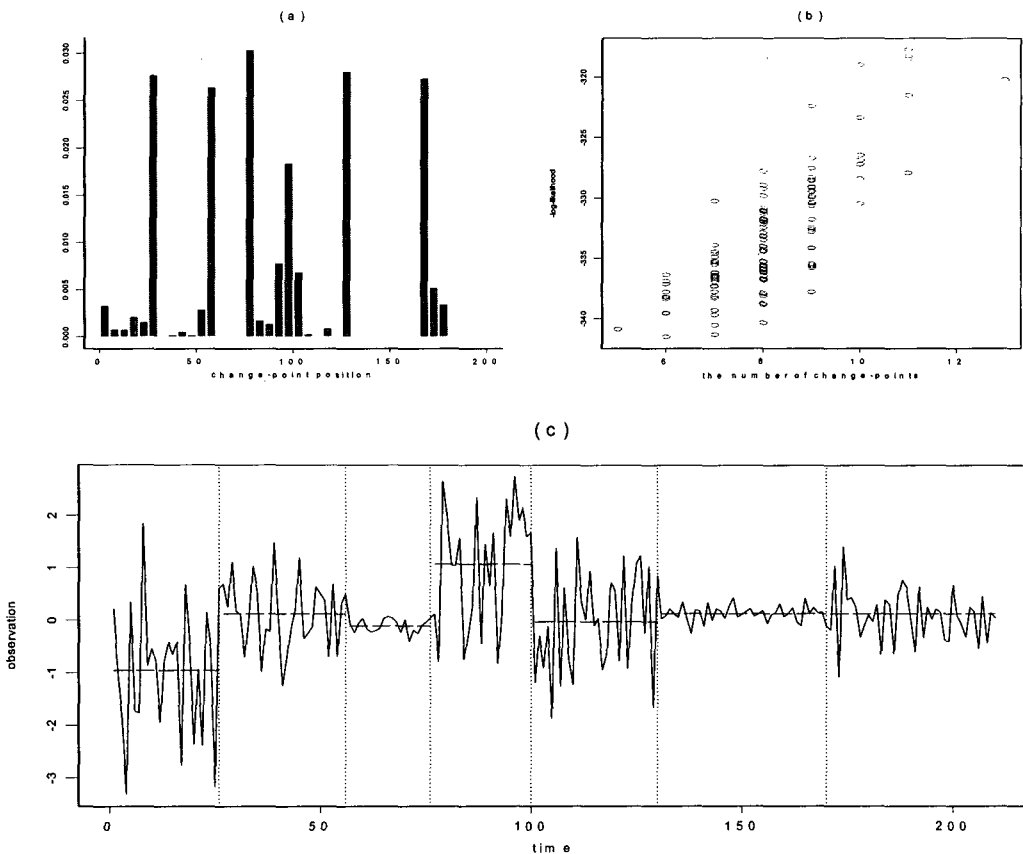
<Figure 2(b)> shows that the Maximum posterior model was sampled. In addition, many models with close log-posterior were sampled. <Figure 2(a)> is the histogram of the posterior of the identified change-point positions. It shows that most of the models with high posterior probabilities include 6 change-points. This figure shows that the six most likely change points are around 26, 56, 76, 100, 130 and 170. Note that there is much uncertainty around the fourth cluster of the histogram bars in <Figure 2(a)>. This is consistent with the results shown in <Table 1>: the models with the first three change points at 26, 56 and 76, and the third change point being around 98 (92 - 100) have very close log-posterior values. <Figure 2(c)> shows the Maximum posteriori estimate of the change patterns of the data. The maximum posteriori estimate of the change points is (26, 56, 76, 100, 130, 170), which is exactly the same with the true model. This result is strongly supported by the data. The EMC simulation seems to work well in identifying some plausible models with multiple change-points according to the posterior likelihood.

<Table 1> The 10 models with the largest log-posterior values sampled by EMC. The underlined model is true and is ranked 1 in log-posterior values among all models sampled by EMC. The second column shows the differences of the log-posterior values of the models from the true model.

No	Log-posterior*	Number of change-points	Change patterns
1	0.0000	6	(26, 56, 76, 100, 130, 170)
2	-0.2003	7	(26, 56, 76, 94, 100, 130, 170)
3	-0.6526	5	(26, 56, 76, 130, 170)

<Table 1> Continued

4	-0.9902	7	(26, 56, 76, 100, 130, 170, 174)
5	-1.1797	8	(26, 56, 76, 94, 100, 130, 170, 174)
6	-1.8435	7	(26, 56, 76, 78, 100, 130, 170)
7	-1.9585	6	(26, 56, 76, 102, 130, 170)
8	-2.0017	6	(26, 56, 76, 98, 130, 170)
9	-2.2120	7	(26, 56, 76, 92, 100, 130, 170)
10	-2.6722	8	(26, 56, 76, 94, 100, 130, 170, 176)

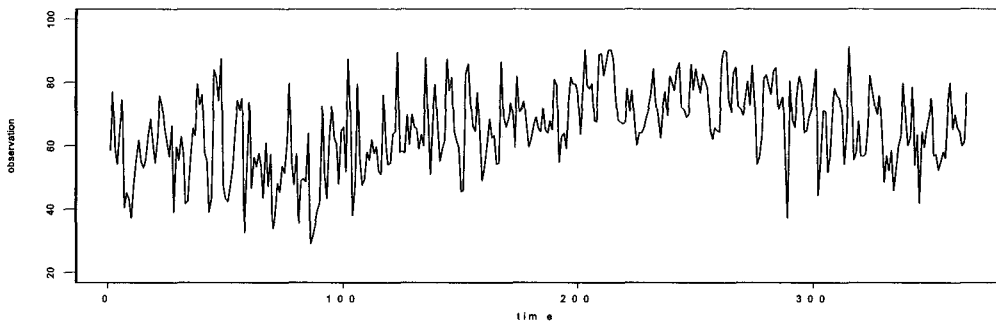


<Figure 2> The simulation results of EMC for the change-point example: (a) The posterior histogram of change-point positions sampled by EMC; (b) The scatter plot of the negative log-posterior values among all models sampled by EMC vs the number of change-points (c) A comparison of the Maximum posteriori estimate of the change-point positions and the true change-point positions (the vertical (dotted) lines indicated the change-point positions identified by the Maximum posteriori model, the horizontal (dashed) lines indicate the mean value of observations separated by change-point positions of the true model).

4.2. Change-point Estimation with Humidity Data of Seoul in 1999

Meteorological data and air pollutants were measured hourly from January 1, 1999 to December 31, 1999 at 27 sites monitored by Korean Ministry of Environment. The hourly 24 averages on 27 sites in Seoul were computed. Meteorological data including temperature, precipitation, relative humidity, wind speed, and wind direction were obtained from Korea Meteorological Administration.

We would like to estimate the several change-points of humidity in a year to see the change trend in Seoul. Bayesian models express appropriate uncertainty about the change-points and give a plausible description about the observations. The time plot is shown in <Figure 3>. Since each observation is an averaged value, it can be assumed to follow the normal distribution by the central limit theorem. Here we didn't consider the dependency as a time series. Considering time series change-point model is left as the next research.



<Figure 3> A Time Series Plot of the Humidity in 1999.

For a Bayesian analysis, the hyperparameters should be decided from the previous information. We assume that there are no more than 181 change-points in the observation sequence, and the change-points only occur after even observations, i.e., $d = 181$, and $c_i \in \{2, 4, \dots, 362\}$. In the simulation, we used the exactly same setting as that used for above Normal data, except that we set $\lambda = 3$, $\gamma = 1$ and $\delta = 0.003$ which corresponds to a vague prior on σ_i^{-2} . The overall acceptance probabilities of mutations, crossover and exchange operations were 0.41, 0.05 and 0.57, respectively. Table 2 lists the 10 models with the largest log-posterior values found by EMC. Other results of the run are shown in <Figure 4>.

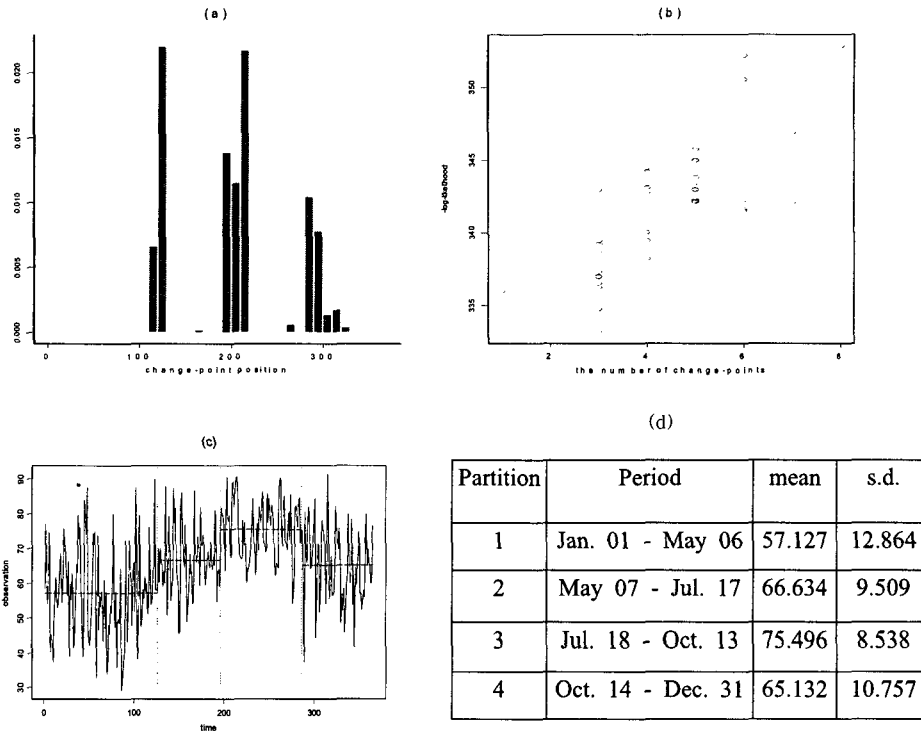
Many models were sampled for which the log-posterior values were very close. <Figure 4(a)> is the posterior histogram of the identified change-point positions.

It shows that most of the models with high posterior probabilities include three change-points. This figure shows that the three most likely change points are around 120, 196 and 286. Note that there is much uncertainty around all clusters of the histogram bars in <Figure 4(a)>. We used Bayesian Information Criterion (BIC), a popular criterion for model selection. <Table 2> indicates that the models with the first one change point being around 120 (114-126), the second change-point being around 200 (194-210) and the third change-point being around 300 (284 - 310) have very close log-posterior values. <Figure 4(c)> shows the Minimum BIC estimate of the change patterns of the data. The Minimum BIC estimate of the change points is (126, 196, 286); i.e., (May 06, Jul. 15, Oct. 13). Our analysis supports changes in humidity at the start of May, mid July, mid October which reflect the trend of humidity in Seoul, Korea. Considering the climate in Korea, July and August in the summer season have higher temperature and humidity than other days in a year. And in the winter and early spring, the humidity is lower than the summer season.

<Table 2> The 10 models with the largest log-posterior values sampled by EMC. The second column shows BIC (Bayesian Information criterion) values for model selection. The fifth column indicates the log-posterior values of the posteriori models.

No	BIC*	# change-points	Change patterns	Log-posterior*
1	502.5761	3	126(5/06), 198(7/17), 286(10/13)	-336.2414
2	502.5764	3	124(5/04), 198(7/17), 286(10/13)	-336.5378
3	502.6183	3	116(4/26), 196(7/15), 286(10/13)	-333.1895
4	502.6188	3	114(4/24), 196(7/15), 286(10/13)	-334.7499
5	512.9199	3	126(5/06), 198(7/17), 284(10/11)	-336.9836
6	512.9202	3	124(5/04), 198(7/17), 284(10/11)	-337.2800
7	512.9213	3	118(4/28), 198(7/17), 284(10/11)	-336.9681
8	514.3762	4	124(5/04), 126(5/06), 198(7/17), 286	-337.4471
9	514.4183	4	114(4/24), 116(4/26), 196(7/15), 286	-335.2888
10	1408.135	1	124(5/04)	-335.9760

Therefore, those estimated change-points explain the humidity situation in Seoul. <Figure 4(c)> indicates that the data have a great deal of noise, and it is very difficult to decide if an observation comes from one or the other of the two neighboring distributions.



<Figure 4> The simulation results of EMC for the change-point example: (a) The posterior histogram of change-point positions sampled by EMC; (b) The scatter plot of the negative log-posterior values among all models sampled by EMC vs the number of change-points (c) A plot of the Minimum BIC estimates of the change-point positions (the vertical (dotted) lines indicate the change-point positions, and the mean values of observations separated by change-point positions identified by the Minimum BIC model) (d) The table of the mean and standard deviation (s.d.) values of observations separated by change-point positions of the Minimum BIC model.

5. Discussion

The multiple change-point estimation problem was solved using an Evolutionary Monte Carlo algorithm with the normal observations with the Bayesian methodology. The methodology also can be extended to observations from other distributions and thus offers solutions to multiple change-point problems. For calculation of the posterior probability of the multiple change-points,

EMC algorithm works in the way that the population is updated by mutation, crossover and exchange operators, and the updates are accepted or rejected according to the Metropolis rule. The EMC algorithm works very well in change-point identification in the Bayesian setup.

The effectiveness of the algorithm is summarized by two things. First, the algorithm has incorporated the learning ability of the genetic algorithm by evolving with crossover operators. The "learning" or adaptive nature of the algorithm plays an important role in the simulation of EMC, especially in the early stage of the simulation. EMC works on a population of Markov chains. Second, the algorithm has incorporated the attractive feature of simulated annealing by sampling along a temperature ladder. The simulation at high temperature can help the system escape from local minima, and this substantially accelerates the mixing of the system. It is known that selection is an essential ingredient in all evolutionary processes.

In this paper, EMC algorithm is adopted for the Bayesian change-point identification, it is already applicable in many other areas such as the statistical model selection problem (George and McCulloch (1993), Brown and Vannucci (1998)) and real parameter problems (Liang and Wong (1999b)). All computational results show the effectiveness of the crossover operator, and thus EMC algorithm.

Finally, the multiple change-point problems in other situations such as non-parametric Bayesian model, or time series model can be considered and further research is expected in the change analysis.

Acknowledgements

We thank Prof. Kim Ho for providing and helping to arrange the meteorological data.

References

- [1] Barry, D. and Hartigan, J.A. (1993). A Bayesian Analysis for Change Point Problems. *Journal of the American Statistical Association*, Vol. 88, 309-319.
- [2] Brown, P.J. and Vannucci, M. (1998). Multivariate Bayesian Variable Selection and Prediction. *Journal of Royal. Statistics Society B*, Vol. 60, 627-641.
- [3] Chernoff, H. and Zacks, S. (1964). Estimating the Current Mean of a Normal Distribution which is subject to Changes in Time. *Annals of*

- Mathematical Statistics*, Vol. 35, 999-1018.
- [4] Chen, J. and Gupta, A.K. (2000). Parametric Statistical Change Point Analysis, *Birkhauser*. New York.
- [5] George, E.I. and McCulloch, R.E. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, Vol. 88, 881-889.
- [6] Geyer, C.J. (1991). Markov chain Monte Carlo Maximum Likelihood. In Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface (Edited by E.M. Keramigas), 156-163. Interface Foundations. Fairfax.
- [7] Green, P.J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, Vol. 82, 711-732.
- [8] Holland, J.H. (1975). Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor.
- [9] Hukushima, K. and Nemoto, K. (1996). Exchange Monte Carlo Method and Application to Spin Glass Simulations. *Journal. Physics. Society. in Japan*. Vol. 65, 1604-1608.
- [10] Liang, F. and Wong, W.H. (1999). Real Parameter Evolutionary Monte Carlo with Applications in Bayesian Neural Networks. Technical Report, Department of Statistics and Applied Probability, NUS.
- [11] Liang, F. and Wong, W.H. (2000) Evolutionary Monte Carlo Applications to Cp Model Sampling and Change Point Problem. *Statistica Sinica*, Vol. 10, 317-342.
- [12] Liang, F. and Liu, C. (2005). Efficient MCMC Estimation of Discrete Distributions. *Computational Statistics & Data Analysis*, Vol. 49, 1039-1052.
- [13] Marinari, E. and Parisi, G. (1992). Simulated Tempering: a New Monte Carlo Scheme. *Europhysics. Letters*, Vol. 19, 451-458.
- [14] Spears, W.M. (1992). Crossover or mutation? In Foundations of Genetic Algorithms 2 (Edited by L. D. Whitley). Morgan Kaufmann, San Mateo.
- [15] Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions (with discussion). *Annals of Statistics*, Vol. 22, 1701-1762.
- [16] Yao, Y.C. (1984). Estimation of a Noisy Discrete-time Step Function: Bayes and Empirical Bayes Approaches. *Annals of Statistics*, Vol. 12, 1434-1444.