Computational Statistics & Machine Learning

Lecture 11

MCMC in Infinite Dimensional Space

Mark Girolami
mag92@cam.ac.uk

Department of Engineering
University of Cambridge

April 26, 2022

Overview

- Revisit Metropolis-Hastings Algorithm
- Define MH in Infinite-Dimensional Space
- Consider correctness of MH in Hilbert space
- Define Markov chain sampler in Hilbert space

Metropolis-Hastings Algorithm

- To simulate from a target distribution with density $\pi(\cdot)$
- **for** $j = 1 \rightarrow N$ **do**
    Simulate z from $q(x^{(j)}, \cdot)$
    Simulate u from $U(0, 1)$
    if $u \leq \alpha(x^{(j)}, z) = min\left(\frac{\pi(z)q(z,x^{(j)})}{\pi(x^{(j)})q(x^{(j)},z)}, 1\right)$ **then**
        Set $x^{(j+1)} = z$
    else
        Set $x^{(j+1)} = x^{(j)}$
    end if
  end for
  Return $\{x^{(1)}, x^{(2)}, x^{(3)}, \cdots, x^{(N)}\}$

Metropolis-Hastings Algorithm

- Consider, for $x \in H$, an infinite dimensional Hilbert space, the proposal $q(x^{(j)}, \cdot) = \mathcal{N}(x^{(j)}, I)$
- There is a problem already as the norm of all $x^{(j)}$ is going to be unbounded and hence ill defined
- The random variables $x_n \sim \mathcal{N}(0, 1)$ will almost surely diverge
- This problem can be resolved by employing a trace class covariance operator C so that $q(x^{(j)}, \cdot) = \mathcal{N}(x^{(j)}, C)$
- Now the proposal is well defined and random draws will have finite norm and converge
- What about the acceptance probability?

Defining MH in Infinite-Dimensional Space

- Andrew Stuart FRS - Caltech

Metropolis-Hastings Algorithm

- Consider the definition of the acceptance probability

$$\alpha(x^{(j)}, z) = min\left(\frac{\pi(z)q(z, x^{(j)})}{\pi(x^{(j)})q(x^{(j)}, z)}, 1\right)$$

- Recall that $\pi(z)$ is our target density, which would be defined by a Radon-Nikodym derivative with respect to Lebesgue Measure
- We have already seen that Lebesgue measure is undefined in an infinite dimensional space
- Bayes formula has now been redefined as a Radon-Nikodym derivative with respect to Gaussian measure

$$\frac{d\mu^y}{d\mu^0}(x) \propto \mathbb{L}(y|x)$$

Metropolis-Hastings Algorithm

- Bayes formula has now been redefined as a Radon-Nikodym derivative with respect to Gaussian measure

$$\frac{d\mu^y}{d\mu^0}(x) \propto \mathbb{L}(y|x)$$

- The acceptance probability can then be defined in terms of the appropriate measures replacing $\pi(z)q(z, x^{(j)})$ with $\mu^y(dz)q(dz, x^{(j)})$
- Noting that for the case $\mu^0 = \mathcal{N}(0, C)$ and the proposal is $v = u + \beta\xi$ where $\xi \sim \mathcal{N}(0, C)$ show that, if $\mathbb{L}(y|v) = \exp(-\Phi(v))$ the acceptance probability follows as $\min\left(\exp(J(v) - J(u)), 1\right)$ where

$$\begin{aligned} J(v) &= \log\left(\mathbb{L}(y|v)\right) - \frac{1}{2}(v, C^{-1}v) \\ &= -\Phi(v) - \frac{1}{2}|C^{-\frac{1}{2}}v|^2 \end{aligned}$$

Metropolis-Hastings Algorithm

- With an acceptance probability $\min(\exp(J(v) - J(u)), 1)$ where $J(v) = -\Phi(v) - \frac{1}{2}|C^{-\frac{1}{2}}v|^2$ is this well defined in an infinite dimensional space?
- The question to be asked is whether $\mu^y(dv)q(v, du)$ is absolutely continuous with respect to $\mu^y(du)q(u, dv)$ what does it suggest for the algorithm if they are not?
- In fact the ratio is not well defined in an infinite dimensional space (for any fixed β)
- The i'th coordinate of $C^{-\frac{1}{2}}u$ is $\mathcal{N}(0, 1)$
- The i'th eigenvalue of $C^{-\frac{1}{2}}$ is $\frac{1}{\lambda_i}$ also $u_i \sim \mathcal{N}(0, \lambda_i^2)$ so $\lambda_i^{-1}u_i \sim \mathcal{N}(0, \lambda_i^{-2}\lambda_i^2) = \mathcal{N}(0, 1)$
- The expectation of the norm will diverge as $\mathbb{E}|C^{-\frac{1}{2}}u|^2 = \mathbb{E}\sum_i(\lambda_i^{-1}u_i)^2 = \sum_i = \infty$
- This is not good....

Metropolis-Hastings Algorithm

- Change the proposal from $v = u + \beta\xi$ with $\xi \sim \mathcal{N}(0, C)$ and β a scaling to

$$v = \sqrt{1-\beta^2}u + \beta\xi$$

  with $\xi \sim \mathcal{N}(0, C)$
- If $u \sim \mathcal{N}(0, C)$ then the proposal also is $v \sim \mathcal{N}(0, C)$ preserving the prior measure unlike the previous proposal a seemingly innocuous change
- Show that this proposal yields an acceptance probability of the form

$$\min\left(\exp(\Phi(u) - \Phi(v)), 1\right)$$

- This acceptance probability is well defined in an infinite dimensional space, provided $\Phi(\cdot)$ is well defined
- We now have an MCMC method that will not decay in performance as dimensionality increases to the infinite dimensional

Metropolis-Hastings Algorithm

- This is of course not pulled out of a bag like some magicians sleight of hand
- Consider the following stochastic differential equation $du = -uds + \sqrt{2C}db$ with b standard brownian motion
- It can be shown that this will have $\mu^0 = \mathcal{N}(0, C)$ as an invariant measure
- By discretising the SDE will produce a discrete time Markov chain that can be used in MCMC
- The Theta discretisation scheme yields

$$v - u = -\delta((1-\theta)u + \theta v) + \sqrt{2C\delta}\xi_0$$

- With u the current position, v the next position, δ the discrete time difference and ξ_0 is a standard Normal

Metropolis-Hastings Algorithm

- Rearranging

$$v - u = -\delta((1-\theta)u + \theta v) + \sqrt{2C\delta}\xi_0$$

gives

$$v = (I + \delta\theta)^{-1}\left(u - \delta(1-\theta)u + \sqrt{2\delta C}\xi_0\right)$$

- By setting $\theta = \frac{1}{2}$ we obtain

$$v = \left(1 + \frac{\delta}{2}\right)^{-1}\left(\frac{2-\delta}{2}u + \sqrt{2\delta C}\xi_0\right)$$

Set $\beta = \sqrt{2\delta}(1 + \frac{\delta}{2})^{-1}$ and finally

$$v = \sqrt{1-\beta^2}u + \beta\xi$$

with $\xi \sim \mathcal{N}(0, C)$, and $u = u_0 \sim \mathcal{N}(0, C)$

Metropolis-Hastings Algorithm

- Now the question is whether this ensures that the two measures defining the acceptance probability are equivalent
- It is the case the measures are equivalent only for the case where $\theta = \frac{1}{2}$
- The way to prove this is to look at each element of v i.e. v_i
- We have

$$v = (I + \delta\theta)^{-1} \left(u - \delta(1-\theta)u + \sqrt{2\delta C}\xi_0 \right)$$

- and

$$v_i = \frac{1 - \delta(1-\theta)}{(I + \delta\theta)} u_i + \frac{\sqrt{2\delta\lambda_i^2}}{(I + \delta\theta)} g_i$$

with g_i standard Normal

Metropolis-Hastings Algorithm

- It is clear that $v_i = \frac{1-\delta(1-\theta)}{(I+\delta\theta)} u_i + \frac{\sqrt{2\delta\lambda_i^2}}{(I+\delta\theta)} g_i$ has zero mean
- The variance follows straighforwardly as

$$\left(\frac{1-\delta(1-\theta)}{(I+\delta\theta)}\right)^2 \lambda_i^2 + \frac{2\delta\lambda_i^2}{(I+\delta\theta)^2}$$

- We are interested in the ratio of the variance of v_i and u_i as the convergence criterion

$$\sum_{i=1}^{\infty}\left(\frac{\sigma(v_i)}{\sigma(u_i)} - 1\right)^2 < \infty$$

needs to be satisfied for two product Gaussian measures to be equivalent

Metropolis-Hastings Algorithm

- The required ratio, remembering that variance of u_i is λ_i, follows as

$$\left(\frac{1-\delta(1-\theta)}{(I+\delta\theta)}\right)^2 + \frac{2\delta}{(I+\delta\theta)^2}$$

- This needs to be unity for the inequality to be satisfied and so solving $1 + 2\theta\delta + \delta^2(1-\theta)^2 = (1+\delta\theta)^2$ for θ
- $\theta = \frac{1}{2}$
- We have shown that the proposal mechanism yields a well defined acceptqnce probability in infinite dimensional space and the method will not suffer from degeneracy in high dimensions as it is properly defined in infinite dimensions