Controlled MCMC for Optimal Sampling

Christophe ANDRIEU & C.P. ROBERT

Maths. Department, University of Bristol, UK

Ceremade - Université Paris-Dauphine, France.

1.1– Problem and Principle of MCMC algorithms.

- (Typical) **Problem:** efficiently estimate

$$I(f) = \int_{\mathcal{X}} f(x)\,\pi(x)\,dx.$$

- **Solution**: build an ergodic Markov chain (MC) x_i $(i = 0, 1, \ldots)$ with invariant distribution π and compute the estimator

$$\widehat{I}_N(f) = \frac{1}{N+1}\sum_{i=0}^{N} f(x_i).$$

1.2– Metropolis-Hastings is the main building block of MCMC.

- Most, if not all MCMC algorithms, rely on the Metropolis-Hastings (MH) algorithm.
- It requires the definition of a proposal distribution q.
- Given that the MC is at x, y is proposed from the conditional distribution $q(x,y)$. Then, the chain
  1. goes to y with probability

$$\alpha(x,y) = \min\left\{1, \frac{\pi(y)\,q(y,x)}{\pi(x)\,q(x,y)}\right\}.$$

  2. Otherwise stays at x.

1.2– Metropolis-Hastings is the main building block of MCMC.

- As pointed out in Gareth's talk, the choice of q is key to the success of the MCMC approach.
- Examples with Laird's applets (`www.lbreyer.com`).
- Question: does anyone know where Laird is?

2.1– Choice of the proposal distribution.

- Typically q depends on a parameter, say θ. We introduce the notation q_θ to point out this dependence.
- Example: a symetric Gaussian random walk

$$q_\theta(x,y) = \frac{1}{\sqrt{2\pi\theta^2}}\exp\left(\frac{-1}{2\theta^2}(y-x)^2\right).$$

2.2– Problem addressed.

- We want to choose θ in an "**optimal**" manner.
- We want this choice to be automatic, *i.e.* minimise human intervention.
- We want to use a single run of the algorithm, *i.e.* the algorithm will learn from its past history.

2.3– A more general problem.

- **The Truth**: real MCMC algorithms consist of mixtures of MH algorithms:

$$K(x,dy) = \sum_{i=1}^{n}\omega_i K_i(x,dy),$$

where $\sum_{i=1}^n \omega_i = 1$, $\omega_i \geq 0$ and the K_i's are "simple" MH kernels.

- In practice the weights from the mixture and the proposal distribution depend on some parameters θ_i, *i.e.* $\omega_{i,\theta_i}, q_{i,\theta_i}$, and from now on $\theta = \{\theta_i; i = 1, \ldots, n\}$.