Intractable likelihoods with the pseudo-marginal MCMC

Matti Vihola

OxWaSP Computational Statistics and Statistical Computing module, 14 Oct 2014

The standard MCMC setting

- Aim: Approximate

$$\pi(f) = \int_{\mathsf{X}} f(x)\pi(x)\mathrm{d}x,$$

where the probability density π is known up to a normalising constant.

- The standard approach: Run MCMC $(X_k)_{k\geq 1}$ (typically Metropolis-Hastings) with target density π, and compute

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \approx \pi(f).$$

The standard MCMC algorithm

This is the Metropolis-Hastings algorithm we usually implement

Marginal algorithm P

- Draw sample from the proposal density, $Y_n \sim q(X_{n-1}, \,\cdot\,)$.
- Accept the move (set $X_n \leftarrow Y_n$) with probability

$$\min\left\{1, \frac{\pi_u(Y_n)}{\pi_u(X_{n-1})}\frac{q(Y_n, X_{n-1})}{q(X_{n-1}, Y_n)}\right\},$$

otherwise set $X_n \leftarrow X_{n-1}$.

π_u is the unnormalised density, $\pi_u(x) = c\pi(x)$

- For example, $\pi_u(x) = p(y_{\text{obs}} \mid x)p(x)$ and $\pi(x) = p(x \mid y)$.

The standard MCMC algorithm

This is the Metropolis-Hastings algorithm we usually implement

Marginal algorithm P

- Draw sample from the proposal density, $Y_n \sim q(X_{n-1}, \,\cdot\,)$.
- Accept the move (set $X_n \leftarrow Y_n$) with probability

$$\min\left\{1, \frac{\pi_u(Y_n)}{\pi_u(X_{n-1})} \frac{q(Y_n, X_{n-1})}{q(X_{n-1}, Y_n)}\right\},$$

otherwise set $X_n \leftarrow X_{n-1}$.

π_u is the unnormalised density, $\pi_u(x) = c\pi(x)$

- For example, $\pi_u(x) = p(y_{\text{obs}} \mid x)p(x)$ and $\pi(x) = p(x \mid y)$.

What if $\pi_u(\,\cdot\,)$ cannot be computed?

Marginal inference with MCMC

- Suppose the unnormalised density is defined through an integral (over the latent variables)

$$\pi_u(x) = \int \pi'_u(x, z)\mathrm{d}z.$$

  - For example $\pi'_u(x, z) = p(y_{\mathsf{obs}} \mid x, z)p(x, z) \propto p(x, z \mid y)$ and $\pi(x) = p(x \mid y)$.
- The standard approach: Run MCMC $(X_k, Z_k)_{k\geq 1}$ targeting a joint probability density $\pi'(x, z) \propto \pi'_u(x, z)$, and then compute

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \approx \pi(f).$$

Problems with marginal inference

- High-dimensional latent variables 'Z_k' $\implies$ slowly mixing MCMC.
  - Generally difficult to design efficient MCMC in high-dimensional situations.
- The latent variables may be impossible to simulate (e.g. infinite-dimensional...)

Problems with marginal inference

- High-dimensional latent variables 'Z_k' $\implies$ slowly mixing MCMC.
  - Generally difficult to design efficient MCMC in high-dimensional situations.
- The latent variables may be impossible to simulate (e.g. infinite-dimensional. . .)

Naive idea: Approximate $\pi_u(\,\cdot\,)$ in the marginal algorithm. . .

Pseudo-marginal MCMC

Suppose we can generate non-negative unbiased estimates:

$$T_x \geq 0, \qquad \mathbb{E}[T_x] = \pi_u(x) \qquad \forall x \in \mathsf{X}$$

Pseudo-marginal algorithm $\tilde{P}$

- Draw sample from the proposal density, $Y_n \sim q(X_{n-1}, \,\cdot\,)$ and generate $S_n \geq 0$ with $\mathbb{E}[S_n] = \pi_u(Y_n)$.
- Set $(X_n, T_n) \leftarrow (Y_n, S_n)$ with probability

$$\min\left\{1, \frac{S_n}{T_{n-1}} \frac{q(Y_n, X_{n-1})}{q(X_{n-1}, Y_n)}\right\},$$

otherwise set $(X_n, T_n) \leftarrow (X_{n-1}, T_{n-1})$.

If the estimates are perfect, $T_x \equiv \pi_u(x)$, then $S_n = \pi_u(Y_n)$ and $T_{n-1} = \pi_u(X_{n-1})$
$\implies$ $\tilde{P}$ coincides with the marginal algorithm P.

Example run of a pseudo-marginal

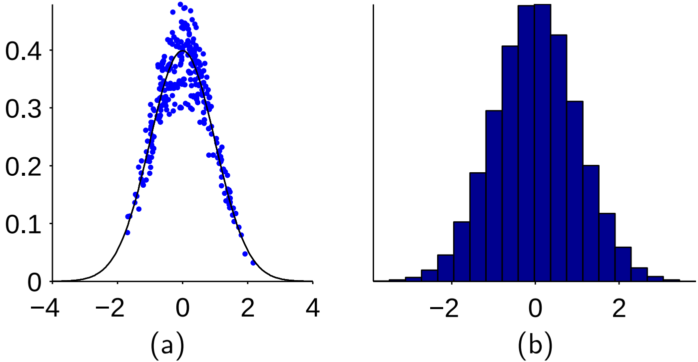

Figure: (a) Samples (X_k, T_k) (blue) and the true density π (black) (b) Histogram over 100000 samples (X_k).

The pseudo-marginal algorithm is valid MCMC

- Straightforward to check that the pseudo-marginal chain has a unique target distribution $\tilde{\pi}(x,t)$ satisfying

$$\pi(x) = \int \tilde{\pi}(x,t)\mathrm{d}t. \qquad \text{Correct marginal}$$

- Consequently,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{n\to\infty} \pi(f) \qquad (a.s.)$$

(given that the chain is $\tilde{\pi}$-irreducible, for which it is enough that the marginal chain is π-irreducible.)

- Despite of the approximations of $T_x \approx \pi_u(x)$, the method is exact! (in the sense of the strong law above).

Grouped independence Metropolis-Hastings (GIMH) (Beaumont, *Genetics*, 2003)

- Unbiased estimates from importance sampling

$$T_x = \frac{1}{m} \sum_{j=1}^{m} \frac{\pi'_u(x, Z_j)}{h_x(Z_j)} \quad \text{where} \quad Z_j \overset{\text{i.i.d.}}{\sim} h_x(\,\cdot\,).$$

- The *importance densities* $h_x(\,\cdot\,)$ for each x must satisfy $\mathrm{supp}(h_x) \supset \mathrm{supp}(\pi'_u(x,\,\cdot\,))$[1]

Prove that T_x is unbiased...

[1] $\mathrm{supp}(f) = \{x \,:\, f(x) > 0\}$.

Approximate Bayesian Computation MCMC (Marjoram, Molitor, Plagnol & Tavaré, *PNAS*, 2003)

- Interested in $\pi(x) \propto p(y_{\textsf{obs}} \mid x)p(x)$.
- Easy to simulate samples Y from $p(y \mid x)$.
- Consider a modified *approximate* posterior

$$\pi_\epsilon(x) \propto p(x) \int \mathbb{I}\{d(y, y_{\textsf{obs}}) \leq \epsilon\} p(y \mid x) \mathrm{d}y,$$

  where $\epsilon > 0$ is a *tolerance* parameter and $d(y, y')$ is some *distance* metric between two 'data' y and y'.
- It is possible to do inference over the ABC posterior $\pi_\epsilon(x)$ by pseudo-marginal MCMC:

$$T_x = p(x)\left(\frac{1}{m}\sum_{k=1}^{m} \mathbb{I}\{d(Y_k, y_{\textsf{obs}}) \leq \epsilon\}\right), \qquad Y_k \overset{\textsf{i.i.d.}}{\sim} p(y \mid x)$$

ABC ingredients

The distance metric

- Usually $d(y, y') = \|\theta(y) - \theta(y')\|$, where $\theta(y) \in \mathbb{R}^d$ are some statistics calculated from the data y.
- Often $\theta(y)$ are *not sufficient* (and $d \ll \dim(y)$)
  $\implies$ already a (coarse) approximation made here!

The tolerance

- The smaller $\epsilon > 0$ is, the smaller the (further) approximation error is.

  If θ are sufficient & further regularity conditions hold, then $\pi_\epsilon \to \pi$ as $\epsilon \to 0$.
- The smaller $\epsilon > 0$ is, the more inefficient the MCMC is (acceptance rate goes down).

The MCMC algorithm

- Some guidelines available from theoretical findings. . .

Other examples of pseudo-marginal algorithms

- Particle marginal Metropolis-Hastings (Andrieu, Doucet & Holenstein, *JRSS B* read paper, 2010).
- Statistical inference in diffusion models (Beskos, Papaspiliopoulos, Roberts & Fearnhead, *JRSS B* read paper, 2006).
- Model selection (Andrieu & Roberts, *Ann. Statist.*, 2009).
- ...

Practical

- Take a look at the original papers:
  - Beaumont, *Genetics*, 2003
  - Marjoram, Molitor, Plagnol & Tavaré, *PNAS*, 2003
- Implement the GIMH and the ABC-MCMC on some problem.
  - You can, for example, look at the following toy example

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(- \frac{x^2}{2\sigma_x^2} \right),$$
$$p(y \mid x) = \frac{1}{\sqrt{2\pi}} \exp\left(- \frac{(y-x)^2}{2\sigma_y^2} \right).$$

  (Feel free to study a more interesting model!)
- Test how different importance densities h_x perform in the GIMH. What seems the best?
- Test how choosing different values for $\epsilon > 0$ affect the accuracy of the ABC & your simulation efficiency.

References

M. A. Beaumont. Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164:1139–1160, 2003.

P. Marjoram, J. Molitor, V. Plagnol and S. Tavaré. Markov chain Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA*, 100: 15324–15328, 2003.

C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, 37(2):697–725, 2009.

L. Bornn, N. Pillai, A. Smith and D. Woodard. One Pseudo-Sample is Enough in Approximate Bayesian Computation MCMC. Preprint arXiv:1404.6298, 2014.