Markov Chain Monte Carlo

some more advanced methodology

Introduction

- Markov chain Monte Carlo (MCMC):
  - We need an ergodic and π-invariant Markov probability kernel K
  - Use sample averages from this Markov chain to compute expectations
- So far have seen only basics
  - Metropolis Hastings, Gibbs sampling
- Numerous possible extensions and algorithms

Outline & Reading List

- Computing the normalising constant
  - Chib & Jeliazkov (2001) Marginal Likelihood From the Metropolis–Hastings Output, JASA, Vol 96, Issue 453, 270-281
- Adaptive MCMC
  - Andrieu, C. & Thoms, J. (2008). A tutorial on adaptive MCMC. Stat. & Comp., 18(4), 343-373.
- Pseudo-marginal MCMC
  - Andrieu and Roberts, (2009) The pseudo-marginal approach for efficient Monte Carlo computations, Ann. Stat., Vol, No 2, 697-725.

Computing the normalising constant

- In contrast to Importance sampling computing the normalising constant is not natural here.
- One possibility is harmonic mean estimator:
  - set test function $\varphi = \frac{q}{\gamma}$, integrate and invert:
  - quite problematic, can lead even to infinite estimator variance.
- A more interesting approach appeared in
  - (Chib 95, Chib and Jeliazkov 01)
  - and extended in (Mira and Nicholls 04)

Computing the normalising constant

- Basic idea: consider the identities for any point $x' \neq x$

$$Z = \frac{\gamma(x')}{\pi(x')} \tag{1}$$

  and use detailed balance to construct

$$\pi(x') = \frac{\int \alpha(x, x') q(x, x') \pi(x) dx}{\int \alpha(x', x) q(x', x) dx}$$

- Numerator is integral w.r.t π (so can use MCMC output), denominator integral w.r.t q

Computing the normalising constant

- Pick a point x' near the mode of π and estimate $\pi(x')$ as

$$\frac{\frac{1}{N}\sum_{i=1}^{N}\alpha(X_i, x')q(X_i, x')}{\frac{1}{N'}\sum_{j=1}^{N'}\alpha(x', Y_j)}$$

  and use this in (1).
- Here:
  - X_n is output of MCMC chain
  - Y_j can be i.i.d samples of $q(x', \cdot)$
- One can construct a similar method for Gibbs samplers using full conditionals and Rao Blackwelisation
  - Chib, S. (1995), "Marginal Likelihood from the Gibbs Output," Journal of the American Statistical Association, 90, 1313–1321

Adaptive MCMC

- Another question very relevant in the multivariate case is how to choose the scaling in the proposals?
  - theory suggests a good proposal should have a similar covariance with the target
- Consider

$$Q(x, \cdot) = \mathcal{N}(x, \frac{\varrho^2}{d}\Sigma)$$

  Is it possible to use the output of the chain to construct Σ?

Adaptive MCMC

- (Haario , Saksman , Tamminen 01) proposed to use stochastic approximation:

$$\mu_n = \mu_{n-1} + \gamma_n (X_n - \mu_{n-1})$$
$$\Sigma_n = \Sigma_{n-1} + \gamma_n \left((X_n - \mu_n) (X_n - \mu_n)^T - \Sigma_{n-1} \right)$$

- The next question is: given we know what value α should be, is it possible to vary on-line the step size ϱ to achieve this?
  - Use again Robbins Monro

$$\varrho_n = \varrho_{n-1} + \gamma_n (\alpha(\varrho_{n-1}) - 0.234)$$

Adaptive MCMC

- Originally $\gamma_n = \frac{1}{n}$ but it was soon noticed that the convergence of this algorithm can be unstable
  - (X_n) process is no longer a Markov Chain
- This was sorted using a more careful adaptation approach: "diminishing adaptation",
  - see (Andrieu & Thoms 08) for a tutorial.
- Interesting topic related to theory and practice:
  - see works from Andrieu, Vihola, Moulines, Atchade, Latuszynski, Roberts, Rosenthal, Haario, Saksman , Tamminen,....

The pseudo-marginal sampler

- In many cases one cannot compute γ pointwise
- Common case is intractable likelihoods in Bayesian inference:

$$\gamma(x) = p(y|x)p(x)$$

and

$$p(y|x) = \int p(y, z|x)dz$$

- This is a very common situation, but often one may have available unbiased estimates:
  - importance sampling
  - particle filtering (Particle MCMC)

The pseudo-marginal sampler

- (Andrieu & Roberts 09, extending Beaumont 03)
- Let $\hat{\mathcal{Z}}_x$ be an unbiased estimate of $\mathcal{Z}_x = p(y|x)$.
- In the MH algorithm replace

$$\alpha(x, x') = 1 \wedge \frac{\mathcal{Z}_{x'} p(x') q(x', x)}{\mathcal{Z}_x p(x) q(x, x')}$$

  with

$$\widetilde{\alpha}(x, x') = 1 \wedge \frac{\hat{\mathcal{Z}}_{x'} p(x') q(x', x)}{\hat{\mathcal{Z}}_x p(x) q(x, x')}$$

  (x' is proposed state here)

- It turns out that such a algorithm has the right invariant distribution
  - more precisely admits π as a marginal

The pseudo-marginal sampler

- $\hat{\mathcal{Z}}_x$ can be obtained using IS
  - Sample $Z_n^i \sim q(\cdot|x),\ i = 1, \ldots, L$
  - Compute

$$\hat{\mathcal{Z}}_x = \frac{1}{L}\sum_{i=1}^{L} w(y, Z_n^i, x), \quad \text{with } w(y, z, x) = \frac{p(y, z|x)}{q(z|x)}$$

- Other methods also possible as long as $\hat{\mathcal{Z}}_x$ is unbiased

The pseudo-marginal sampler

- In fact we are running an MCMC algorithm with state being

$$(X_n, Z_n^1, \ldots, Z_n^L)$$

where $Z_n^i \sim p(\cdot|x')$ are the ingredient in computing $\mathcal{Z}_{x'}$

- Consider this target

$$\tilde{\pi}(x, z^1, \ldots, z^L) \propto p(x) \left(\frac{1}{L} \sum_{i=1}^{L} w(y, z^i, x) \right) \prod_{i=1}^{L} q(z^i|x)$$

The pseudo-marginal sampler

- Note because of unbiasedness of the normalising const.

$$\pi(x) = \int \tilde{\pi}(x, z^1, \dots, z^L) dz^1 \cdots dz^L$$

- So we can MCMC algorithm targetting $\tilde{\pi}$
  - in practice marginalisation means we ignore samples for $z^1, \dots, z^L$
  - At each iteration we propose

    $$(X_n' = x', Z_n^1, \dots, Z_n^L)$$

    using

    $$q(x_{n-1}, x') \prod_{i=1}^{L} q(Z_n^i | x')$$

  - Each of the $Z_n^i \sim q(\cdot|x')$ are used in computing $\hat{\mathcal{Z}}_{x'}$

The pseudo-marginal sampler

Next write the MH acceptance ratio

$$
\begin{aligned}
\widetilde{\alpha}(x,x') =& \\
& 1 \wedge \frac{p(x')\prod_{i=1}^{L} q(Z_n^i|x')\left(\frac{1}{L}\sum_{i=1}^{L} w(y, Z_n^i, x')\right)}{p(x)\prod_{i=1}^{L} q(Z_{n-1}^i|x)\left(\frac{1}{L}\sum_{i=1}^{L} w(y, Z_{n-1}^i, x)\right)} \\
& \times \frac{q(x',x)\prod_{i=1}^{L} q(Z_{n-1}^i|x)}{q(x,x')\prod_{i=1}^{L} q(Z_n^i|x')} \\
=& 1 \wedge \frac{\hat{\mathcal{Z}}_{x'}\, p(x')q(x',x)}{\hat{\mathcal{Z}}_x\, p(x)q(x,x')}
\end{aligned}
$$

The pseudo-marginal sampler

- MCMC kernel is reversible to

$$\tilde{\pi}(x, z^1, \ldots, z^L)) \propto p(x) \left(\frac{1}{L} \sum_{i=1}^{L} p(y|z^i, x) \right) \prod_{i=1}^{L} p(z^i|x)$$

and due to unbiasedness of $\hat{\mathcal{Z}}_{x'}$ integrating z^i-s gives π.

The pseudo-marginal sampler

- The mixing of the pseudo-marginal algorithm depends on
  - The mixing of the true-marginal MCMC algorithm.
  - The variance in the estimators of the marginal likelihood
- Question between adding more computation via number of MCMC iterations N or auxiliary variables L
  - (Doucet, Pitt, Deligiannidis & Kohn 15),
  - (Sherlock, Thiery, Roberts, Rosenthal 15)
- Correlating can improve dependence on size of y
  - (Deligiannidis, Doucet, Pitt 17)

Discussion

- These topics were chosen on the basis
  - simplicity of presentation and connection with previous topics
  - convey basic concepts
  - general usefulness and possibility to extend
- Various more advanced extensions are available in many directions