Handout for "Dynamic Striated Metropolis-Hastings Sampler for High-Dimensional Models"

Minsu Chang

April 25, 2015

This paper develops a generic posterior simulator called the dynamic striated Metropolis-Hastings (DSMH) sampler. Standard Markov Chain Monte Carlo (MCMC) methods, such as the Metropolis-Hastings algorithm, work well for estimating models with likelihoods or posterior distributions that have smooth Gaussian shapes. For high-dimensional economic and statistical models, however, the likelihood or the posterior distribution can be non-Gaussian with highly irregular shapes and multiple peaks. These problems can severely compromise the accuracy of previous MCMC samplers for Bayesian inference.

Dynamic striated Metropolis-Hastings (DSMH) sampler draws the strengths of two recently developed samplers: equi-energy (EE) (Kou, Zhou, and Wong, 2006) and sequential Monte Carlo (SMC) (Chopin, 2004; Durham and Geweke, 2012; Herbst and Schorfheide, 2014) simulators. The basic idea behind these two techniques is to start with a tractable initial distribution one can sample from and then to transform this initial distribution gradually to the desired posterior distribution through a sequence of stages. At each stage the sample from the previous stage is used to form a new sample for the current stage. At the final stage the sample comes from the desired posterior distribution.

They divide the target distribution into various levels and define a striation as a parameter space in which the posterior kernel is between the two adjacent levels. Striations are dynamically adjusted when we move from one stage to another. This dynamic adjustment ensures that each striation remains fully populated by independent striated draws within the striation.

Generic Algorithm

Let $Y_T = (y_1, ..., y_T)$ denote the observable variables, where T is the total number of observations and y_t denotes an $n \times 1$ vector of variables observed at time t. The likelihood function is denoted by

$p(Y_T \mid \theta)$. Combining the likelihood and the prior probability density $\pi(\theta)$, we obtain the posterior kernel $p(Y_T \mid \theta)\pi(\theta)$. To simplify notation, denote the posterior kernel by $p(\theta)$.

(1) Stages

We transform the posterior distribution by tempering the probability kernel $p(\theta)$. For any real number λ satisfying $0 < \lambda \leq 1$, define $f_\lambda(\theta) = p(\theta)^\lambda$. To define stages, choose λ_i, for $1 \leq i \leq H$, such that $0 < \lambda_1 < \ldots < \lambda_{H-1} < \lambda_H = 1$. For $1 \leq i \leq H$,the target distribution for the ith stage is $f_{\lambda_i}(\theta)$. Note that the final target distribution $f_{\lambda_H}(\theta)$ is the posterior kernel as required.

(2) Striations

A striation is the set of all values of θ that have similar posterior values. Striations at the ith stage are defined by a sequence of M + 1 levels, denoted by $L_{i,k}$, satisfying $0 = L_{i,0} < L_{i,1} < \ldots < L_{i,M-1} < L_{i,M} = \infty$. For $1 \leq k \leq M$,the kth striation is the set

$$S_{i,k} = \{\theta \in \Theta \mid L_{i,k-1} \leq p(\theta) \leq L_{i,k}\}.$$

We choose the levels so that the probability that $\theta \in S_{i,k}$ is equal to 1/M. This probability is with respect to the distribution at the previous stage. If

$$I_{i-1} = \int_{\theta\in\Theta} f_{\lambda_{i-1}}(\theta) d\theta,$$

the levels are chosen to satisfy

$$\frac{1}{M} = \int_{\theta\in S_{i,k}} \frac{f_{\lambda_{i-1}}(\theta)}{I_{i-1}} d\theta$$

(3) Metropolis-Hastings

The proposal distribution is a mixture of a Gaussian and $f_{\lambda_{i1}}(\theta)$, the distribution at the previous stage. If $\theta^* \equiv \theta^{(i,l)}$ is the most recent draw from the Metropolis-Hastings sampler at the ith stage, the proposal density of θ given θ^* is

$$g_i(\theta, \theta^*) = (1-p)\phi_{c_i\Omega_i}(\theta - \theta^*) + p\chi(\theta, \theta^*)\frac{M f_{\lambda_{i-1}}(\theta)}{I_{i-1}}$$

where $\phi_{c_i\Omega_i}(\cdot)$ is the density of the mean-zero Gaussian distribution with variance $c_i\Omega_i$ and $\chi(\theta, \theta^*)$ is the indicator function that returns one if θ and θ^* are in the same striation and zero otherwise. The mixture form of $g_i(\theta, \theta^*)$ indicates that with probability $1 - p, \theta$ is drawn from the Gaussian distribution centered at θ^* and with probability p, θ is drawn from the distribution at the previous stage but from the same striation that contains θ^*.[1]

[1] In the paper, they also have theoretical justification of this sampler's convergence and practical issues related to tuning parameter choice, parallelization, and so on.

Application

They present two simultaneous-equation high-dimensional models for the purpose of testing the DSMH sampler: a SVAR model and a Markov-switching SVAR model. They show that the exact Gibbs sampler exists at every stage of the sampler. This allows to obtain accurate posterior draws from the Gibbs sampler at each stage and compare this true distribution to the distribution simulated from the DSMH sampler.

TABLE 2. Estimated log integral constants ($\log I_i$) and log marginal data densities (at the final stage)

Stages	"Truth"	DSMH	IW	$NSE^{(IW)}$
15	993.04	992.84	1059.77	0.28
16	978.07	977.83	1045.33	0.21
17	963.20	963.00	1030.54	0.16
18	948.45	948.32	1015.64	0.14
19	933.84	933.69	1000.85	0.15
20	919.42	919.30	986.25	0.15
21	905.23	905.11	971.89	0.13
22	891.35	891.21	957.86	0.17
23	877.84	877.71	944.20	0.15
24	864.82	864.70	931.00	0.16
25	852.44	852.29	918.44	0.14
26	840.84	840.72	906.68	0.16
27	830.30	830.16	895.93	0.17
28	821.09	820.94	886.55	0.18
29	813.66	813.52	878.96	0.18
30	808.53	808.38	873.69	0.17
31	806.50	806.33	871.47	0.18
32	808.57	808.40	873.37	0.20
33	816.29	816.06	880.88	0.18
34	831.73	831.52	896.17	0.21
35	858.23	858.02	922.46	0.20
36	900.88	900.63	964.93	0.24
37	968.34	968.11	1032.19	0.27
38	1076.13	1075.74	1139.70	0.32
39	1255.27	1254.95	1318.74	0.46
40	1578.83	1578.28	1642.04	0.60
41	2266.42	2265.48	2328.81	1.23
42	4421.96	4421.17	4479.37	2.23