

Research Memorandum No.556

Apr. 18, 1995

On the resampling scheme in the filtering
procedure of the Kitagawa Monte Carlo Filter

T. Higuchi

The Institute of Statistical Mathematics, Tokyo 106, Japan

Research Memorandum No.556

Apr. 18, 1995

On the resampling scheme in the filtering
procedure of the Kitagawa Monte Carlo Filter

T. Higuchi

The Institute of Statistical Mathematics, Tokyo 106, Japan

On the resampling scheme in the filtering procedure of the Kitagawa Monte Carlo Filter

T. Higuchi

The Institute of Statistical Mathematics, Tokyo 106, Japan

April 18, 1995

Key words and phrases: Bayesian approach, Generalized state space model (GSSM), Non-Gaussian time series model, Genetic Algorithm (GA), Resampling

Abstract.

The major objective of this paper is to address the close relationship between the Genetic algorithm, which is usually used to get an optimal solution for a nonlinear multiparameter optimization problem, and Kitagawa's Monte Carlo filter, (Kitagawa 1995), that was developed to manage a large dimensional state vector in a generalized state space model. The Monte Carlo filter is interpreted from the perspective of the Genetic algorithm. On the other hand, an essential part of the Genetic algorithm can be viewed as a special case of the Monte Carlo filter, and then a Bayesian interpretation of the Genetic algorithm leads to an examination of the performance of several ideas to improve the optimization. Based on an analogy between the Genetic algorithm and the Monte Carlo filter, we investigate a resampling scheme involved in the filtering procedure of the Monte Carlo filter. Several schemes for resampling are applied to a simple problem and their performance is evaluated in terms of an estimation of the likelihood.

1 Introduction

The modern Bayesian approach to data analysis is characterized by using an explicit probability model to incorporate *a priori* knowledge such as analyst's inspection, physical reasoning, well-known results, and sometimes his/her assumptions. The basic principle

behind this approach is to extract as much useful information from the given data, even when the obtained data are contaminated with various noises including outliers, artificial noise, and missing values. A procedure referred to as the Bayesian approach has a long historical background and its basic concept may be traced back to a suggestion given by Whittaker (1923) to solve the *smoothing problem*. The recent rapid progress in electronic computing power has encouraged us to apply the Bayesian approach to increasingly complicated problem. For example, although Hunt (1973) treated Bayesian image restoration, the seminal paper by Geman and Geman (1984) stimulated the more recent interest in Bayesian image restoration, (or Bayesian image analysis), (see Besag (1986) and the special issue of *Applied Statistics* edited by Mardia (1989)), and now involves many researchers in the Bayesian approach.

Preceding the Geman and Geman paper, Akaike (1980) successfully used a Bayesian approach in the modeling of nonstationary time series and stimulated its applicability to a wide variety of data analysis problems. Since this approach facilitates fitting a large amount of the time series model parameters, it allows us to construct a flexible model for a time series. As seen in many successful applications to various field data, the Bayesian approach even in a framework of the linear and Gaussian model is still useful (see the references in Gersch and Kitagawa (1988)). Nevertheless, it is inevitable to treat a non-linear and non-Gaussian model in the Bayesian approach, because we frequently face the situations where the prior and/or data distribution should be non-Gaussian and where parameters are mutually related with high non-linearity.

The non-Gaussian and non-linear time series model has been realized by an extension of the standard Kalman filter state space model, (Kalman 1960), in what is referred to as the extended Kalman filter (EKF), (Anderson and Moore 1979). Kitagawa (1987) realized a general formulation for a state space analysis of the non-Gaussian and non-linear model, that he referred to as the *generalized state space model* (GSSM). In the context of a nonlinear smoothing problem, Kitagawa (1991) demonstrated the superiority of the GSSM model approach to the EKF approach. The GSSM modeling is based on the intrinsic structure of time series in such a way that an observation is sequentially obtained and that, as a general rule, a prior knowledge is numerically described by a local constraint between adjacent data points.

The GSSM approach which enables us to use recursive formulations together with an evaluation of the likelihood, has a unified framework. However, the GSSM still requires, for its practical application to data analysis, computationally extensive and difficult tasks due to the relatively high dimensionality of the state vector. In a case with a low dimension (for an example $k \leq 2$), a simple but flexible approach to approximate any conditional distribution by a first order spline or simple step functions is feasible. Several applications of this approach can be found in Kitagawa (1987, 1991). Also West and Harrison (1989), Carlin *et al.* (1992), Fahrmeir (1992), Frühwirth-Schnatter (1994a, b) and references therein.

For $2 < k$ an implementation along this primitive approach needs the complicated and sophisticated treatments that is rarely practicable because of the inherent programming complexity. In addition, there is an apparent problem due to the storage limitation (Fahrmeir (1992), Frühwirth-Schnatter (1994a)). To handle the high dimensionality Kitagawa (1993) introduced a Monte Carlo method for filtering and smoothing. We refer to it as the *Monte Carlo Filter* (MCF). A treatment of the high dimensionality within the GSSM framework is not an issue in the present paper.

We begin with a brief explanation of the methodology of the MCF together with an introduction to GSSM. Then we investigate some relationships between the MCF and the *Genetic Algorithm*, (GA), the adaptive search procedure inspired by evolution (Holland (1975), Goldberg (1989), Davis (1991), Whitley (1994)). Making use of an analogy between the MCF and the GA, we interpret the MCF from the perspective of the GA in Section 3. Section 4 discusses several ideas to improve the original MCF algorithm in terms of an evaluation of the likelihood. In particular we concentrate on the resampling scheme that is an essential part of the filtering procedure of the MCF, and examine typical resampling schemes by considering a very simple problem.

2 Generalized State Space Model (GSSM)

2.1 Generalization of the state space model

The generalized state space model (GSSM) is defined by

$$\begin{aligned}
\text{system model} \quad x_n &= g(x_{n-1}, v_n) \\
\text{observation model} \quad y_n &= h(x_n, w_n),
\end{aligned} \tag{1}$$

where x_n and y_n are k -dimensional *state vector* and univariate observation at discrete-time of n . v_n and w_n are l -dimensional and 1-dimensional white noise sequences with unknown probability density function $q(v|\lambda_s)$ and $r(w|\lambda_o)$, respectively. λ_s and λ_o are parameter vectors, if necessary, for describing $q(\cdot)$ and $r(\cdot)$, respectively, and called *hyperparameter* in Bayesian terminology (Lindley and Smith (1972)). See Kitagawa (1987, 1991, 1993) for examples of the GSSM. Any dynamic generalized linear model (DGLM) (e.g. West *et al.* (1985), West and Harrison (1989), Fahrmeir (1992), Frühwirth-Schnatter (1994a)) may be expressed as a special case of the GSSM.

The time series model expressed by (1) implies that the conditional distributions satisfy the following Markov properties:

$$\begin{aligned}
p(x_n|X_{n-1}, Y_{n-1}) &= p(x_n|x_{n-1}) \\
p(y_n|X_n, Y_{n-1}) &= p(y_n|x_n),
\end{aligned} \tag{2}$$

where X_n and Y_n are $X_n \equiv [x_1, \dots, x_n]$ and $Y_n \equiv [y_1, \dots, y_n]$, respectively. Assume that $y_n = h(x_n, w_n)$ has the following implicit function: $w_n = h^{-1}(y_n, x_n)$. Then the conditional density $p(y_n|x_n)$ is replaced by

$$p(y_n|x_n) = r(w_n) \left| \frac{dh^{-1}}{dy_n} \right| = r(w_n)w'(y_n, x_n), \tag{3}$$

where we set $w'(y_n, x_n) = |dh^{-1}/dy_n|$.

2.2 Recursive formulas

Equation (1) with (2) and (3) yields useful recursive formulas for the estimation of the conditional probability distribution of the state vector x_i given data Y_j , $p(x_i|Y_j)$, which are formed by a set of the following three steps: *prediction*, *filtering*, and *smoothing* (e.g. Kitagawa (1987), Harvey (1989)).

prediction

$$p(x_n|Y_{n-1}) = \int p(x_{n-1}|Y_{n-1}) \cdot p(x_n|x_{n-1})dx_{n-1} \tag{4}$$

filtering

$$p(x_n|Y_n) = \frac{p(y_n|x_n) \cdot p(x_n|Y_{n-1})}{p(y_n|Y_{n-1})} = \frac{p(y_n|x_n) \cdot p(x_n|Y_{n-1})}{\int p(y_n|x_n) \cdot p(x_n|Y_{n-1}) dx_n} \quad (5)$$

smoothing

$$p(x_n|Y_N) = p(x_n|Y_n) \cdot \int \frac{p(x_{n+1}|Y_N) \cdot p(x_{n+1}|x_n)}{p(x_{n+1}|Y_n)} dx_{n+1}. \quad (6)$$

where N is the total number of data points, and an initial distribution $p(x_0|Y_0)$ is defined *a-priori*.

For the fixed (distribution) forms of $q(\cdot)$ and $r(\cdot)$ the optimal values of hyperparameters, λ^* , are selected by assessing the log-likelihood $\log p(Y_N|\lambda)$ (Good (1965)),

$$\begin{aligned} l(\lambda) = \log p(Y_N|\lambda) &= \log \prod_{n=1}^N p(y_n|Y_{n-1}, \lambda) \\ &= \sum_{n=1}^N \log p(y_n|Y_{n-1}, \lambda) \end{aligned} \quad (7)$$

where $p(y_n|Y_{n-1}, \lambda)$ is the conditional distribution of y_n , given data Y_{n-1} , and $\lambda \equiv [\lambda_s^T, \lambda_o^T]^T$. We note that $p(y_n|Y_{n-1}, \lambda)$ is the one-step-ahead predictive value which appears in (5). The best function forms among competing candidates are chosen so as to maximize (7) in a same manner.

3 Monte Carlo Filter (MCF)

Here we first give a brief introduction to MCF with emphasis on clarifying the close relationship between the MCF and GA algorithms. That follows with an analogy of the MCF with the GA based on an explanation of the MCF.

3.1 Basic idea and algorithm

A basic idea of the MCF (Kitagawa (1993)) is summarized as follows:

- We approximate an arbitrary conditional probability density function by many of its realizations.
- At each n , the number of realizations, m , is fixed to make it easy to implement the MCF algorithm.

For example, we express $p(x_n|Y_{n-1})$ by a set of m realizations: $Z_P^{(n)} \equiv \{z_{P,i}^{(n)}|i = 1, \dots, m\}$. Similarly, $p(x_n|Y_n)$ is approximated by a set of m realizations: $Z_F^{(n)} \equiv \{z_{F,i}^{(n)}|i = 1, \dots, m\}$. The subscripts, P and F , stand for the prediction and filtering, respectively. The superscript (n) , denotes a discrete-time instant n .

The recursive calculations of (4) and (5) are respectively defined in the following manner:

prediction

Realize $Z_P^{(n)}$ each element of which is obtained by $z_{P,i}^{(n)} = g(z_{F,i}^{(n-1)}, v_i^{(n)})$.

filtering

Realize $Z_F^{(n)}$ by resampling $z_{P,i}^{(n)}$ with probability proportional to

$$\begin{aligned} p(z^{(n)} = z_{P,i}^{(n)}|Y_n) &= \frac{r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)}) \cdot p(z^{(n)} = z_{P,i}^{(n)}|Y_{n-1})}{\sum_{i=1}^m r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)}) \cdot p(z^{(n)} = z_{P,i}^{(n)}|Y_{n-1})} \\ &= \frac{r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)}) \cdot (1/m)}{\sum_{i=1}^m r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)}) \cdot (1/m)} \\ &= \frac{r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)})}{\sum_{i=1}^m r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)})} \end{aligned} \quad (8)$$

Here $v_i^{(n)}$ is a realization sampled from $q(\cdot)$ and $w_i^{(n)} = h^{-1}(y_n, z_{P,i}^{(n)})$. In the MCF the value of $p(y_n|Y_{n-1})$ appearing in (5) is approximated by $(1/m) \sum_{i=1}^m r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)})$. It is worthy to note that when $w'(\cdot)$ does not depend on x_n , $w'(y_n, z_{P,i}^{(n)})$ in (8) need not be calculated. For example, there is no dependency of x_n on $w'(\cdot)$ for DGLM.

In the GSSM framework, a final or smoothed estimation of the state vector is usually given through (6). In the MCF Kitagawa (1993) proposed two alternative formulas for the smoothing algorithm: storing the state vector and a two-filter formula (Kitagawa (1994)). It is not feasible to apply the former as it is, and so a modification is necessary for a workable algorithm. The latter approach is essentially the same as the *Baum-Welsh* algorithm with the forward-backward algorithm in HMM (Baum *et al.* (1970), Huang *et al.* (1990)).

3.2 Analogy with a genetic algorithm

In this subsection, we address the idea that the MCF and the GA are quite similarly structured. For example, $z_{P,i}^{(n)}$ and $z_{F,i}^{(n)}$ identified as *particles* in the MCF is considered

as the *string* at n -th generation in GA. $z_{F,i}^{(n-1)}$ and $z_{P,i}^{(n)}$ are regarded as the parent and offspring, respectively, because $z_{F,i}^{(n-1)}$ creates $z_{P,i}^{(n)}$ which undergoes selection. From this point, it follows that the MCF is close to the canonical GA (Whitley (1994)) (sometimes referred as the simple GA (SGA) (Goldberg (1989))) among many algorithms based on a population-based search.

The filtering procedure of the MCF exactly corresponds to the *selection* (or reproduction) of GA by regarding $r(\cdot)w'(\cdot)/m$ as the evaluation function. Accordingly in the GA, the equivalent of $r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)}) / \sum_{i=1}^m r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)})$ can be called the fitness of the string $z_{F,i}^{(n)}$. It should be noticed that whereas $r(\cdot)w'(\cdot)/m$ is obviously dependent of time due to the presence of data y_n , the evaluation function in GA without any modification such as *scaling* (stretching) (e.g., Goldberg (1989)) is independent of n .

$p(y_n|Y_n)$ approximated by $(1/m) \sum_{i=1}^m r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)})$ in the MCF is therefore regarded as the *population fitness at the n -th generation* in GA. The log likelihood in the MCF, $l^{\text{MCF}}(\lambda)$, is the logarithm of the value which we call *family fitness*:

$$\begin{aligned} l^{\text{MCF}}(\lambda) &= \sum_{n=1}^N \log p(y_n|Y_{n-1}, \lambda) \\ &= \sum_{n=1}^N \log (\text{family fitness}). \end{aligned} \quad (9)$$

Along the line of this interpretation, the maximum likelihood principle is interpreted as a rule to choose the model that maximizes the family fitness under a circumstance of given data Y_N .

Now consider the GA from the perspective of the MCF. The scaling procedure which is frequently used to keep appropriate level of competition throughout a simulation can be viewed as a dynamic control of hyperparameters for $r(\cdot)$. Namely in this case we consider the time series model with hyperparameter λ_0 being dependent on time.

An essential idea to involve the operations of the *crossover* and *mutation* in the GA is to escape from local maximum of the optimized function. We note that these may be interpreted from the Bayesian point of view. In particular, a fluctuation caused by a mutation is exactly regarded as the system noise which stems from a non-Gaussian probability density function. When we represent a model using a binary code and use the canonical GA, a closed form of this density function can be obtained.

In the MCF there exists an apparent movement of the population at each step which results from the systematic behavior of the particle driven by $g(\cdot)$ with $v_n \equiv 0$ in the system model. The GA disfavors such drift in model space that is independent of any performance of optimization, called *genetic drift*. The correspondence between MCF's and GA's terminology is summarized in Table 1.

4 Resampling scheme of the filtering

4.1 Selection scheme

One of the advantages of using the MCF for a solution of GSSM is to economize on the computer memory space necessary for approximating the conditional probability density function. On the other hand, the MCF requires the evaluation of the likelihood which plays the critical role in specifying the model. The easiest way to deal with this problem is an enormous increase in the number of particles m . Another candidate for this problem is the use of parallel processing in which l^{MCF} is given by taking an average of the value of the log-likelihood at the i th parallel processor, l_i^{MCF} ($i = 1, \dots, N_p$). N_p is the number of the parallel processing. We note that the MCF is quite amenable to a simple parallel processing.

Here, we focus our attention on the resampling scheme that is an essential part of the filtering procedure of the MCF. As mentioned above, the resampling form appearing in the filtering of the MCF is identical to one in the selection procedure of the GA. Consequently we can choose the resampling scheme best suited for estimating the log-likelihood, from the several different ways to do selection (Goldberg (1989), Whitley (1994)). Here we compare three typical selection methods among selection alternatives classified by Brindle (1981):

1. Deterministic sampling
2. Stochastic sampling with replacement (Roulette wheel selection)
3. Remainder stochastic sampling with replacement

The resampling method known as *stochastic universal sampling* (Baker (1987) is employed here to implement the remainder stochastic sampling efficiently (Whitley (1994)).

4.2 Application of the MCF to the simplest model

We discuss three resampling methods in the context of the following very simple model of a time series:

$$\begin{aligned} \text{system model} \quad t_n &= t_{n-1} + v_n \\ \text{observation model} \quad y_n &= t_n + w_n, \end{aligned} \tag{10}$$

where t_n is a trend component at time of n , t_n , thereby setting the 1-dimensional state vector $x_n = [t_n]$, and v_n is the 1-dimensional white noise sequences.

Figure 1 shows the artificially generated data which is obtained by adding i.i.d. white Gaussian noise with a variance of 0.1 to the step function with a jump of 1 at $n = 51$, when $N = 100$. We consider the simplest case in which both $q(\cdot)$ and $r(\cdot)$ are Gaussian. For this case, the Kalman filter yields a ‘‘true’’ value of the likelihood. The true value of the log-likelihood based on the Kalman filter is specified by l^{true} henceforth in which we omit λ for simplicity. In this case λ^* can be easily obtained by maximizing the likelihood that has a very simple form as a function of τ^2 , the ratio of the variance of the system noise to that of the observation noise. The solid curve is obtained by applying the fixed interval smoother with $\tau^{2*} = 0.25$.

Figure 2 shows the log-likelihood of y_n for the model with τ^{2*} , $\log p(y_n|Y_{n-1}, \lambda^*)$. Here we denote $\log p(y_n|Y_{n-1}, \lambda)$ by $l^{\text{true}(n)}$. Correspondingly, an approximation by the MCF to $\log p(y_n|Y_{n-1}, \lambda)$ is denoted by $l_i^{\text{MCF}(n)}$, where i indicates the i th trial. It is seen that $l^{\text{true}(n)}$ takes a minimum value at $n = 51$, indicated by the arrow, where there is a discontinuous change in the given trend.

We concentrate on the discrepancy of l_i^{MCF} from l^{true} , Δl_i , as well as $\Delta l_i^{(51)} = l_i^{\text{MCF}(51)} - l^{\text{true}(51)}$. The reason why we pay special attention to $\Delta l_i^{(51)}$ is that a sparse distribution of $z_{p,i}^{(51)}$ around y_{51} could result in poor approximation of $l_i^{\text{MCF}(51)}$ to $l^{\text{true}(51)}$. Figure 3(a) shows a distribution of $(\Delta l_i, \Delta l_i^{(51)})$ for a case where the deterministic sampling is adopted to implement the resampling in the filtering procedure of the MCF. The number of trials is $N_p = 5000$ and the particle number is fixed $m = 1000$ for each trial. The vertical and horizontal lines indicate the average values of Δl_i and $\Delta l_i^{(51)}$, $\langle \Delta l_i \rangle$ and $\langle \Delta l_i^{(51)} \rangle$, respectively. We call $\langle \Delta l_i \rangle$ the bias. An apparent positive correlation seen in this figure (correlation coefficient 0.737) supports our conjecture that the poor approximation of the

MCF to $l^{\text{true}(51)}$ is significant in controlling Δl_i . The straight line obtained by the least squares fit to these plots has a slope of 0.265. From these results, it follows that there remains a systematic error in approximation of $l^{\text{MCF}(n')}$ for $51 < n'$. Namely $Z_P^{(51)}$ with negative $\Delta l_i^{(51)}$ leads $Z_P^{(52)}$ giving negative $\Delta l_i^{(52)}$ through $Z_F^{(51)}$.

Figures 3(b) and (c) demonstrate a distribution of $(\Delta l_i, \Delta l_i^{(51)})$ for the roulette wheel selection and for the stochastic universal sampling, respectively. As in Figure 3(a), we can see the same tendency such that Δl_i depends mainly on $\Delta l_i^{(51)}$. Whereas the sign of the bias is common to three cases ($\langle \Delta l_i \rangle < 0$), its magnitude for the deterministic sampling is apparently larger than others. It is clearly seen that plots of $(\Delta l_i, \Delta l_i^{(51)})$ for the Roulette wheel selection show a diffuse distribution compared with plots for the stochastic universal sampling. Table 2 summarizes the bias, standard deviation of Δl_i , correlation coefficient, and slope of the line using the least squares fit, for each resampling scheme. Based on a simple statistics, we suggest that the stochastic universal sampling is most suitable for the resampling necessary for the filtering procedure of the MCF.

5 Concluding remarks

In this article we have discussed close relationship between the GA and the MCF. First the MCF and GA algorithms were shown to be similarly structured. The MCF was then interpreted from the perspective of the GA. Next the GA was cast in a Bayesian framework to yield the interpretation of crossover and mutation as the system noise. Also, some ideas to improve the optimization performance of the GA were explained. Table 1 summarized the analogy between the MCF and GA.

By making use of this analogy, numerical experiments were conducted to investigate the resampling schemes of the MCF that is identical to the selection scheme of the GA. Based on numerical experience, we recommended the stochastic universal sampling for the resampling in the filtering procedure of the MCF, because among typical selection schemes, it provides us with a good estimate of the log-likelihood. This recommendation is supported by the fact that Δl_i for the stochastic universal sampling has the smallest variance with least biased log-likelihood value.

Acknowledgement

We would like to thank Prof. Will Gersch of University of Hawaii for his valuable and useful comments on the manuscript.

References

- Akaike, H. (1980). Likelihood and the Bayes procedure (with discussion), in *Bayesian Statistics*, edited by J. M. Bernardo, M. H. De Groot, D. V. Lindley, and A. F. M. Smith, pp. 143–166, University Press, Valencia, Spain.
- Anderson, B. D. O. and Moore, J.B. (1979). *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, N.J.
- Baker, J. (1987). Reducing bias and inefficiency in the selection algorithm, In *Proceedings of the Second International Conference on Genetic Algorithms and Their Applications*, ed. J. Grefenstette, Lawrence Erlbaum.
- Baum, L.E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *Ann. Math. Stat.*, **41**, 164–171.
- Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion), *J. R. Statist. Soc. B*, **48**, 259–302.
- Brindle, A. (1981). *Genetic algorithm for function optimization*, Unpublished doctoral dissertation, University of Alberta, Edmonton.
- Carlin, B.P., Polson, N.G. and Stoffer, D.S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling, *J. Am. Statist. Assoc.*, **87**, No. 418, 493–500.
- Davis, L.D. (1991). *Handbook of Genetic Algorithm*, Van Nostrand Reinhold, New York.
- Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear model, *J. Am. Statist. Assoc.*, **87**, No. 418, 501–509.
- Frühwirth-Schnatter, S. (1994a). Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering, *Statistics and Computing*, **4**, 259–269.
- Frühwirth-Schnatter, S. (1994b). Data augmentation and dynamic linear models, *J. Time Series Analysis*, **15**, No. 2, 183–202.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.*, **6**, 721–741.
- Gersch, W. and Kitagawa, G. (1988). Smoothness priors in time series, in *Bayesian Analysis of Time Series and Dynamic Models*, edited by J.C. Spall, pp. 431–476, Marcel Dekker,

- INC., New York and Basel.
- Goldberg, D.E. (1989). *Genetic Algorithm in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company, Massachusetts.
- Good, I. J. (1965). *The Estimation of Probabilities*, Cambridge, Mass.:MIT Press.
- Harvey, A. (1989). *Forecasting, Structural Time Series Models, and Kalman Filter*, Cambridge: Cambridge University Press.
- Holland, J.H. (1975) *Adoption in natural and artificial systems*, Ann Arbor: The University of Michigan Press.
- Huang, X.D., Ariki, Y., and Jack, M.A. (1990). *Hidden Markov models for speech recognition*, Edinburgh University Press, Edinburgh.
- Hunt B.R. (1973), The application of constrained least-squares estimation to image restoration by digital computer, *IEEE Transactions on Computers*, C-22, 805-812.
- Kalman, R.E. (1960). A new approach to linear filtering and prediction problems, *Trans. ASME, J. Basic Eng.*, **82D**, 35-45, 1960.
- Kitagawa, G. (1987). Non-Gaussian state space modeling of nonstationary time series (with discussion), *J. Am. Stat. Assoc.*, **79**, 1032-1063.
- Kitagawa, G. (1991). A nonlinear smoothing method for time series analysis, *Statistica Sinica*, **1**, No.2, 371-388.
- Kitagawa, G. (1993). Monte Carlo Filtering and smoothing method for non-Gaussian Non-linear state space models, *Inst. Statist. Math. Res. Memo.*, **462**.
- Kitagawa, G. (1994). The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother, *Ann. Inst. Statist. Math.*, **46**, No. 4, 605-623.
- Lindley, D. V. and Smith, A.F.M. (1972). Bayes estimates of the linear model (with discussion), *J. R. Statist. Soc. B*, **34**, 1-41.
- Mardia, K.V. (1989). *SPECIAL ISSUE: Statistical Methods in Image Analysis*, technical editor K.V. Mardia, *J. Applied Statistics*, **16**, No. 2.
- West, M. and Harrison, P.J. (1989). *Bayesian forecasting and dynamic model*, Springer-Verlag, New York.
- West, M. Harrison, P.J., and Migon, H.S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion), *J. Am. Statist. Assoc.*, **80**, No. 389, 73-97.
- Whitley, D. (1994). A genetic algorithm tutorial, *Statistics and Computing*, **4**, 63-85.
- Whittaker, E. T. (1923). On a new method of graduation, *Proceeding of the Edinburgh Mathematical Society*, **41**, 63-75.

Table 1. Analogy between the MCF and GA

Monte Carlo Filter (MCF)	Genetic Algorithm (GA)
$z_{P,i}^{(n)}$ and $z_{F,i}^{(n)}$: particle	string
(n) : time	generation
$g(\cdot, v_n = 0)$	genetic drift
system noise	crossover mutation
filtering	selection
$r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)})/m$	evaluation function
$r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)})/\sum_{i=1}^m r(w_i^{(n)})w'(y_n, z_{P,i}^{(n)})$	fitness
$p(y_n Y_{n-1}, \lambda)$	population fitness
$p(Y_N \lambda)$	family fitness
y_n : observation	environment
Y_N : data	history of environment

Table 2. Comparison of selection schemes: $N_p = 5000$ trials

Sampling Scheme	$\langle \Delta l_i \rangle$	s.d. of Δl_i	c.c.	slope
Deterministic	-0.344	0.392	0.737	0.265
Roulette wheel	-0.127	0.496	0.723	0.234
stochastic universal	-0.0952	0.455	0.776	0.249

Figure Captions

Fig. 1. Artificially generated data y_n . The smooth curve is obtained by applying the fixed interval smoother with the best value of hyperparameter, τ^{2*} , that is determined by maximizing the log-likelihood.

Fig. 2. The log-likelihood of y_n of the best model, given Y_{n-1} , $\log p(y_n|Y_{n-1})$. The arrow indicates a jump point of the step function where $\log p(y_n|Y_{n-1})$ takes a minimum value.

Fig. 3. (a) Distribution of $(\Delta l_i, \Delta l_i^{(51)})$, where Δl_i is defined by a discrepancy of l_i^{MCF} from l^{true} . l^{true} and l_i^{MCF} are values of the log-likelihood of Y_N , $\log p(Y_N|\lambda)$, obtained by the Kalman filtering and the MCF, respectively. i indicates the the i th trial. $l^{\text{true}(51)}$ and $l_i^{\text{MCF}(51)}$ denotes the values of $\log p(y_n|Y_{n-1}, \lambda)$ based on the Kalman filtering and the MCF, respectively. The deterministic sampling scheme is used for the resampling appearing in the filtering process of the MCF.

(b) Distribution of $(\Delta l_i, \Delta l_i^{(51)})$ for the resampling based on the roulette wheel selection.

(c) Distribution of $(\Delta l_i, \Delta l_i^{(51)})$ for the resampling based on the stochastic universal sampling.

Figure 1

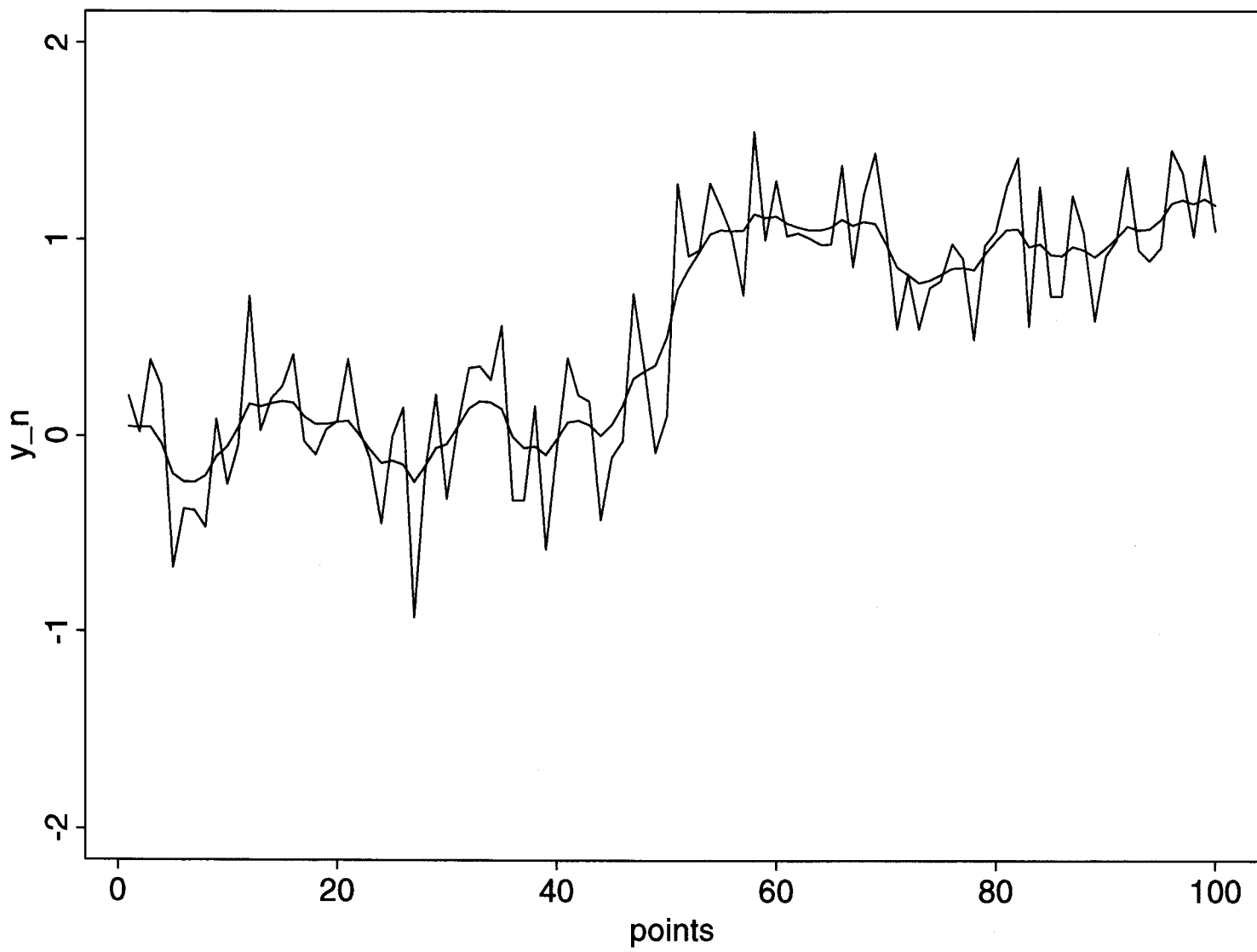


Figure 2

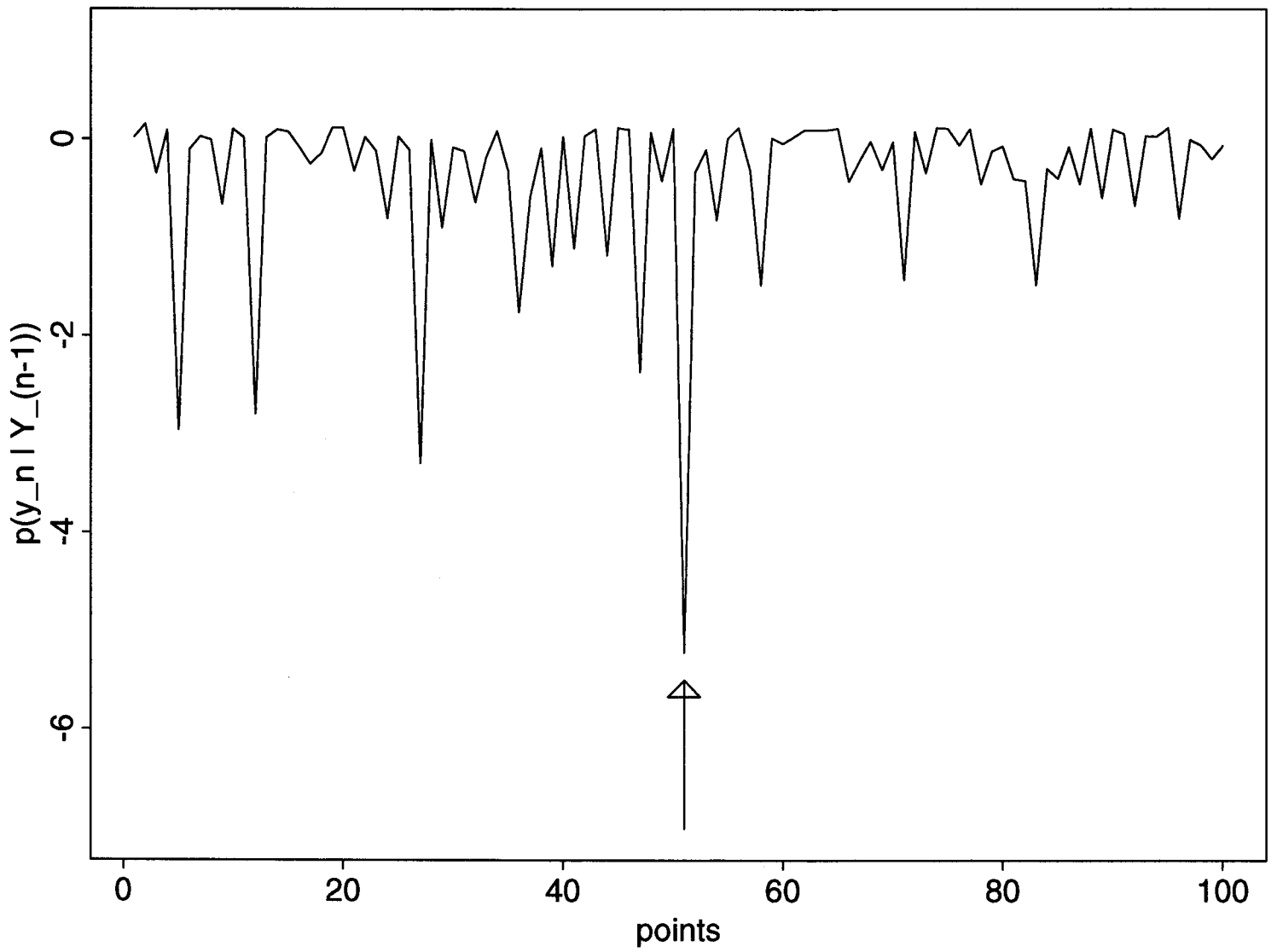


Figure 3(a) 5000 trials

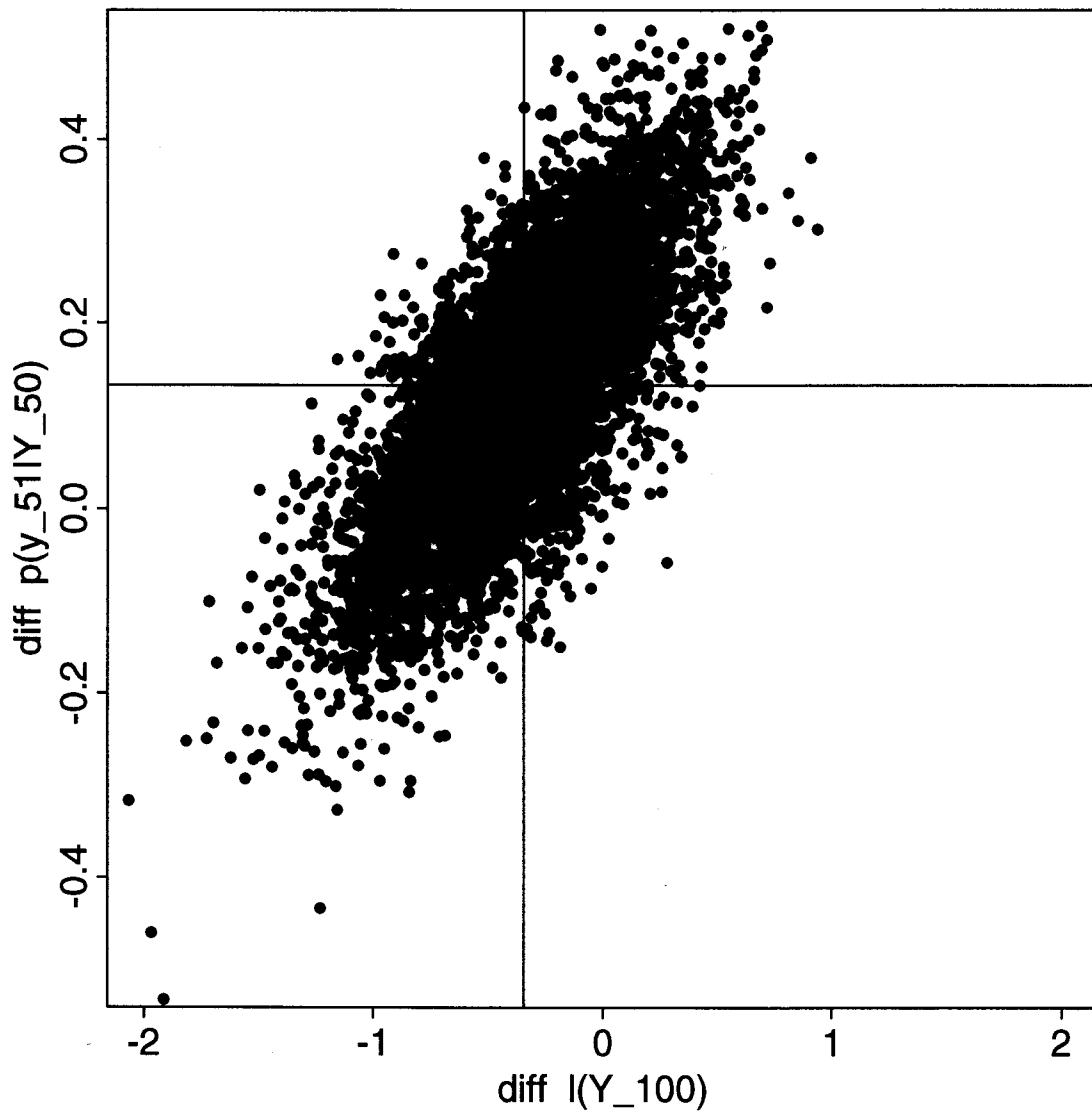


Figure 3(b) 5000 trials

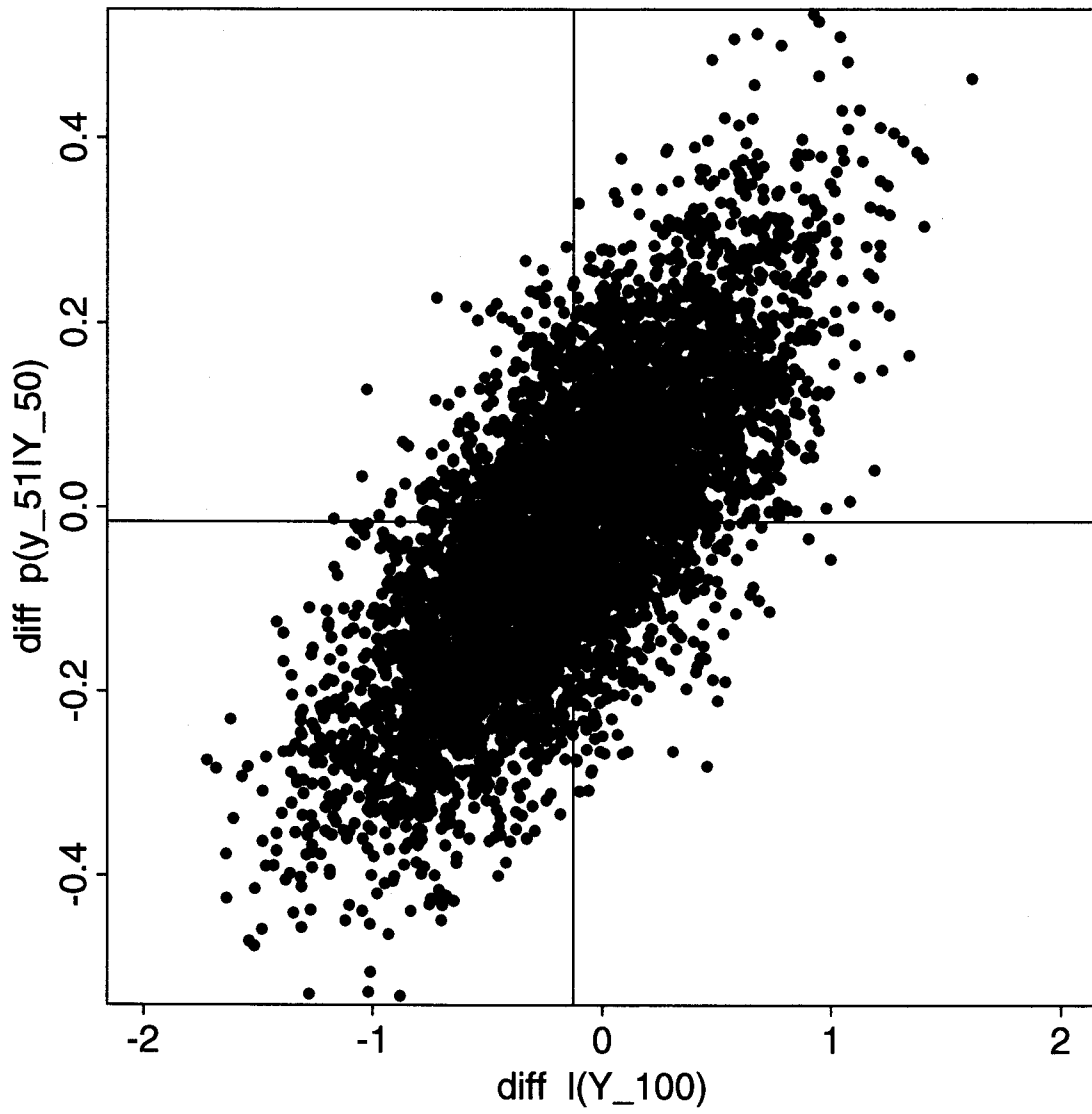
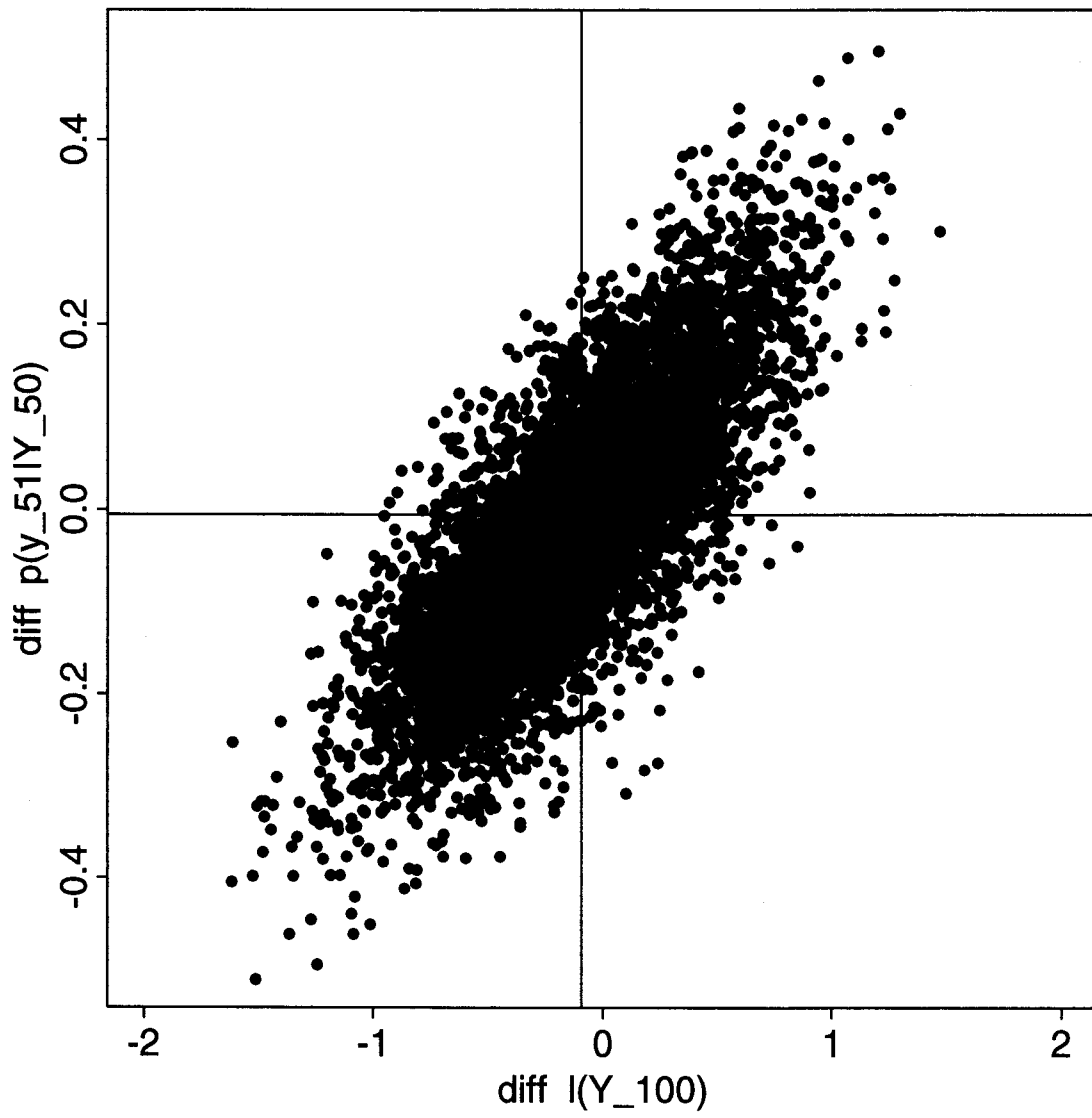


Figure 3(c) 5000 trials



REGISTRATION
of
Research Memorandum

Research Memorandum NO. 556 Received on April 18, 1995 by
(month) (day) (year)

Section of Research Information, the Institute of Statistical Mathematics

(Do not write above this line)

(Fill in the blanks below and return to Section of Research Information)

Title
On the resampling scheme in the filtering procedure of the Kitagawa Monte Carlo Filter
Author(s) and Affiliation(s)
T. Higuchi The Institute of Statistical Mathematics, Tokyo 106, Japan
Key Words
Key words and phrases: Bayesian approach, Generalized state space model (GSSM), Non-Gaussian time series model, Genetic Algorithm (GA), Resampling
Abstract
The major objective of this paper is to address the close relationship between the Genetic algorithm, which is usually used to get an optimal solution for a nonlinear multiparameter optimization problem, and Kitagawa's Monte Carlo filter, (Kitagawa 1995), that was developed to manage a large dimensional state vector in a generalized state space model. The Monte Carlo filter is interpreted from the perspective of the Genetic algorithm. On the other hand, an essential part of the Genetic algorithm can be viewed as a special case of the Monte Carlo filter, and then a Bayesian interpretation of the Genetic algorithm leads to an examination of the performance of several ideas to improve the optimization. Based on an analogy between the Genetic algorithm and the Monte Carlo filter, we investigate a resampling scheme involved in the filtering procedure of the Monte Carlo filter. Several schemes for resampling are applied to a simple problem and their performance is evaluated in terms of an estimation of the likelihood.