MVE550 2019 Lecture 9

Petter Mostad

Chalmers University

December 4, 2019

Overview

- Review of the Metropolis Hastings algorithm.
- Example.
- Gibbs sampling in the Ising model.
- Perfect sampling.
- Total Variation Distance and card shuffling.

The Metropolis Hastings algorithm

- Assume a density (or probability mass function) $\pi(\theta)$ is provided.
- We also assume given a *proposal function* $q(\theta_{new} \mid \theta)$, which, for every given θ, provides a probability distribution (or probability mass function) for a new θ_{new}.
- Finally, define, for θ and θ_{new}, the acceptance probability

$$a = \min\left(1, \frac{\pi(\theta_{new})q(\theta \mid \theta_{new})}{\pi(\theta)q(\theta_{new} \mid \theta)}\right)$$

- The Metropolis Hastings algorithm is: Starting with some initial value θ_0, generate $\theta_1, \theta_2, \ldots$ by, at each step, proposing a new θ based on the old using the proposal function and accepting it with probability a. If it is not accepted, the old value is used again.
- If this defines an ergodic Markov chain, its unique stationary distribution is $\pi(\theta)$.

Example

- Assume that a model has the real parameter θ, and that the posterior for θ has been found to be

$$\pi(\theta \mid \text{data}) = 0.3\,\text{Normal}(\theta; 2, 0.5^2) + 0.7\,\text{Normal}(\theta; 6, 1^2).$$

  As a toy example, compare a sample simulated directly from this distribution to one simulated using Metropolis Hastings. Use as starting value 1 and proposal function $\pi(\theta' \mid \theta) = \text{Uniform}(\theta'; \theta - 0.5, \theta + 0.5)$.
- Assume we would like find the predictive distribution for y when $y \mid \theta \sim \text{Normal}(\theta, 0.3^2)$ and θ has the distribution above.
  - Do this first by using a sample from generated by Metropolis Hastings.
  - Then, compute and compare to the theoretical distribution.

Gibbs sampling

- A version of Metropolis Hastings with a special type of proposal functions: For each component of $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$, use the conditional distribution where all but one of the components are fixed.
- It is straightforward to show that the acceptance probability becomes 1.
- When conditional distributions are easy to derive, this is a popular choice for proposal functions.
- Convergence is not always fast.

The Ising model

- The *configurations* σ consist of nodes in a grid, where in each node v the configuration has value $\sigma_v = 1$ or $\sigma_v = -1$.
- The energy of a configuration is defined as
$$E(\sigma) = -\sum_{v \sim w} \sigma_v \sigma_w$$
where the sum is over all *neighbour* pairs v and w.
- The Gibbs distribution on the set of all configurations has probaility mass function
$$\pi(\sigma) = \frac{\exp\left(-\beta E(\sigma)\right)}{\sum_\tau \exp\left(-\beta E(\tau)\right)}$$
where β is a real parameter.
- Gibbs sampling works well as a simulation method for the Gibbs distribution.
- (One can observe a “phase transition” at a particular value of β.)

Perfect sampling

Given ergodic Markov chain with finite sample space of size k and limiting distribution π.

- Idea: Given n, prove that X_n actually has the limit distribution.
- Method: Prove that the distribution at X_n is independent of the starting value at X_0.
- How: Construct k Markov chains that are dependent ("coupled") but which are marginally Markov chains as above. If they all start at the k possible values at X_0 but have identical values at X_n, we are done.
- Note: *n cannot* be determined as the first value where the k chains meet; it must be determined beforehand!
- Thus usually one wants to generate a chain $X_{-n}, X_{-n+1}, \ldots, X_0$ where X_0 has the limiting distribution, and we stepwise increase n to make all chains *coalesce* to one chain.

Using same source of randomness for all chains

Consider the chains $X_{-n}^{(j)}, \ldots, X_0^{(j)}$ for $j = 1, \ldots, k$.

- Instead of simulating $X_{i+1}^{(j)}$ based on $X_i^{(j)}$ independently for each j, we define a function g so that $X_{i+1}^{(j)} = g(X_i^{(j)}, U_i)$ for all j, where $U_i \sim \text{Uniform}(0, 1)$.
- Thus if two chains have identical values in X_i, they will also be identical at X_{i+1}.
- See Figure 5.10 in Dobrow.