EL2320 - MCMC

John Folkesson

Markov Chain Monte Carlo
(Advanced Topic)

An Introduction to MCMC ... Andrieu et al.

MCMC are a class of algorithms that have been shown to solve otherwise intractable problems (given enough time)

PF is an example.

Metroplis-Hastings algorithm is the most important.

Monte Carlo Principle

Monte Carlo methods share the representation of the target distribution as

$$p(\mathbf{x}) = \frac{1}{M}\sum_{i=1}^{M}\delta(\mathbf{x} - \mathbf{x}^{(i)}) \tag{1}$$

It replaces integrals with sums to compute expectation values

$$E[f(\mathbf{x})]_p = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x} \approx \frac{1}{M}\sum_{i=1}^{M} f(\mathbf{x}^{(i)}) \tag{2}$$

Generating the Samples = Estimation

The general idea is to generate samples from a target distribution we can not sample $p(\mathbf{x})$ but we can evaluate $\tilde{p}(\mathbf{x})$ at any given state:

$$p(\mathbf{x}) = \frac{\tilde{p}(\mathbf{x})}{\int \tilde{p}(\mathbf{x})dx} \tag{3}$$

Method of the PF says use a two sample method where we first sample from a proposal distribution $\mathbf{x}^{(i)} \sim q(\mathbf{x})$ then resample with the importance weights.

$$w^{(i)} \propto \frac{\tilde{p}(\mathbf{x}^{(i)})}{q(\mathbf{x}^{(i)})} \tag{4}$$

As we gave several examples of we can chose different proposal distributions to improve our estimates.

One Simple Method - Rejection Sampling

Find an easy to sample distribution and a number K that satisfy:

$$\tilde{p}(\mathbf{x}) < Kq(\mathbf{x}) \tag{5}$$

Sample:

$$x^{(i)} \sim q(\mathbf{x}) \tag{6}$$

$$u \in [0, 1] \tag{7}$$

Where u is uniformly distributed. Then if

$$u < \frac{\tilde{p}(\mathbf{x}^{(i)})}{Kq(\mathbf{x}^{(i)})} \tag{8}$$

We 'accept' $\mathbf{x}^{(i)}$ else try again.

One Simple Method - Rejection Sampling

We see that the two independent samples of u and $\mathbf{x}^{(i)}$) mean that the effective a sampling is the product of the two sample distibutions

$$\mathbf{x}^{(i)} \sim \frac{\tilde{p}(\mathbf{x}^{(i)})}{Kq(\mathbf{x}^{(i)})} q(\mathbf{x}^{(i)}) = \frac{\tilde{p}(\mathbf{x}^{(i)})}{K} \tag{9}$$

Problem if K is very large then too many samples get rejected and this takes too long. Typical issue for high dimensional spaces.

Define a stocastic process with the Markov property:

$$p(\mathbf{x}^{(i)}|\mathbf{x}^{(i-1)}, ..., \mathbf{x}^{(1)}) = T(\mathbf{x}^{(i)}|\mathbf{x}^{(i-1)}) \tag{10}$$

We also want T to not depend on i. Then if the T have three properties we can sample them to generate a sequence of states that 'converge' to samples from our target distribution.

- Irreducibility - There is a positive probability of visiting any state starting at any state.
- Aperiodicity - The chain can not get traped in cycles, (even ones that visit every state).
- Detailed balance - $T(\mathbf{x}^{(i-1)}|\mathbf{x}^{(i)})p(\mathbf{x}^{(i)}) = T(\mathbf{x}^{(i)}|\mathbf{x}^{(i-1)})p(\mathbf{x}^{(i-1)})$.

$$p(\mathbf{x}^{(i)}) = \sum_{\mathbf{x}^{(i-1)}} T(\mathbf{x}^{(i)}|\mathbf{x}^{(i-1)})p(\mathbf{x}^{(i-1)}) \tag{11}$$