DRAM – efficient adaptive MCMC

Heikki Haario, Marko Laine, Antonietta Mira, Eero Saksman
University of Lappeenranta, Finland   University of Helsinki, Finland   University of Insubria, Italy   University of Jyväskylä, Finland

The success of MCMC methods, in general, depends on how well the proposal distribution fits the target distribution. We propose to combine two quite powerful ideas that have recently appeared in the Markov chain Monte Carlo literature: adaptive Metropolis samplers and delayed rejection.

DR – Delayed Rejection

Delayed rejection is a way of modifying the standard Metropolis-Hastings algorithm to improve efficiency of the resulting MCMC estimators relative to Peskun asymptotic variance ordering. The basic idea is that, upon rejection in a MH, instead of advancing time and retaining the same position, a second stage move is proposed. The acceptance probability of the second stage candidate is computed so that reversibility of the Markov chain relative to the distribution of interest is preserved. The second stage proposal is allowed to depend on the current position of the chain and also on what we have just proposed and rejected. The process of delaying rejection can be iterated for a fixed or random number of stages. In its basic formulation, DR employs a given number of fixed proposals that are used at the different stages.

AM – Adaptive Metropolis

The basic idea is to create a Gaussian proposal distribution with a covariance matrix calibrated using the sample path of the MCMC chain. The crucial point is how the covariance of the proposal distribution depends on the history of the chain. We take, possibly after an initial non-adaptation period, the proposal to be centered at the current position of the Markov chain, X_t, and compute the covariance from an incerasing part of the generated chain. In order to start the adaptation procedure an arbitrary strictly positive definite initial covariance, C_0, is chosen according to a priori knowledge (which may be quite poor). A time index, $t_0 > 0$, defines the length the initial non-adaptation period and we let

$$C_t = \begin{cases} C_0, & t \le t_0 \\ s_d \mathrm{Cov}(X_0, \ldots, X_{t-1}) + \varepsilon I_d, & t > t_0. \end{cases} \tag{1}$$

Scaling parameter s_d depends only on the dimension d of the state space where π is defined and $\varepsilon > 0$ is a constant that we may choose very small. This form of adaptation was proved to be ergodic in [2].

DRAM

The success of the DR strategy depends largely on the fact that at least one of the proposals is successfully chosen. The intuition behind adaptive strategies is to learn from the information obtained during the run of the chain, and, based on this, to tune the proposals to work more efficiently.

In the example, we shall combine AM adaptation with an m–stages DR algorithm in the following way:

- The proposal at the first stage of DR is adapted just as in AM: the covariance C_n^1 is computed from the points of the sampled chain, no matter at which stage these points have been accepted in the sample path.
- The covariance C_n^i of the proposal for the i:th stage $(i = 2, \ldots, m)$ is always computed simply as a scaled version of the proposal of the first stage, $C_n^i = \gamma_i C_n^1$.

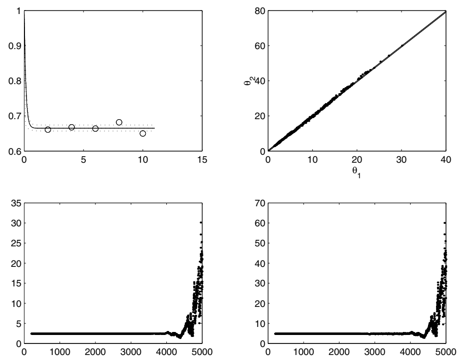
Example

This example presents a situation where neither AM nor DR works properly alone, but the combination DRAM has no difficulties. Consider a simple chemical reaction $A \underset{k_2}{\overset{k_1}{\rightleftarrows}} B$, where a component A goes to B in a reversible manner, with reaction rate coefficients k_1, k_2. So the dynamics is given by the ODE system

$$\frac{dA}{dt} = -k_1 A + k_2 B, \quad \frac{dB}{dt} = k_1 A - k_2 B$$

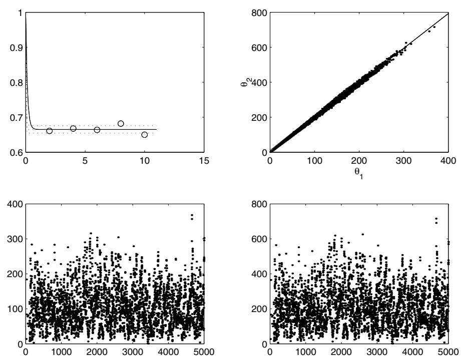
estimation task would be to find values for k_1 and k_2 when data for, e.g., $A(t) = k_2/(k_1 + k_2) + (A_0 - k_2/(k_1 + k_2))e^{-(k_1+k_2)t}$ has been obtained at given sampling times of t. Suppose that the data has been sampled too late, in the sense that the reaction already has reached a steady–state equilibrium at the sampling times.It is clear that from such data the values of the parameters can not be separately determined, only the ratio k_1/k_2 may be identified, as well as lower bounds for k_1 and k_2. With wide Gaussian prior on the parameters, the posterior distribution for k_1 and k_2 will be a thin, practically infinite 'zone' in a direction where k_1/k_2 is constant. As a test case, we try to find this posterior with DR, AM and DRAM.

AM



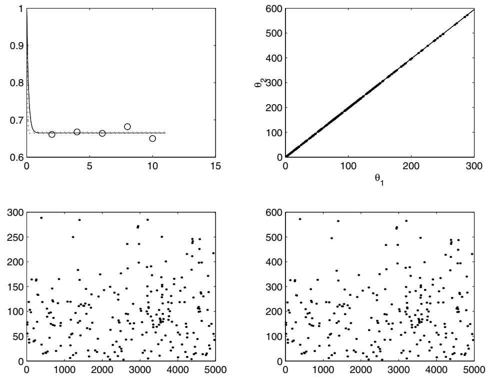
When using AM the chain starts to mix only after enough points has been gathered to allow the adaptation to start.

DRAM



Combining DR and AM makes the chain to adapt from right the begining, bacause DR ensures that enoug good points are acceppted.

DR



DR will mix from the begining of the run, but if none of the proposals are right the mixing can be slow.

References

[1] Green, P. J. and Mira, A. *Delayed rejection in reversible jump Metropolis-Hastings*. Biometrika, Vol. 88, pp. 1035-1053, 2001

[2] A. Mira: On Metropolis-Hastings algorithms with delayed rejection. *Metron*, Vol. LIX, n. 3-4, pp. 231-241, 2001.

[3] H. Haario, E. Saksman and J. Tamminen: *An adaptive Metropolis algorithm* Bernoulli 7, 223-242, 2001.

[4] H. Haario, M. Laine, A. Mira and E. Saksman: DRAM: Efficient adaptive MCMC, University of Helsinki, Department of Mathematics, preprint 374, 2003, submitted to *Statistics and Computing*.

[5] L. Tierney and A. Mira. Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine*, 18, pp. 2507-2515, 1999.

ADAPSki Bormio, Italy, January 9–11 2005