On stability and convergence of adaptive MC[MC]

Discussion by Christian P. Robert

Université Paris-Dauphine, IuF, and CREST
http://xianblog.wordpress.com

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

The adaptive algorithm

Focussing on the scale matrix of an adaptive random-walk Metropolis Hastings scheme is theoretically fascinating, especially when considering

- removal of containment for LLN [intuitive but hard to prove]
- potential link with renewal theory [in the fixed component version]
- no lower bound on Σ_m
- verifiable assumptions

[Roberts & Rosenthal, 2007]

The adaptive algorithm (2)

Limited applicability of the adaptive scheme when using a random-walk Metropolis Hastings scheme because of

- strong tail [or support] conditions
- more generaly, dependence on parameterisation
- curse of dimensionality [ellip. symm. less & less appropriate]
- unavailability of the performance target [e.g., acceptance of 0.234]

The adaptive algorithm (2)

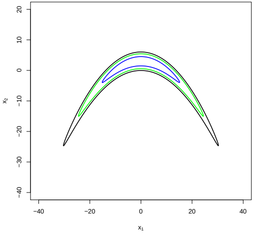
Limited applicability of the adaptive scheme when using a random-walk Metropolis Hastings scheme because of

- strong tail [or support] conditions
- more generaly, dependence on parameterisation
- curse of dimensionality [ellip. symm. less & less appropriate]
- unavailability of the performance target [e.g., acceptance of 0.234]

Congratulations and thanks for adding a software to the theoretical development!

Banana benchmark

Twisted $\mathscr{N}_p(0, \Sigma)$ target with $\Sigma = \mathrm{diag}(\sigma_1^2, 1, \ldots, 1)$, changing the second co-ordinate x_2 to $x_2 + b(x_1^2 - \sigma_1^2)$

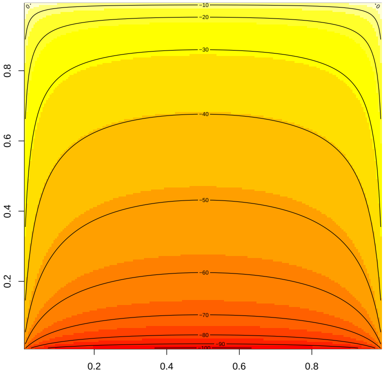


$$p = 10,\ \sigma_1^2 = 100,\ b = 0.03$$

[Haario et al. 1999]

Banana benchmark (2)

Reparameterisation of the above within the unit hypercube by a logit transform



$$p = 2,\ \sigma_1^2 = 100,\ b = 0.01$$

Iterating importance sampling

Population Monte Carlo (PMC) offers a solution to the difficulty of picking the importance function q through adaptivity:
Given a target π, PMC produces a sequence q^t of importance functions $(t = 1, \ldots, T)$ aimed at improving the approximation of π

[Cappé, Douc, Guillin, Marin & X., 2007]

Adaptive importance sampling

Use of mixture densities

$$q^t(x) = q(x; \alpha^t, \theta^t) = \sum_{d=1}^{D} \alpha_d^t \varphi(x; \theta_d^t)$$

[West, 1993]

where

- $\alpha^t = (\alpha_1^t, \ldots, \alpha_D^t)$ is a vector of adaptable weights for the D mixture components
- $\theta^t = (\theta_1^t, \ldots, \theta_D^t)$ is a vector of parameters which specify the components
- φ is a parameterised density (usually taken to be multivariate Gaussian or Student-t)

PMC updates

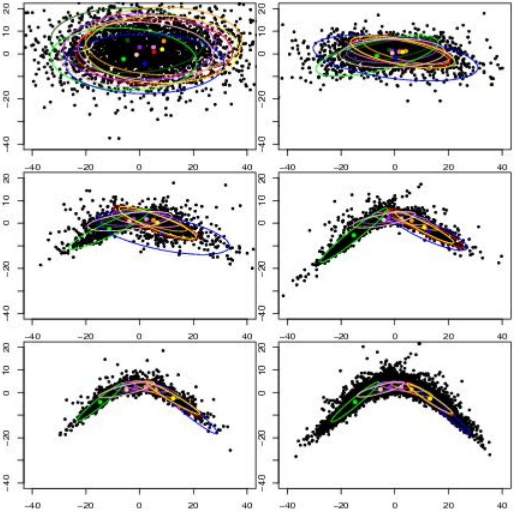
Maximization of $L^t(\alpha, \theta)$ leads to closed form solutions in exponential families (and for the t distributions)
For instance for $\mathscr{N}_p(\mu_d, \Sigma_d)$:

$$\alpha_d^{t+1} = \int \rho_d(x; \alpha^t, \mu^t, \Sigma^t)\pi(x)\mathrm{d}x,$$

$$\mu_d^{t+1} = \frac{\int x\rho_d(x; \alpha^t, \mu^t, \Sigma^t)\pi(x)\mathrm{d}x}{\alpha_d^{t+1}},$$

$$\Sigma_d^{t+1} = \frac{\int (x-\mu_d^{t+1})(x-\mu_d^{t+1})^{\mathrm{T}}\rho_d(x; \alpha^t, \mu^t, \Sigma^t)\pi(x)\mathrm{d}x}{\alpha_d^{t+1}}.$$

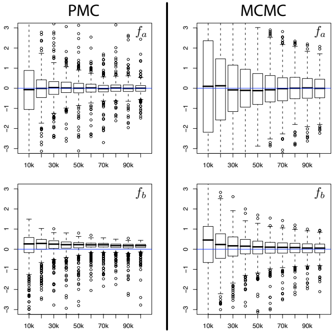
Simula



Comparison to MCMC

Adaptive MCMC: Proposal is a multivariate Gaussian with Σ updated/based on previous values in the chain. Scale and update times chosen for optimal results.

[Kilbinger, Wraith et al., 2010]



Evolution of $\pi(f_a)$ (top panels) and $\pi(f_b)$ (bottom panels) from 10k points to 100k points for both PMC (left panels) and MCMC (right panels).