STA 4103/5107 Computational Methods in Statistics II

Department of Statistics
Florida State University

Class 9
February 7, 2017

Review: Metropolis-Hastings Algorithm

- **Goal: generating samples of a random variable X distributed according to the density $f(x)$.**

- In addition to $f(x)$, we will assume having a density $q(y|x)$ that satisfies the following properties:

  1. It is easy to sample from $q(\cdot|x)$ for all x.

  2. The support of q contains the support of f.

  3. The functional form of $q(y|x)$ is known or $q(y|x)$ is symmetric in y and x. It is not necessary to know the normalizing constant in $q(y|x)$ as long as it does not depend upon x.

Review: M-H Algorithm

- **Algorithm 34 (Metropolis-Hastings Algorithm)** Given $f(x)$ and a choice of $q(y|x)$ that satisfies the above mentioned properties, we choose an initial condition X_0 in the support of $f(x)$. The Markov chain $X_1, X_2, \ldots, X_n$ is constructed iteratively according to the steps:

  1. Generate a candidate $Y \sim q(y|X_t)$.
  2. Update the state to X_{t+1} according to:

$$X_{t+1} = \begin{cases} Y & \text{with probability } \rho(X_t, Y) \\ X_t & \text{with probability } 1 - \rho(X_t, Y) \end{cases}$$

  where $\rho(x, y) = \min\{[f(y)q(x \mid y)]/[f(x)q(y \mid x)], 1\}$.

  $q(y|x)$ is called the proposal density and $\rho(x, y)$ is called the acceptance-rejection function.

7.4 Gibbs Sampler

Gibbs Sampler

- Gibbs sampler is another commonly used tool for generating Markov chains with suitable asymptotic properties.

- By construction, Gibbs sampler applies only to the problem of sampling from multivariate densities.

- Let $X = (X_1, X_2, \ldots, X_p) \in \boldsymbol{R}^p$ be a vector of random variables with the joint density function given by $f(x_1, x_2, \ldots, x_p)$.

- Our goal is to generate samples from f and we will do so by constructing a Markov chain on $\boldsymbol{R}^p$.

Gibbs Sampler Algorithm

- *Assumption*: we know the conditional densities
$$f_i(x_i \mid x_j, j \neq i), i = 1, \ldots, p,$$
and have method(s) to sample from each of these. These conditional densities are called the *full conditionals*.

- **Algorithm 36 (Gibbs Sampler)** Let $X^{(t)} = [X_1^{(t)} \ \cdots \ X_p^{(t)}] \in R^p$ be the value of Markov chain at time t. Following steps describe an update from $X(t)$ to $X(t+1)$.

  1. Generate $\quad X_1^{(t+1)} \sim f_1(x_1 \mid X_2^{(t)}, X_3^{(t)}, \ \cdots \ X_p^{(t)})$

  2. Generate $\quad X_2^{(t+1)} \sim f_2(x_2 \mid X_1^{(t+1)}, X_3^{(t)}, \ \cdots \ X_p^{(t)})$

  ... ...

  p. Generate $\quad X_p^{(t+1)} \sim f_p(x_p \mid X_1^{(t+1)}, X_2^{(t+1)}, \ \cdots \ X_{p-1}^{(t+1)})$

Properties

- Gibbs sampler by construction applies only to multivariate densities.
- An important property of the Gibbs sampler is that even for large values of p, one samples from a univariate density at each step.
- This makes Gibbs sampler very attractive for large dimensional problems such as image analysis.
- A Gibbs sampler can be considered a special case of the M-H algorithm with the proposal density given by the full conditionals.
- Also, in Gibbs sampler one always accepts the proposed state as opposed to the acceptance/rejection of the M-H algorithm.

Bivariate Gibbs sampler

- A specific case for $p = 2$ is the bivariate Gibbs sampler.
- Let X and Y be two scalar random variables with the joint density function $f(x, y)$ and the full conditionals: $f_1(x|y)$ and $f_2(y|x)$.
- Gibbs sampler can be constructed as follows:
- Start with some initial condition (x_0, y_0) and iterate according to:
- **Algorithm 37 (Bivariate Gibbs Sampler)**

  1. Generate $x_{t+1} \sim f_1(x|y_t)$.
  2. Generate $y_{t+1} \sim f_2(y|x_{t+1})$.
  3. Set $t = t + 1$ and go to Step 1.

Normal Density

- To illustrate this special case, consider the bivariate normal density:

$$(X,Y) \sim N(0,\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}).$$

The full conditionals:

$$f_1(x|y) = N(\rho y,\ 1-\rho^2),\ f_2(y|x) = N(\rho x,\ 1-\rho^2)\ .$$

Markov Random Fields

- An important applications of Gibbs sampler is in generating samples from **Markov Random Field (MRF) models**.
- Let S be a collection of indexed locations, e.g. indices on a uniform lattice, in a plane.
- For each site $s \in S$, we assign a random variable $X(s) \in \mathbf{R}$, and the collection $X \equiv \{X(s), s \in S\}$ as a field of random variables.
- Elements of S are often called the pixel locations and the random variable $X(s)$ is called the pixel value at s.
- Before we introduce specific models, we introduce some additional notation.

Notation

- We will consider the sites geographically close to $s \in S$ as neighbors of s, denoted by $N_s \subset S$.
- For example, in a uniform, square lattice, we can choose the sites immediately north, south, east and west to s to form a neighborhood of s.
- Alternatively, one can include the diagonal neighbors also.
- Depending on the application, different neighborhoods can be chosen for a random field. With this notation, a MRF is defined as follows.

Definition

- **Definition 28 (Markov Random Field)** If the conditional probability density of a random variable $X(s)$, given the remaining variables $\{X(r), r \neq s\}$, is the same as the conditional probability density of $X(s)$ given only its neighbors

$$\{X(s)|X(r), r \in N_s\},$$

for all s, then X is called a Markov random field. That is, for all $s \in S$,

$$f(X(s)|X(r), r \neq s) = f(X(s)|X(r), r \in N_s) .$$

- A simple example of MRF is the **Ising model** where the pixel values $X(s)$ are allowed only two values 1 or -1.

Ising Model

- **Example 8 (Ising Model)** For constants H and J, the joint probability distribution of all pixel values in an Ising model is given by: for $X \equiv \{X(s) \in \{-1, 1\}, s \in S\}$,

$$P(X) = \frac{1}{Z} e^{H\sum_s X(s) + \frac{J}{2}\sum_s X(s)(\sum_{r \in N_s} X(r))}$$

where Z is the normalizing constant.

The conditional probability distribution is given by:

$$P(X(s) \mid X(r), r \in N_s) = e^{(H+J(\sum_{r \in N_s} X(r)))(1+X(s))} / [1 + e^{2(H+J(\sum_{r \in N_s} X(r)))}].$$

- The Gibbs sampler algorithm for sampling from this Ising model is: for any fixed ordering of indices, generate a sample from $P(X(s)|X(r), r \in N_s)$ and use it to update $X(s)$.