Looking at Markov Samplers through Cusum Path Plots: a simple diagnostic idea

Bin Yu*

Department of Statistics, University of California at Berkeley

Per Mykland†

Department of Statistics, University of Chicago

Abstract

In this paper, we propose to monitor a Markov chain sampler using the cusum path plot of a chosen 1-dimensional summary statistic. We argue that the cusum path plot can bring out, more effectively than the sequential plot, those aspects of a Markov sampler which tell the user how quickly or slowly the sampler is moving around in its sample space, in the direction of the summary statistic. The proposal is then illustrated in four examples which represent situations where the cusum path plot works well and not well. Moreover, a rigorous analysis is given for one of the examples. We conclude that the cusum path plot is an effective tool for convergence diagnostics of a Markov sampler and for comparing different Markov samplers.

KEY WORDS: Convergence diagnostic; Cusum path plot, Markov sampler; Mixing; Sequential plot; Summary statistic.

*Research supported in part by ARO Grant DAAL03-91-G-007.

†Research supported in part by NSF Grant DMS-9305601.

1 Introduction

As Markov chain Monte Carlo methods become more popular among Bayesians and non-Bayesians alike, the more acute becomes the problem of reliable convergence diagnostics of a Markov sampler. Previous works addressing this problem include Cui et al (1992), Gelman and Rubin (1992b), Liu et al (1992), Roberts (1992, 1993), and Yu (1994). Both Gelman and Rubin (1992b) and Yu (1994) rely on additional information to that from a single run, though in different ways.

In this paper, we propose to use the cusum (path) plot from quality control (cf. Kotz and Johnson (1988), p. 233-236) to monitor the performance of the Markov sampler. The cusum path plot is based on a single run and a chosen 1-dim summary statistic. We argue that this plot can bring out, more effectively than the sequential plot, those aspects of a Markov sampler which tell the user how quickly or slowly the sampler is moving around in the sample space, in the direction of the summary statistic. From the examples we implemented, we can see that, for some Markov samplers, we can get more information than previously believed about the convergence or mixing behavior using only a single run. We conclude that the cusum plot is an effective tool for convergence diagnostics of a Markov sampler.

This paper is organized as follows. Section 2 contains the cusum proposal and its basic supporting arguments. Section 3 contains 4 examples where the proposal is implemented. The first three examples are chosen to represent situations where the cusum path plot works well: the bimodal case, the "cigar" case (Gilks and Roberts, 1993), and the "slow-mixing" data from the literature (Gelman and Rubin, 1992a). These three examples have the same feature: the sampler moves around in the sample space in a relatively homogeneous way, in the direction of the summary statistic. In other words, we believe that the driving force behind the bimodal and "cigar" phenomena are the same: the sampler is moving slowly, both within the same mode and among different modes if there are more. The proposed cusum plot brings out nicely this "slowness" or "quickness" of a sampler and therefore is effective even when the sampler gets stuck at one mode for a particular run. This issue is further investigated rigorously in Section 4 in the context of the cigar example. In a way, we feel that we have found a good pair of glasses, the cusum path plot, to look at Markov sampler with, and different samplers can be compared under this pair of glasses. However, the cusum path plot is no panacea. The last example is the "uniform" variant of Witch's Hat (cf. Cui et al, 1992). It gives a situation where the sampler has a split mixing behavior and the cusum path plot might not work well: the sample space is divided into two regions between which the communicativity almost does not exist and the sampler mixes more quickly in one region than in the other. If we start the sampler in the better mixing region, the cusum path plot won't be able to detect the other region before it moves there. In this particular example, the difficulty of detection is increased tremendously by the fact that the worse-mixing region is very tiny. Obviously, the sequential plot (single run or multiple runs), the Gibbs stopper in Cui et al (1992), and the L^1 error proposal in Yu (1994) will have a hard time as well: in first case if the sampler never visits the slow-mixing region, in the second case if the candidate proposal in the Metropolis-Hastings algorithm never proposes a point in that region, and in the third case if the grid used in the evaluation of the estimated L^1 error is not fine enough.

2 Cusum path plot – a good pair of glasses

The idea of using the cusum path plot to monitor Markov samplers dawned on us when we realized that in the test case of bimodal target densities, a bad Markov sampler tends to stay at one mode for a while and makes a jump and then stays at the other mode for another while. Ignoring the dependence between the iterations and the possible nonstationarity of the chain, this is very much like the "change point" problem which the cusum plot is known to be good for. First we introduce the cusum plot in a formal way and later we provide its supporting arguments based on change point and partial sum literatures when the observations are not necessarily iid.

Since cusum plots can be constructed only for 1-dim quantities, we have to decide beforehand the 1-dim summary statistic(s) of the Markov sampler we want to monitor. We may monitor one summary statistic at a time or several simultaneously. The sampler is monitored only in the direction(s) of the chosen 1-dim statistic(s), in the hope that the convergence behavior of the chosen statistic(s) reflect well that of the sampler.

Assume $X_0, X_1, ..., X_n$ is a single run of a Markov sampler, which may be discrete or continuous. Let $T(X)$ be the 1-dim summary statistic we decide to monitor. Let n_0 be the "burn-in" time and we should construct our cusum statistics based on $T(X_{n_0+1}), ..., T(X_n)$. The reason we don't construct the cusum statistic from the beginning is to avoid the initial bias of the sampler since this bias can be too overwhelming in the cusum plot (a huge linear tread upwards or downwards), and this linear trend could mask other finer features of the statistic. What we hope to get out of the cusum plot is the more detailed information we can't see in the simple sequential plot of $T(X)$ which MCMC users have been plotting all along. *Hence it is an implicit presumption that a sequential plot of $T(X)$ should be done before the cusum plot and the burn-in time should be chosen based on this sequential plot.*

For given starting and ending values n_0 and n, denote the expected sample mean of $T(X)$ and its estimator as

$$\mu := (n - n_0)^{-1} \sum_{n_0+1}^{n} ET(X_j), \;\; \hat{\mu} := (n - n_0)^{-1} \sum_{n_0+1}^{n} T(X_j),$$

the cusum or partial sum

$$S_t :== \sum_{j=n_0+1}^{t} [T(X_j) - ET(X_j)] = \sum_{j=n_0+1}^{t} [T(X_j) - \mu] \;\text{ for }\; t = n_0 + 1, ..., n,$$

and the observed cusum or partial sum

$$\hat{S}_t := \sum_{j=n_0+1}^{t} [T(X_j) - \hat{\mu}] = S_t - (\hat{\mu} - \mu)t \;\text{ for }\; t = n_0 + 1, ..., n.$$

Cusum path plot: Plot $\{\hat{S}_t\}$ against t for $t = n_0 + 1, ..., n$ and connect the successive points with line segments. Since $\sum_t \hat{S}_t = 0$, the cusum path plot ends at 0.

Next we argue that the slowness or the quickness of the mixing rate of $T(X)$ is reflected in the smoothness of the cusum plot path, i.e., the more "hairy" the cusum path is, the faster the mixing rate of $T(X)$; the smoother the cusum path, the slower the mixing rate of $T(X)$.

First, we provide some heuristic arguments. In quality control, cusum plot is used to detect small shifts in the process. When there is a shift, we see an approximate line segment in the cusum plot. Hence if the target density has multiple modes or sticky locations within the same mode where the sampler gets stuck at, (for example, because of the small acceptance probability when using the Metropolis-Hastings algorithm, cf. Hastings, 1970), we may view the process $T(X)$ as having many shifts whose sizes depend on the differences in $T(X)$ which the locations of the sticky points of the sampler induce, hence we expect to observe line segments along the cusum path and hence the path appears smooth.

On the other hand, we can argue more rigorously about the same point based on the result in Lin (1992, Theorem, p. 323), which characterizes the behavior of the increments of partial sums of a ϕ-mixing sequence. The sequence does not have to be stationary. Of course, our sampler might not be ϕ-mixing. Nevertheless, Lin's result tells us quantitatively how the cusum's behavior depends on the variances of sums over blocks of the sequence. The stickiness of the sampler is reflected in the variances of sums in the sense that the stickier the sampler, the faster the variance increases as the block size increases. To be more precise, for a given nondecreasing sequence a_n, assume $ET(X_i) = 0$, and denote

$$\sigma_{nN}^2 = E(T(X_{n+1} + ... + T(X_{n+a_N}))^2,$$

$$f_{nN} = \sigma_{nN}\{2[\log(N/\sigma_{nN}^2) + \log\log N]\}^{1/2},\ S(n,k) = S_{n+k} - S_n.$$

Lin imposed the following conditions to obtain his result.

$$\text{(i)} \liminf_{n\to\infty} \inf_{m\geq 0} E(T(X_{m+1} + ... + T(X_{m+n})^2/n > 0;$$

(ii) there exists t_0, $M > 0$, such that $Ee^{tT(X_k)} \leq M$ for every k and any $|t| \leq t_0$;

(iii) there exists $l > 0$ such that the ϕ mixing coefficient (cf. Lin (1992) for definition)

$$\phi_n = O(n^{-l});$$

(iv) there exists $a > 0$ such that $a\log^d n \leq a_n \leq n$, where $d > 3l + 1/(l-1)$.

Under the above assumptions, Lin asserts that

$$\limsup_{N\to\infty} \max{}_{1\leq n\leq N} f_{nN}^{-1}|S(n,a_N)| = 1 \ \ a.s.$$

$$\limsup_{N\to\infty} \max{}_{1\leq n\leq N} \max{}_{1\leq k\leq a_N} f_{nN}^{-1}|S(n,k)| = 1 \ \ a.s.$$

Hence when N large and $\delta \in (0,1)$ small, we can choose $a_n = \delta n$, then Lin's result implies that the cusum path behaves like f_{nN}, which is an increasing function of σ_{nN}^2 and N. The stickier the sampler, the faster the variance σ_{nN}^2 increases as N and thus a_N increases. Hence the steeper the partial sum path if we connect the partial sum points with line segments, that is, the smoother the cusum path S_t. Also, the stickier the sampler, the larger the variance σ_{nN}^2; thus its cusum path plot S_t takes a relatively bigger excursion before coming back to zero at n. This is consistent with the stickier sampler having a smoother path since to make a big excursion the increments should point in the same direction; since we are connecting the cusum points by lines, this will form longer line segments and hence make for a smoother cusum path. So far, we have pretended that $ET(X_i) = 0$, but in our observed cusum path plot $\hat{S}_t$, the expected value of the sample mean is estimated by $\hat{\mu}$. When $T(X)$ does not mix well, it is likely that $\hat{\mu}$ is biased compared with the expected sample mean. This (observed) bias will add to the theoretical cusum plot (whose smoothness we have argued to be related to the stickiness of the sequence $T(X)$) a linear term with minus the bias as the slope. Fortunately, this will only enhance the smoothness of the observed cusum path plot $\hat{S}_t$. Therefore the smoothness of the observed cusum path plot reflects both the variance and bias of the sampler in the direction of $T(X)$, and remember we want both the variance and bias to be small, i.e. the observed cusum path plot to be hairy. Since the observed cusum path plot starts and ends at 0, if the variance of the sample sum (or mean) is large, we will see large excursions on the observed cusum path plot.

When the (observed) cusum path plot is very small, the solution of the graphic device will interfere with our visual perception of the "smoothness" of the cusum path. We will have a hard time distinguishing a smooth path from a hairy one since they can both look like bold lines on a small plot. It is therefore a good idea to make the cusum path plot as large as we can, say use a whole page, and keep in mind that *small cusum plots can be very non-informative and should be avoided at all cost.*

Another worry is that our perception of "smoothness" or "hairiness" of a path is very subjective. For this reason, one can do a much better job if there is another path one can compare with. Keeping in mind the invariance principle for partial sum processes of weakly dependent sequences (cf. Phillip and Stout, 1975), we suggest to add a "benchmark" cusum path in the same cusum path plot of $T(X)$. The "benchmark" plot is the cusum path plot of an iid sequence of normal random variables with their mean and variance matched with the estimated mean and variance of $\{T(X_j) : j = n_0 + 1, ..., n\}$. That is, for $t = n_0 + 1, ..., n$, let

$$\hat{S}_t^b := \sum_{j=n_0+1}^{t} [Y_j - \hat{\mu}_Y], \;\; \hat{\mu}_Y := (n - n_0)^{-1} \sum_{j=n_0+1}^{n} Y_j,$$

where $Y_{n_0+1}, ..., Y_n$ is an iid sequence of $N(\hat{\mu}_T, s_T^2)$ random variables with $\hat{\mu}_T$ as above and s_T^2 being the sample variance of $\{T(X_j) : j = n_0 + 1, ..., n\}$.

Thus when we only have one sampler, we should

plot $\hat{S}_t$ *and* $\hat{S}_t^b$ *against* t *for* $t = n_0 + 1, ..., n$ *and connect the points by line interpolation for each path.*

By the invariance principle for the partial sums of weakly dependent process, the $\hat{S}^b$ path approximates, to the second order, the "ideal" cusum path of an iid sequence from the same target distribution. If the $\hat{S}$ path is comparable with the $\hat{S}^b$ path in terms of smoothness of the path and size of the excursion, then we conclude that the sampler is mixing well (in the direction specified by $T(X)$, to be precise). Otherwise, we conclude that the sampler is not mixing well, in the direction specified by $T(X)$.

When our intention is to compare two samplers for the same target distribution, we may omit the "benchmark" cusum path plot in which case we just compare the cusum paths from the two samplers in the same plot.

3 Examples

In this section, we demonstrate the cusum path plot in four examples, which represent four situations where slow mixing is known to occur, but in different manners.

Example 1: Bimodal. Mixtures of two normals

Let us take the test case where the target density is a 1-dim bimodal mixture density with two equal modes at 0 and 3 and variance 1 at each mode, cf. Yu (1994).

Three samplers are run up to n=2000 iterations for this target density and the burn-in time is taken to be $n_0 = 1000$. The three samplers are the independent sampler, the MH sampler 1 and the MH sampler 2. The latter two samplers are obtained using Metropolis-Hasting algorithm with an $N(x, \sigma)$ jumping kernel and $\sigma = 4.146$ for sampler 1 and $\sigma = 3$ for sampler 2. N(0,1) is used as the initial distribution. By the analysis in Gelman (1994), we know that sampler 1 mixes faster than sampler 2 since it is "optimal" in the sense specified by Gelman (1994). The sequential plots in Fig. 1 indicate that both sampler 1 and sampler 2 are less desirable than the independent sampler, but have little power of distinguishing between sampler 1 and sampler 2. However, the cusum path plots in Figures 2-4 tell us that, although both sampler 1 and sampler 2 are pretty good (Fig. 3-4), sampler 1 indeed seems better than sampler 2 (Fig. 2).

Example 2: "Cigar". Gibbs sampler for bivariate normal (Gilks and Roberts (1993))

In this example, the target distribution is bivariate normal with marginal means 0 and marginal variances 1, and correlation coefficient ρ. When ρ is close to 1, it is known that the bivariate density has a "cigar" shape along the diagonal line of the first quadrant and the two-component Gibbs sampler moves slowly up and down the "cigar".

We take the first component of the bivariate variable as our 1-dim summary, i.e., $T(X) = X_1$ and $X = (X_1, X_2)$. The independent sampler in the plots is simply the iid $N(0, 1)$ sampler. N(0,1) is used as the initial distribution so our Markov sampler produces a stationary sequence. Again, three samplers are taken: the independent sampler, the Gibbs sampler 1 when $\rho = 0.75$ and the Gibbs sampler 2 when $\rho = 0.80$. Similar observations as those made in Example 1 can be made here based on Figures 5-8.

Example 3: A slowly moving Gibbs sampler for an Ising model (Gelman and Rubin (1992a))

Professor Gelman has kindly provided us the two runs of the 1-dim summary statistic from a Gibbs sampler for an Ising model, that appeared in Gelman and Rubin (1992a). By their analysis, we known that the 1-dim summary statistic seem to have converged on a single run, but multiple runs reveal that the convergence is false.

We present the cusum path plots for both runs for $n_0 = 1000$ and $n = 2000$, with benchmark paths added. It is obvious from each path plot (Fig. 10 or 11) that this summary statistic is sticky, without using the information from the other run. This suggests that the slow-converging property of this summary statistic is due to that fact that the sampler moves around slowly, even within the same mode. Our cusum path plot magnifies this "slowness" to a degree where it can be seen easily, so that a single run gives enough information about the stickiness of the sampler. We are nicely surprised about this example since many MCMC researchers, including the authors, had felt that a single run might not be enough for the convergence diagnostic purpose, mainly because of examples like this one. As will be illustrated by the next example, there are still cases where a single run might not provide enough information. We can learn more, however, about the sampler from a single run by using the cusum path plot than by using just the sequential plot. In other words, if the sampler is slow-mixing homogeneously in the sample space, our proposed cusum path plot has the ability to detect the stickiness within a mode, which the simple sequential plot of $T(X)$ does not have.

Example 4: A "uniform" variant of Witch's Hat (cf. Cui et al (1992))

This example gives a situation where the cusum path plot might fail and so might other diagnostic means, for example, those in Gelman and Rubin (1992b), Cui et al (1992) and Yu (1994), because it represents, in a way, the ultimate "nightmare" of a MCMC user.

Here the target density is an equal weight mixture of two uniforms: $U(0,1)$ and $U(0.5-\delta, 0.5+\delta)$. When δ is small, this density looks like a Witch's Hat with a flat top, instead of the traditional peaky one (cf. Cui et al, 1992). In the simulation, we use $N(x,1)$ as the proposal kernel in the Metropolis-Hastings algorithm and take $\delta = 0.005$.

The unit interval can be divided into two regions: $I_1 = (0, 0.5-\delta] \cup [0.5+\delta, 1)$ and $I_2 = (0.5-\delta, 0.5+\delta)$. On I_1 the rejection rate in the Metropolis-Hastings algorithm is 0 so the sampler moves around following the proposal kernel – mixing relatively fast, until it moves to a point in I_2. Note that this happens with probability less than 0.004 ($= 2\delta/\sqrt{2\pi}$) which is an upper bound on the probability that the proposal kernel suggests a point in I_2. On I_2, the acceptance rate is almost always 1/101 unless the proposal from the Normal kernel gives a candidate in I_2 and this happens with probability less than 0.004. Hence the two regions communicate with each other with probabilities less than 0.004, and I_2 is a sticky or slow-mixing region, while I_1 is a fast-mixing region. Most likely one would start the sampler in I_1 and everything would seem mixing fast until the sampler enters I_2 and it will be stuck there for a long time. We simulate a sample of size $n = 2000$ with $X_0 = 0.1$. If we take the first 500 observations, both the sequential and cusum path plots (Fig. 12) suggest the mixing speed of the sampler is not too bad, but when we continue the simulation up to 2000 iterations, both the sequential and cusum plots (Fig. 13) tell us there is something very sticky: two flat line segments in the sequential plot and two sloppy line segments in the cusum plot. In fact, the transitions between the two regions are just obvious in the sequential plot as in the cusum path plot.

4 The Cigar Chain: an Asymptotic Study

In this section, only for the mathematically tractable cigar case, we establish rigorously the connection between the mixing behaviors of the 1-dim summary statistic and the smoothness properties of the cusum path plots, and we argue that the cusum path plot can be used to infer unseen nonconvergence of the summary statistic, at least in this case. There are no doubt several ways of investigating whether this is actually the case; we here propose to do it with triangular array asymptotics: as the sample size increases, the 1-dim summary statistic $T(X) = X_1$ process is increasingly nonergodic or slow-mixing.

In the cigar case, it is straightforward to verify that $T(X) = X_1$ follows an AR(1) process of the form:

$$X_{1,n+1} = \rho^2 X_{1,n} + (1-\rho^4)^{1/2}\epsilon_{n+1},$$

where the ϵ's are independent standard normal. Obviously, if $\rho < 1$, this process is asymptotically ergodic and fast-mixing, and the distribution of the cusum path plot will approach a Brownian bridge; hence "hairy" cusum path. The almost nonergodic (therefore very slow or non-mixing) case is obtained by considering the cusum plot based on n observations when the correlation is ρ_n, and then let $\rho_n \to 1$ as $n \to \infty$. A similar approach is common when doing inference on AR processes which are close to the boundary of the domain of stationarity, see, e.g., Chan & Wei (1987).

The cutoff between ergodic and nonergodic behavior is as follows. If $1 - \rho_n >> O(n^{-1})$, it is easy to see that $\bar{X}_{1,n}$ (the mean of n observations of X_1 under ρ_n) converges to zero, which is the desired mean. On the other hand, if $\rho_n = 1 - \gamma/n + o(n^{-1})$, then

$$\bar{X}_{1,n} \to N(\mu, \sigma^2)$$

in law, where

$$\mu = X_{1,1}(1 - \exp(-2\gamma))/2\gamma$$

($X_{1,1}$ if $\gamma = 0$) and

$$\sigma^2 = (-e^{-4\gamma} + 4e^{-2\gamma} + 4\gamma - 3)/4\gamma^2$$

(zero if $\gamma = 0$). This is conditional on the first value of X_1, which could either be a fixed number or (if we are dealing with the stationary chain) standard normal.

The test of whether the cusum plot provides the correct diagnostic as suggested in Section 2 is now one of whether it is smoother in the second case than in the first.

The answer is that this is indeed the case. To discuss the asymptotic behavior, assume that the cusum path is standardized to be a function of $t, 0 \leq t \leq 1$, and that it is multiplied by a nonrandom factor to have a nontrivial limiting process. Having done this, it is readily apparent that if $1 - \rho_n >> O(n^{-1})$, the limiting process of the cusum path is a Brownian bridge; hence "hairy". This is the same as in the asymptotically ergodic case ($\rho < 1$). In the other case, if $\rho_n = 1 - \gamma/n + o(n^{-1})$, the limiting process of the cusum path is a smooth random process with nonzero total variation, specifically

$$\int_0^t Y_s ds - t \int_0^1 Y_s ds,$$

where Y_t is the Ornstein-Uhlenbeck process given by the stochastic differential equation

$$dY_t = -2\gamma Y_t dt + dW_t$$

(this is most easily seen using, say, Chan & Tong (1987)).

5 Concluding Remarks

In this paper, we proposed to use the cusum path plot of a chosen 1-dim summary statistic to monitor the mixing or convergence behavior of a Markov sampler. We argued in general and established rigorously in the cigar example, that the smoothness of the cusum path plot corresponds to the slowness of the mixing behavior of the summary statistic. Based on the examples in Section 3 and the theoretical analysis in Section 4, we believe that the cusum path plot can be more effective than the sequential plot so both should be used together, especially when the sequential plot seems to suggest good mixing behavior. Overall, this research endeavor has led us to sense a notion of "local" mixing property of a Markov sampler in relation to its sample space: some regions have fast-mixing properties while others have slow-mixing properties. The cusum path plot might fail when the local mixing property is very inhomogeneous, for example, as in Example 4 in Section 3. We regret that we haven't been able to pin down the precise mathematical meaning of this "local" mixing property, but we do believe that thinking in this term even at a very heuristic level helps us understand different behaviors of Markov samplers.

References

[1] Chan, K. S., and Tong, H. (1987). A note on embedding a discrete parameter ARMA model in a continuous parameter ARMA model. *J. Time Ser. Anal.* **8** 277-281.

[2] Chan, N. H., and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Statist.* **15** 1050-1063.

[3] Cui, L., Tanner, M. A., Sinhua, B., and Hall, W. J. (1992). Comment: Monitoring Convergence of the Gibbs Sampler: Further Experience with the Gibbs Stopper. *Statist. Sci.* **7** 483-486.

[4] Kotz, S. and Johnson, N. (1988). *Encyclopedia of Statistical Sciences*, Vol. 5, Wiley.

[5] Gelman, A. (1993). A Note on Efficient Metropolis Jumping Rules. Manuscript.

[6] Gelman, A. and Rubin, D. B. (1992a). A single series from the Gibbs sampler provides a false sense of security. In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds) 625-632. Oxford Univ. Press.

[7] Gelman, A. and Rubin, D. B. (1992b). Inference from Iterative Simulation Using Multiple Sequences. *Statist. Science* **7** 457-511.

[8] Gilks, W. R. and Roberts, G. O. (1993). Adaptive Markov Chain Monte Carlo. *IMS Bulletin*, Vol. 22, 275.

[9] Hastings, W. K. (1970). Monte-Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57** 97-109.

[10] Lin, Z. Y. (1992). On the Increments of Partial Sums of a ϕ-mixing Sequence. *Theory Probab. Appl.* **36** 316-326.

[11] Liu, C., Liu, J., and Rubin, D. B. (1992). A Variational Control Variable for Assessing the Convergence of the Gibbs Sampler. In *1992 Proc. Statist. Comp. Sec. of ASA* 74-78. *J. Chem. Phys.* **21** 1087-1092.

[12] Philipp, W. and Stout, W. (1975). *Almost sure invariance principles for partial sums of weakly dependent random variables.* Mem. Amer. Math. Soc., 161.

[13] Roberts, G. O. (1993). Methods for Estimating L^2 Convergence of Markov Chain Monte Carlo. Manuscript.

[14] Yu, B. (1994). Estimating the L^1 error of Kernel estimators based on Markov samplers. Technical Report 409, Statistics Department, University of California, Berkeley.

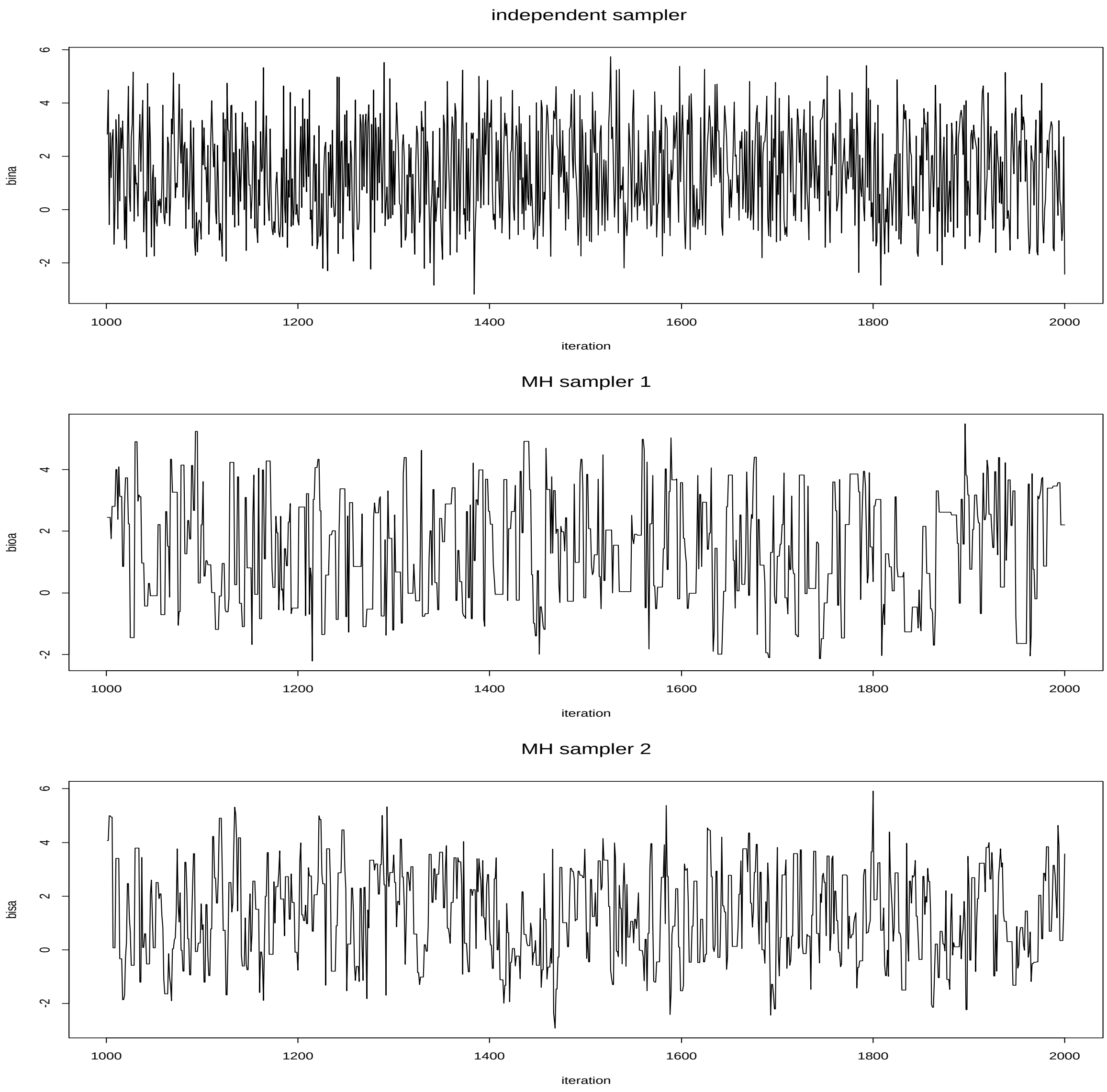

Figure 1: Bimodal example. Sequential plots for three samplers

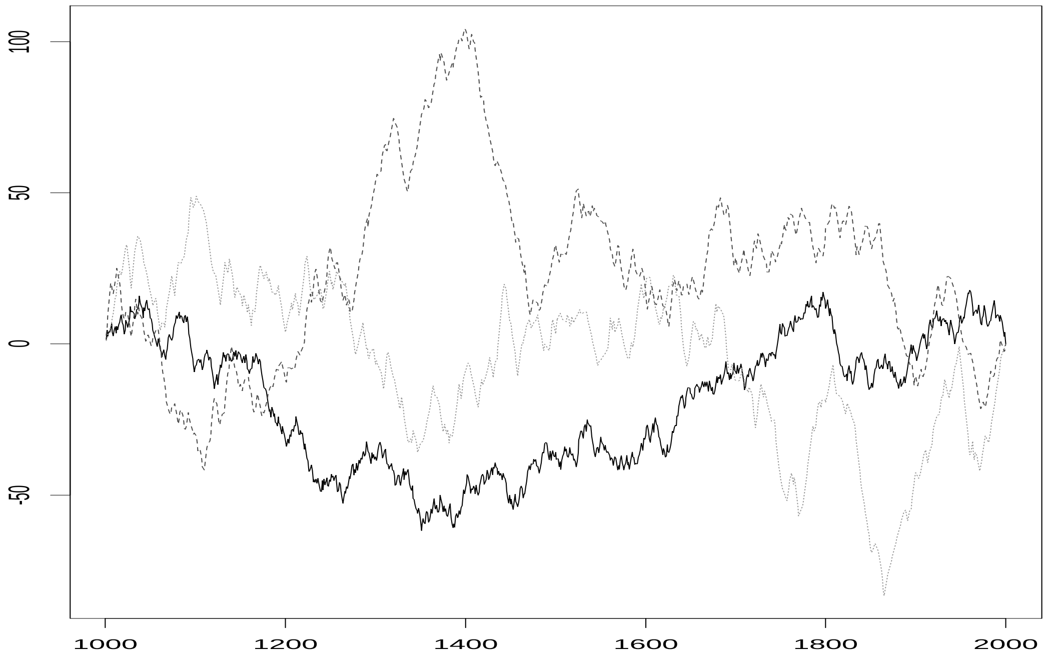

Figure 2: Bimodal example. Solid line –independent sampler; dotted line – sampler 1; dashed line – sampler 2

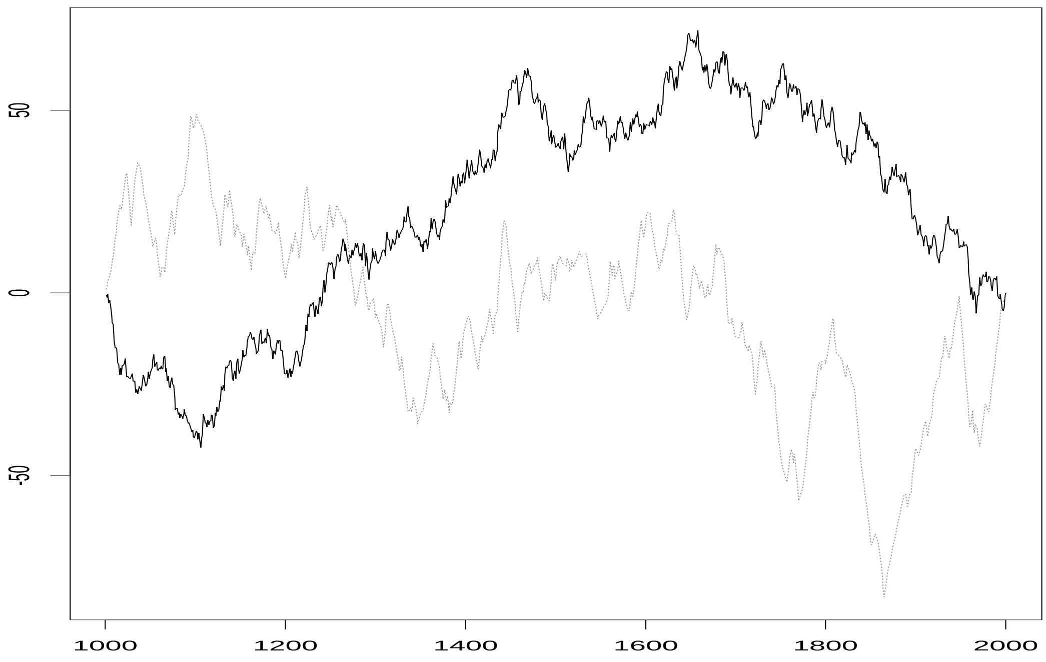

Figure 3: Bimodal example. Solid line – benchmark path; dotted line – sampler 1 path

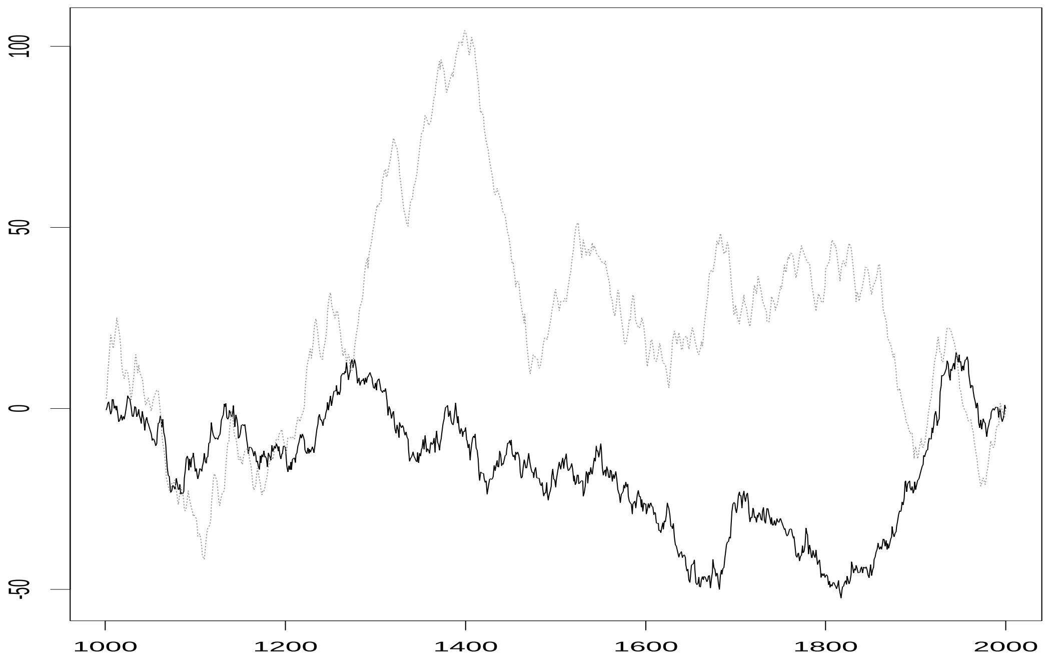

Figure 4: Bimodal example. Solid line – benchmark path; dotted line – sampler 2 path

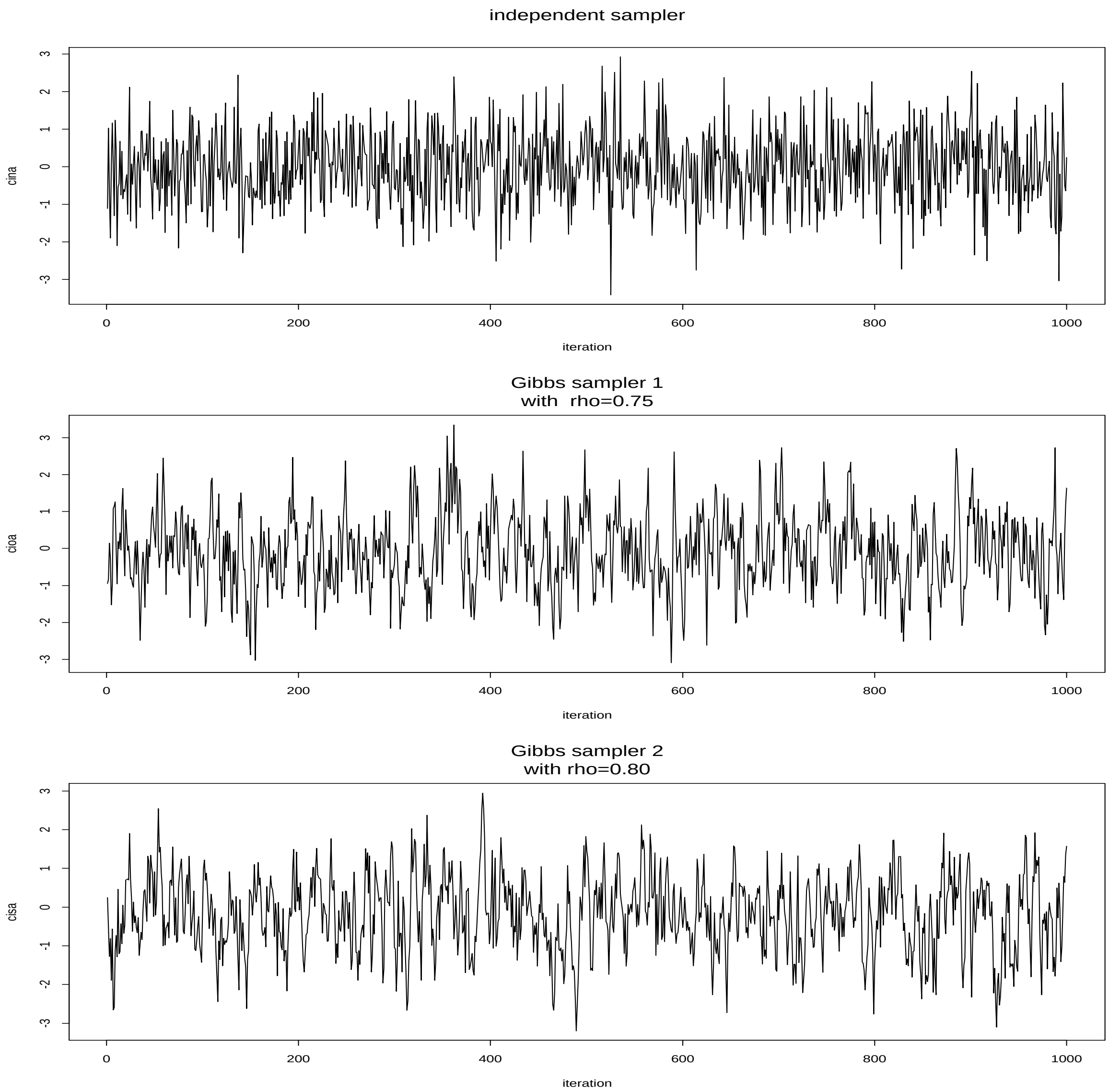

Figure 5: Cigar example. Sequential plots for three samplers

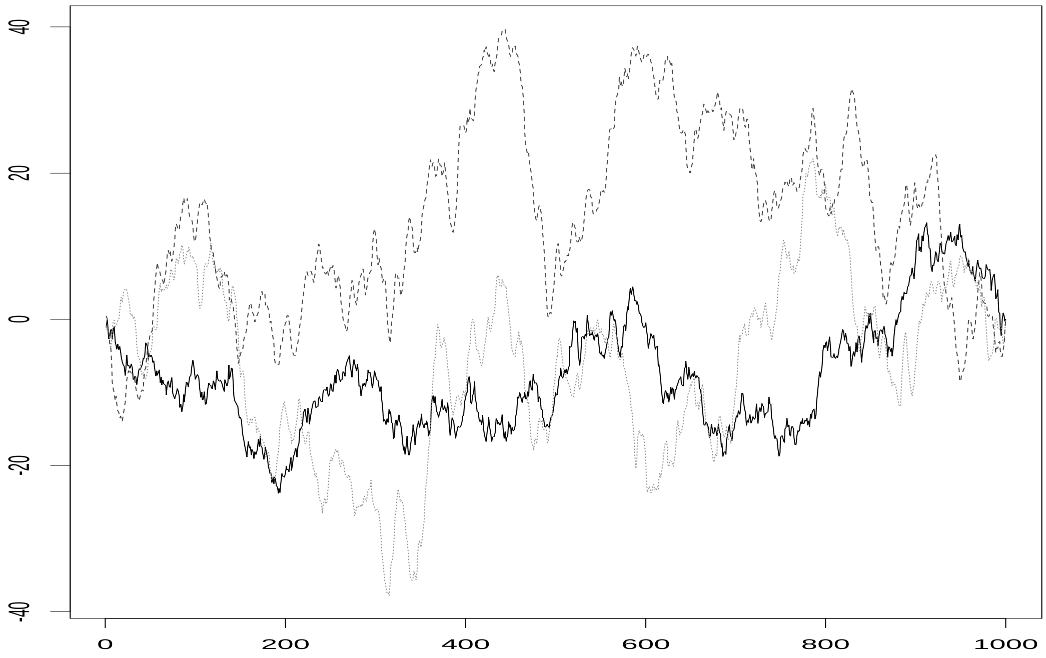

Figure 6: Cigar example. Solid line–independent sampler path; dotted line – Gibbs sampler 1 for rho=0.80; dashed line- Gibbs sampler 2 for rho=0.75

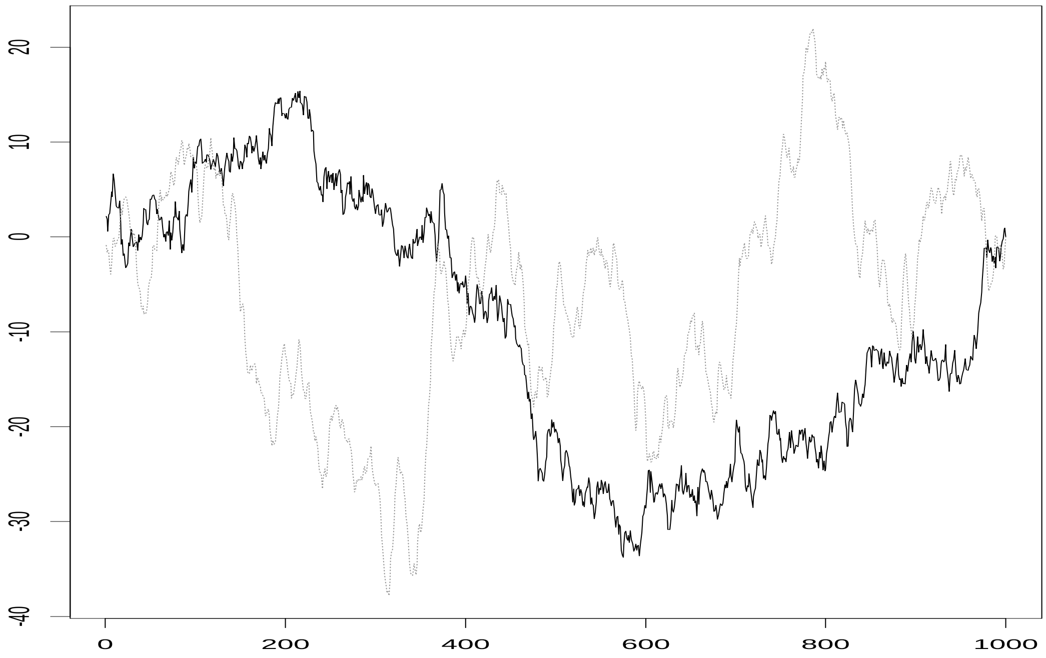

Figure 7: Cigar example with rho=0.75. Solid line–independent sampler path; dotted line – Gibbs sampler 2 path

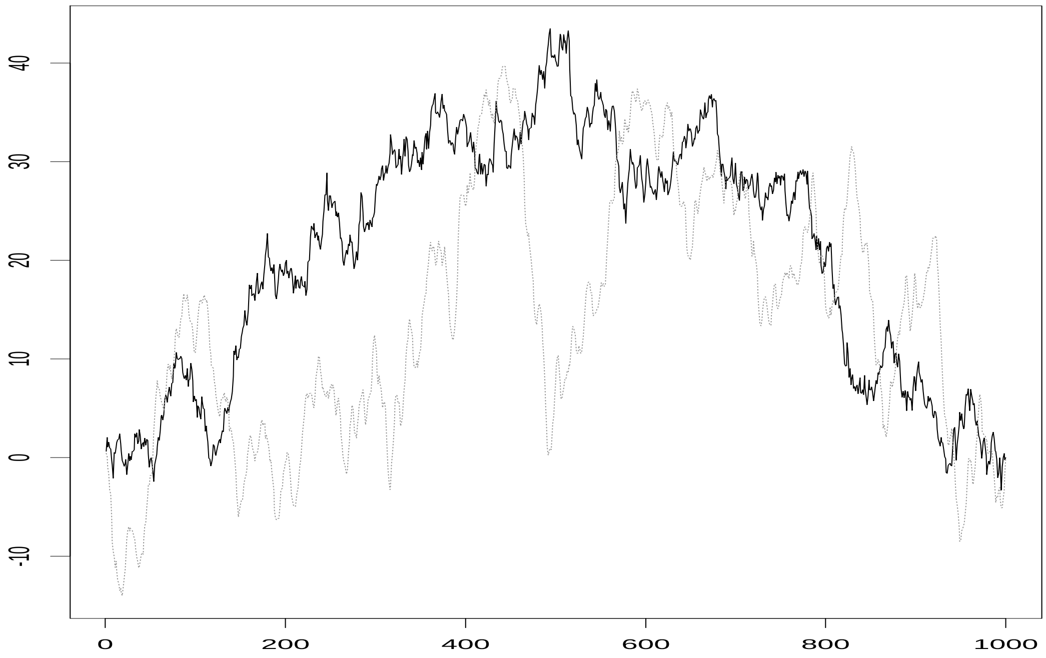

Figure 8: Cigar example with rho=0.80. Solid line–benchmark path; dotted line – Gibbs sampler 1 path

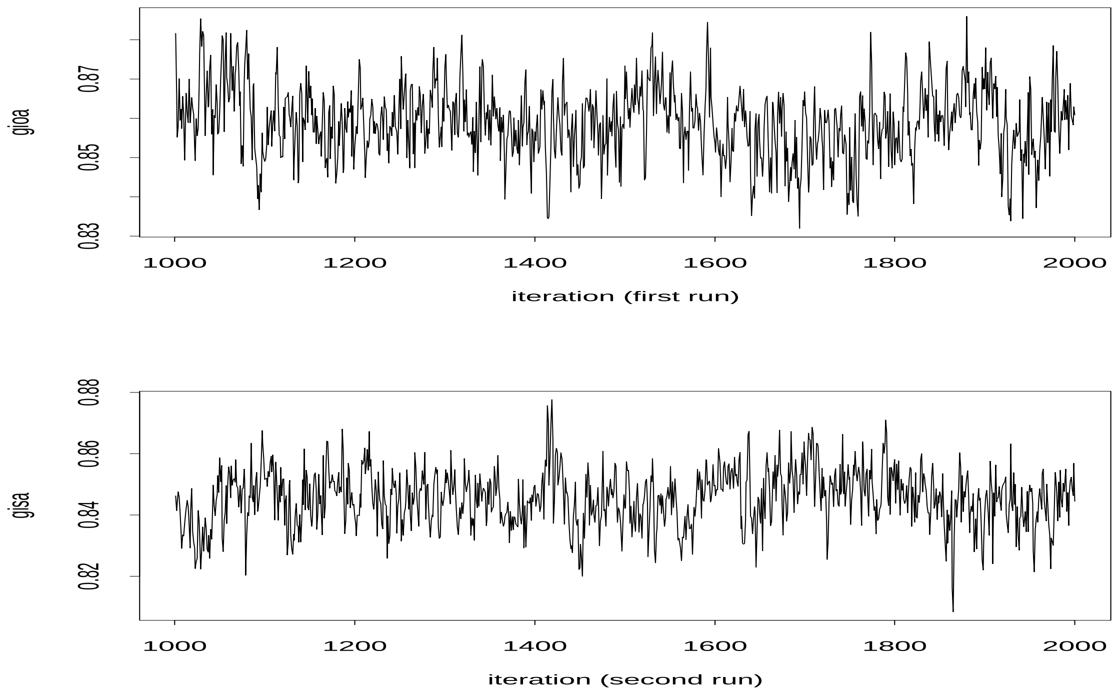

Figure 9: Ising model. Sequential plots for two runs

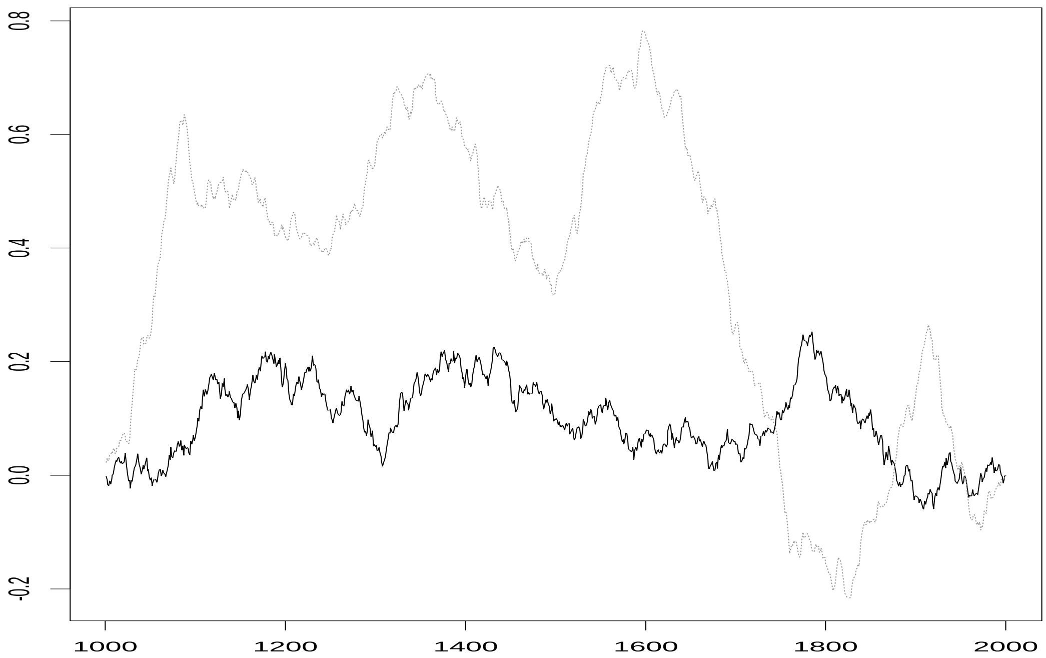

Figure 10: Ising model: first run. Solid line–benchmark path; dotted line – Gibbs sampler path

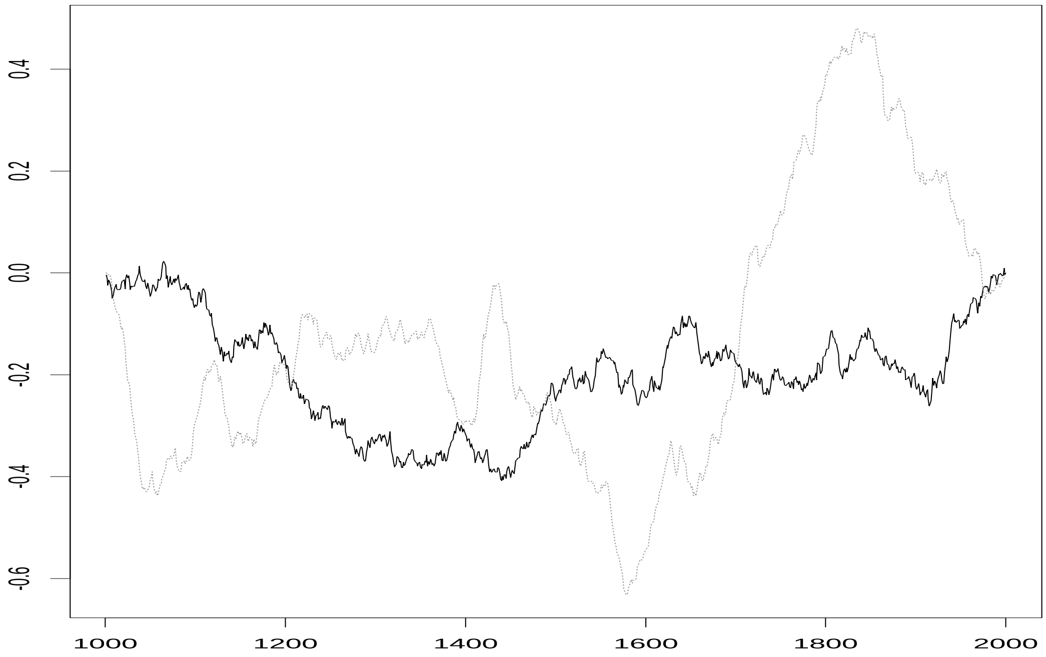

Figure 11: Ising model: second run. Solid line–benchmark path; dotted line – Gibbs sampler path

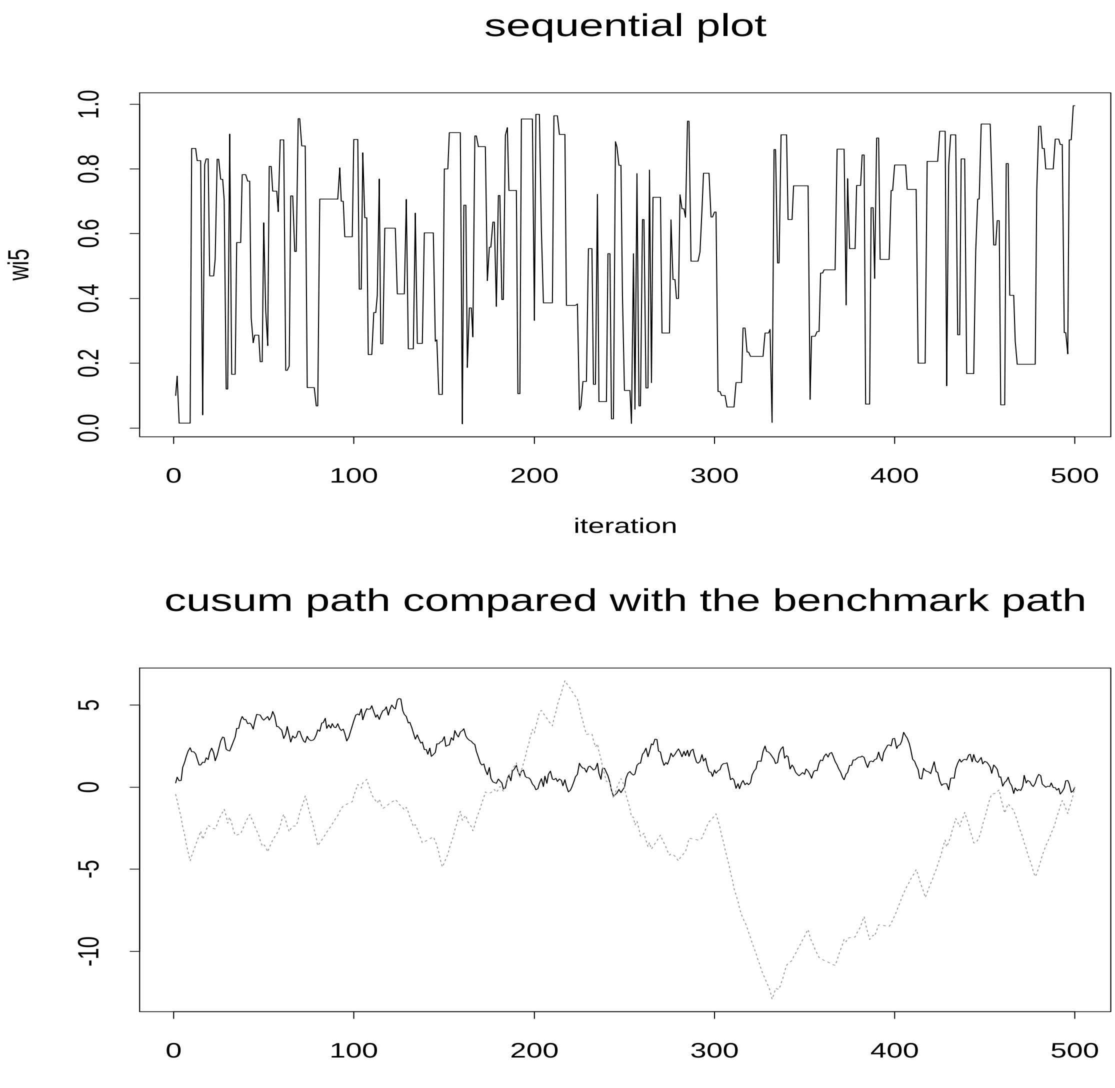

Figure 12: A uniform variant on Wicth's Hat: n=500, n0=1

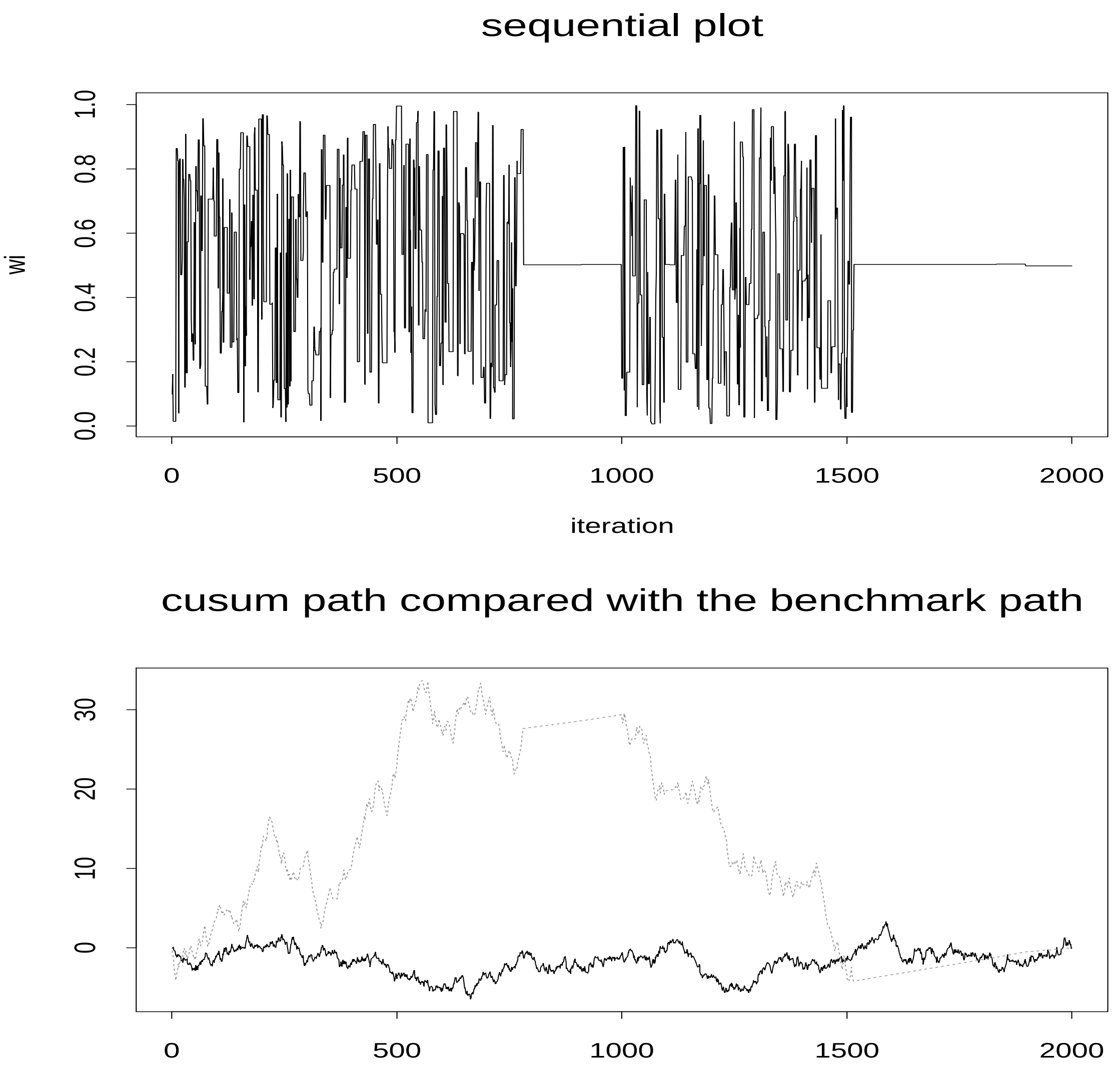

Figure 13: A uniform variant on Wicth's Hat: n=2000, n0=1