An ABC interpretation of the multiple auxiliary variable method

by

Dennis Prangle and

2 Background

2.1 Auxiliary variable methods

This is often referred to as the ABC likelihood. It is proportional to a convolution of the likelihood with a uniform density, evaluated at y. For > 0 this is generally an inexact approximation to the likelihood. For discrete data it is possible to use $= 0$ in which case the ABC likelihood equals the exact likelihood, and so L_{ABC} is unbiased.

For MRFs empirically it is observed that, compared with competitors such as the exchange algorithm (Murray et al., 2006), ABC requires a relatively large number of simulations to yield an e cient algorithm (Friel, 2013).

3 Derivation

3.1 ABC for MRF models

Suppose that the model $f(y|\ )$ has an intractable likelihood but can be targeted by a MCMC chain $x = (x_1; x_2; \ldots; x_n)$. Let represent densities relating to this chain. Then $_n(y|\ ) := \ (x_n = y|\ )$ is an approximation of $f(y|\ )$ which can be estimated by ABC. For now suppose that y is discrete and consider the ABC likelihood estimate requiring an exact match: simulate from $(x|\ )$ and return $(x_n = y)$. We will consider an IS variation on this: simulate from $g(x|\ )$ and return $(x_n = y)\ (x|\ )=g(x|\ )$. Under the mild assumption that $g(x|\ )$ has the same support as $(x|\ )$ (typically true unless n is small), both estimates have the expectation $\Pr(x_n = y|\ )$.

This can be generalised to cover continuous data using the identity

$$_n(y|\ ) = \int_{x_n = y} (x|\ )dx_{1:n\ 1};$$

where $x_{i:j}$ represents $(x_i; x_{i+1}; \ldots; x_j)$. An importance sampling estimate of this integral is

$$w = \frac{(x|\ )}{g(x_{1:n\ 1}|\ )} \qquad (3)$$

where x is sampled from $g(x_{1:n\ 1}|\ )\ (x_n = y)$, with representing a Dirac delta measure. Then, under mild conditions on the support of g, w is an unbiased estimate of $_n(y|\ )$.

The ideal choice of $g(x_{1:n\ 1}|\ )$ is $(x_{1:n\ 1}|x_n;\ )$, as then $w = \ (x_n = y|\ )$ exactly. This represents sampling from the Markov chain conditional on its nal state being y.

3.2 Equivalence to MAV

We now show that natural choices of $(x|\ )$ and $g(x_{1:n\ 1}|\ )$ in the ABC method just outlined

reTJ /R66 7.97011056]TJ -299552 T8586(x)4.0f 12.286(x)4./837.3294]TJ 9 Td [(x)3.93724]TJ /R70 7.9

Here $\pi(x|\theta)$ defines a MCMC chain with transitions $K_i(x_{i+1}|x_i)$. Suppose K_i is as in Section 2.1 for $i \leq a$, and for $i > a$ it is a reversible Markov kernel with invariant distribution $f(\cdot|\theta)$. Also assume $b := n - a \to \infty$. Then the MCMC chain ends in a long sequence of steps targeting $f(\cdot|\theta)$ so that $\lim_{n\to\infty} \pi_n(\cdot|\theta) = f(\cdot|\theta)$. Thus the likelihood being estimated converges on the true likelihood for large n. Note this is the case even for fixed a.

The importance density $g(x_{1:n-1}|\theta)$ specifies a reverse time MCMC chain starting from $x_n = y$ with transitions $K_i(x_i|x_{i+1})$. Simulating x is straightforward by sampling x_{n-1}, then x_{n-2} and so on. This importance density is an approximation to the ideal choice stated at the end of Section 3.1.

The resulting likelihood estimator is

$$w = f_1(x_1|\theta;y) \prod_{i=1}^{n-1} \frac{K_i(x_{i+1}|x_i)}{K_i(x_i|x_{i+1})}.$$

Using detailed balance gives

$$\frac{K_i(x_{i+1}|x_i)}{K_i(x_i|x_{i+1})} = \frac{f_i(x_{i+1}|\theta;y)}{f_i(x_i|\theta;y)} = \frac{\gamma_i(x_{i+1}|\theta;y)}{\gamma_i(x_i|\theta;y)},$$

so that

$$w = f_1(x_1|\theta;y) \prod_{i=1}^{n-1} \frac{\gamma_i(x_{i+1}|\theta;y)}{\gamma_i(x_i|\theta;y)} = \gamma(y|\theta) \prod_{i=2}^{n} \frac{\gamma_{i-1}(x_i|\theta;y)}{\gamma_i(x_i|\theta;y)}.$$

This is an unbiased estimator of $\pi_n(y|\theta)$. Hence

$$v = \prod_{i=2}^{n} \frac{\gamma_{i-1}(x_i|\theta;y)}{\gamma_i(x_i|\theta;y)} = \prod_{i=2}^{a} \frac{\gamma_{i-1}(x_i|\theta;y)}{\gamma_i(x_i|\theta;y)}.$$

is an unbiased estimator of $\pi_n(y|\theta)/\gamma(y|\theta) \to 1/Z(\theta)$. In the above we have assumed, as in Section 3.1, that γ_1 is normalised. When this is not the case then we instead get an estimator of $Z(\tilde{\theta})/Z(\theta)$, as for MAV methods. Also [redacted]ote that in either case a valid estimator is produced for any choice of $\tilde{\theta}$.

The ABC estimate can be viewed by a two sta[redacted]e procedure. First run a MCMC chain of length b with any starting value, targeting $f(\cdot|$ [redacted] $;y)$

References

Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, pages 697–725.