Adaptive MCMH for Bayesian inference of spatial autologistic models

discussion by Pierre E. Jacob & Christian P. Robert

Université Paris-Dauphine and CREST

Adap'skiii, Park City, Utah, Jan. 03, 2011

Context: untractable likelihoods

- Cases when the likelihood function $f(\mathbf{y}|\theta)$ is unavailable and when the completion step

$$f(\mathbf{y}|\theta) = \int_{\mathscr{Z}} f(\mathbf{y}, \mathbf{z}|\theta) \, \mathrm{d}\mathbf{z}$$

is impossible or too costly because of the dimension of $\mathbf{z}$

Context: untractable likelihoods

- Cases when the likelihood function $f(\mathbf{y}|\theta)$ is unavailable and when the completion step

$$f(\mathbf{y}|\theta) = \int_{\mathscr{Z}} f(\mathbf{y}, \mathbf{z}|\theta)\, \mathrm{d}\mathbf{z}$$

is impossible or too costly because of the dimension of $\mathbf{z}$
- © MCMC cannot be implemented!

Context: untractable likelihoods

- Cases when the likelihood function $f(\mathbf{y}|\theta)$ is unavailable and when the completion step

$$f(\mathbf{y}|\theta) = \int_{\mathscr{Z}} f(\mathbf{y}, \mathbf{z}|\theta)\, \mathrm{d}\mathbf{z}$$

  is impossible or too costly because of the dimension of $\mathbf{z}$
- © MCMC cannot be implemented!
- $\Longrightarrow$ adaptive MCMH

Compare ABC with adaptive MCMH?

ABC algorithm

For an observation $\mathbf{y} \sim f(\mathbf{y}|\theta)$, under the prior $\pi(\theta)$, keep *jointly* simulating

$$\theta' \sim \pi(\theta)\,, \mathbf{z} \sim f(\mathbf{z}|\theta')\,,$$

until the auxiliary variable $\mathbf{z}$ is equal to the observed value, $\mathbf{z} = \mathbf{y}$.

[Tavaré et al., 1997]

Compare ABC with adaptive MCMH?

ABC algorithm

For an observation $\mathbf{y} \sim f(\mathbf{y}|\theta)$, under the prior $\pi(\theta)$, keep *jointly* simulating

$$\theta' \sim \pi(\theta)\,, \mathbf{z} \sim f(\mathbf{z}|\theta')\,,$$

until the auxiliary variable $\mathbf{z}$ is equal to the observed value, $\mathbf{z} = \mathbf{y}$.

[Tavaré et al., 1997]

- When y is a continuous random variable, replace the "$z = y$" condition by $\rho(z, y) < \varepsilon$. Then ABC becomes an approximation.

Compare ABC with adaptive MCMH?

ABC algorithm

For an observation $\mathbf{y} \sim f(\mathbf{y}|\theta)$, under the prior $\pi(\theta)$, keep *jointly* simulating

$$\theta' \sim \pi(\theta)\,, \mathbf{z} \sim f(\mathbf{z}|\theta')\,,$$

until the auxiliary variable $\mathbf{z}$ is equal to the observed value, $\mathbf{z} = \mathbf{y}$.

[Tavaré et al., 1997]

- When y is a continuous random variable, replace the "$z = y$" condition by $\rho(z, y) < \varepsilon$. Then ABC becomes an approximation.
- ABC can be applied as long as samples from f can be drawn. $\Longrightarrow$ comparison with adaptive MCMH?

Some questions about the estimator $\hat{R}(\theta_t, \vartheta)$

$$\hat{R}(\theta_t, \vartheta) = \frac{1}{m_0 + m_0 \sum_{\theta_i \in S_t \setminus \{\theta_t\}} I(||\theta_i - \vartheta|| \leq \eta)}$$

$$\times \left\{ \sum_{\theta_i \in S_t \setminus \{\theta_t\}} \left[I(||\theta_i - \vartheta|| \leq \eta) \sum_{j=1}^{m_0} \frac{g(z_j^{(i)}, \theta_t)}{g(z_j^{(i)}, \vartheta)} \right] + \sum_{j=1}^{m_0} \frac{g(z_j^{(t)}, \theta_t)}{g(z_j^{(t)}, \vartheta)} \right\}$$

- Why does it bypass the number of repetitions of the θ_i's?

Some questions about the estimator $\hat{R}(\theta_t, \vartheta)$

$$\hat{R}(\theta_t, \vartheta) = \frac{1}{m_0 + m_0 \sum_{\theta_i \in S_t \setminus \{\theta_t\}} I(||\theta_i - \vartheta|| \leq \eta)}$$
$$\times \left\{ \sum_{\theta_i \in S_t \setminus \{\theta_t\}} \left[I(||\theta_i - \vartheta|| \leq \eta) \sum_{j=1}^{m_0} \frac{g(z_j^{(i)}, \theta_t)}{g(z_j^{(i)}, \vartheta)} \right] + \sum_{j=1}^{m_0} \frac{g(z_j^{(t)}, \theta_t)}{g(z_j^{(t)}, \vartheta)} \right\}$$

- Why does it bypass the number of repetitions of the θ_i's?
- Why resample the $y_j^{(i)}$'s only to inverse-weight them later? Why not stick to the original sample and weight the $y_j^{(i)}$'s directly? This would reduce the variance of $\hat{R}$ and prevent one from having to choose m_0.

Combine adaptations?

Adaptive MCMH looks great, but it adds new tuning parameters compared to standard MCMC!

- It would be nice to be able to adapt the proposal $\mathcal{Q}(\theta_t, \vartheta)$, e.g. to control the acceptance rate, as in other adaptive MCMC algorithms.

Combine adaptations?

Adaptive MCMH looks great, but it adds new tuning parameters compared to standard MCMC!

- It would be nice to be able to adapt the proposal $\mathcal{Q}(\theta_t, \vartheta)$, e.g. to control the acceptance rate, as in other adaptive MCMC algorithms.
- Then could η be adaptive, since it does as well depend on the (unknown) scale of the posterior distribution of θ ?

Combine adaptations?

Adaptive MCMH looks great, but it adds new tuning parameters compared to standard MCMC!

- It would be nice to be able to adapt the proposal $\mathcal{Q}(\theta_t, \vartheta)$, e.g. to control the acceptance rate, as in other adaptive MCMC algorithms.
- Then could η be adaptive, since it does as well depend on the (unknown) scale of the posterior distribution of θ ?
- How would the current theoretical framework cope with these additional adaptations?

Combine adaptations?

Adaptive MCMH looks great, but it adds new tuning parameters compared to standard MCMC!

- It would be nice to be able to adapt the proposal $\mathcal{Q}(\theta_t, \vartheta)$, e.g. to control the acceptance rate, as in other adaptive MCMC algorithms.
- Then could η be adaptive, since it does as well depend on the (unknown) scale of the posterior distribution of θ ?
- How would the current theoretical framework cope with these additional adaptations?
- In the same spirit, could m (and m_0) be considered as algorithmic parameters and be adapted as well? What criterion would be used to adapt them?

Sample size and acceptance rate

- Another MH algorithm where the "true" acceptance ratio is replaced by a ratio with an unbiased estimator is the Particle MCMC algorithm.

Sample size and acceptance rate

- Another MH algorithm where the "true" acceptance ratio is replaced by a ratio with an unbiased estimator is the Particle MCMC algorithm.
- In PMCMC, the acceptance rate is growing with the number of particles used to estimate the likelihood, ie when the estimation is more precise, the acceptance rate increases and converges towards the acceptance rate of an idealized standard MH algorithm.

Sample size and acceptance rate

- Another MH algorithm where the "true" acceptance ratio is replaced by a ratio with an unbiased estimator is the Particle MCMC algorithm.
- In PMCMC, the acceptance rate is growing with the number of particles used to estimate the likelihood, ie when the estimation is more precise, the acceptance rate increases and converges towards the acceptance rate of an idealized standard MH algorithm.
- In adaptive MCMH, is there such a link betweeen m, m_0 and the number of iterations on one side and the acceptance rate of the algorithm on the other side? Should it be growing when the estimation becomes more and more precise?