SI 536 Analysis of Multi-type and Big data

Course Project

Guided by: Prof. Sabnis

Submitted by:

Kushal Padole 205280036

Anamika Mishra 205280019

Priya Kandare 205280001

Gibbs sampler

1 Introduction

The Gibbs sampler is technique for generating random variables from a (marginal) distribution indirectly, without having to calculate the density. Through the use of techniques like Gibbs sampler, we are able to avoid difficult calculations, replacing them instead with sequence of easier calculations.

Suppose we are given a join density $f(x_1, x_2, \dots, x_k)$, and we are interested in obtaining the characteristics of the marginal density (such as mean or variance). However, there are many cases where the computation to get marginal density are extremely difficult to perform, either analytically or numerically. In such cases the Gibbs sampler provides an alternative method for obtaining sample from marginal densities.

A basic Gibbs sampling algorithm is shown below:

a. Assign a vector of starting values $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_k^{(0)})$.
b. Set counter index $i = 0$.
c. We do the sampling as follows
   1. Sample $x_1^{(i+1)}$ from $f(x_1 | x_2^{(i)}, x_3^{(i)}, \dots, x_k^{(i)})$.
   2. Sample $x_2^{(i+1)}$ from $f(x_2 | x_1^{(i+1)}, x_3^{(i)}, \dots, x_k^{(i)})$.

   …

   k. Sample $x_k^{(i+1)}$ from $f(x_k | x_1^{(i+1)}, x_3^{(i+1)}, \dots, x_{k-1}^{(i+1)})$.
d. Form $x^{(i+1)} = (x_1^{(i+1)},\ x_2^{(i+1)},\ \dots, x_k^{(i+1)})$.
e. Set $i = i + 1$ and return to step c.

In words, the Gibbs sampling algorithm defined above generates a random draw from each of the full conditional distributions. However, as the algorithm progresses, the value of the conditioning variable are sequentially updated. Thus, the next draw depends on the previous one. It turns out that under reasonably general conditions, for k large enough, $x_i^{(k)}$ can be thought of as effectively a sample point from marginal density $f_i(x)$. Moreover, a key aspect of Gibbs sampling is that the full expression of conditional distribution does not need to be known; it only needs to be known up to a normalizing constant.

2 Demonstrating the Properties of Gibbs Sampling

Suppose we are given a joint density of (X, λ) as,

$$f(x,\lambda) \propto \lambda^{\alpha} e^{-\lambda x} e^{-\lambda\beta}, \quad x > 0, \lambda > 0$$

Where α, β are known constants.

Suppose we are interested in calculating some characteristics of marginal distribution $f(x)$ of X. The Gibbs sampler allows us to generate a sample from this marginal as follows:

It can be seen from the form of joint density that,

$$f(x|\lambda) \sim exp(\lambda)$$

$$f(\lambda|x) \sim gamma(\alpha + 1, x + \beta)$$

We run the following Gibbs sampler algorithm,

1. Select arbitrary initial values $x^{(0)}$ and $\lambda^{(0)}$.
2. Set counter index $i = 0$.
3. Sample $x^{(i+1)}$ from $exponential(\lambda^{(i)})$.
4. Sample $\lambda^{(i+1)}$ from $gamma(\alpha + 1,\ x^{(i+1)} + \beta)$.
5. Set $i = i + 1$ and return to step 3.

For illustrative purposes, assume $\alpha = 5$ and $\beta = 100$. Using the principles of Gibbs sampling as shown above, 10,000 random numbers are generated for X and λ. Figure 1 shows the last 100 sampled valued for X and Figure 2 shows the last 100 sampled values for λ.

Figure 1.

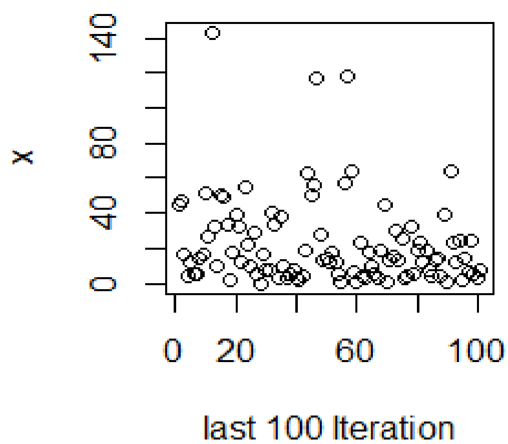
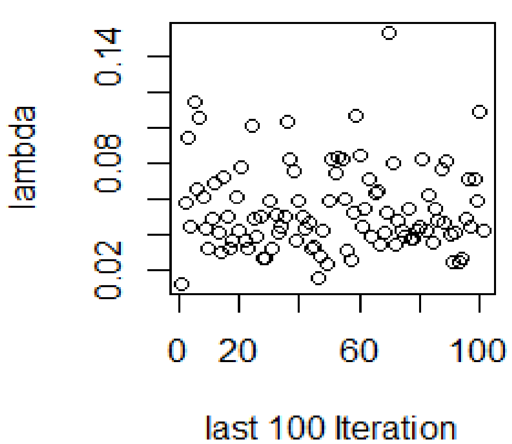
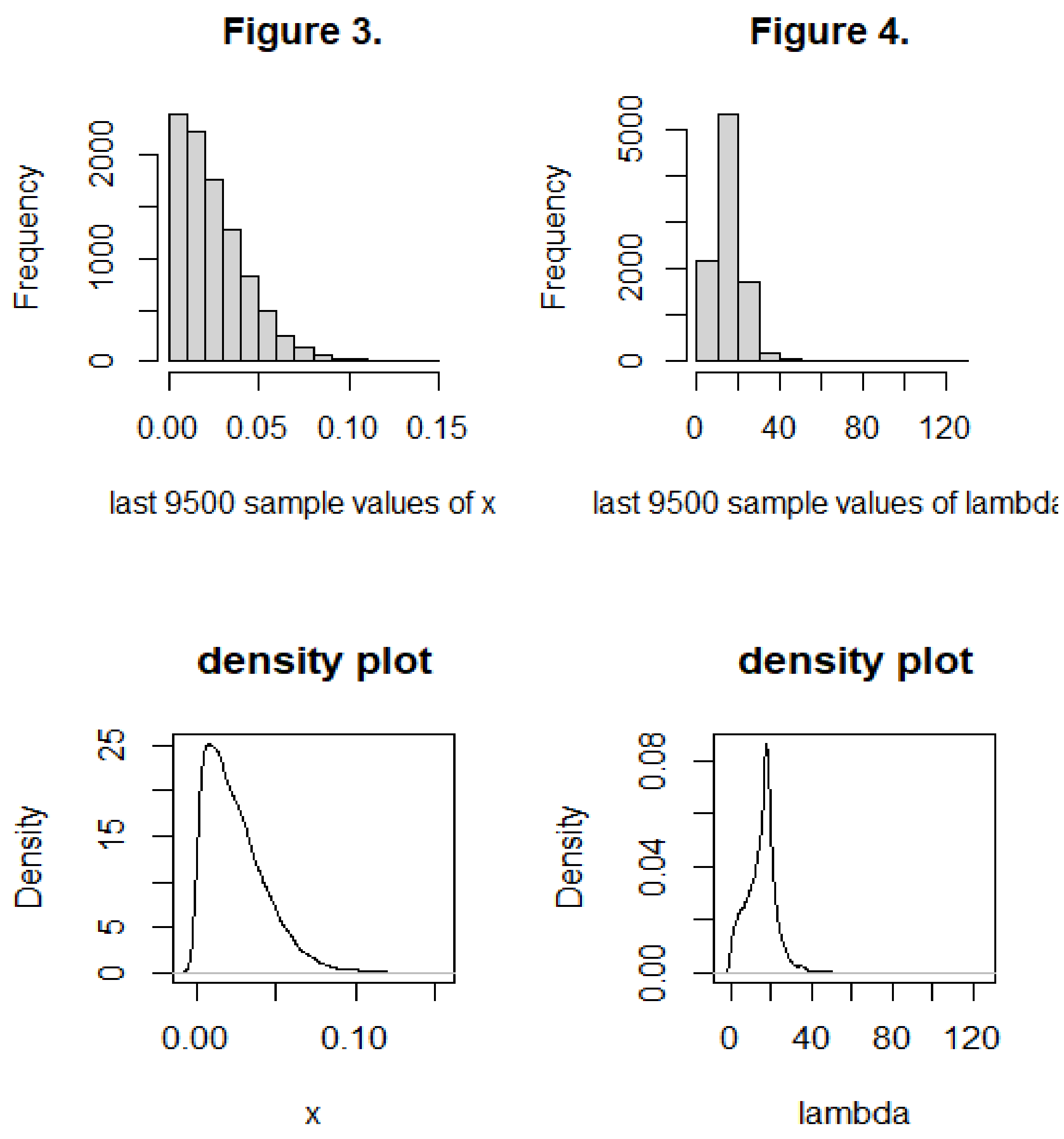


Figure 2.



As one can see from Figure 1 and Figure 2, there is no pattern among the generated random numbers. Therefore, they can be considered as independent random samples.

Next, below figure shows the histogram and density of last 9500 sampled values of X and λ. The first 500 values of each sequence are discarded as these are considered to be burn-in iterations.



These values especially at tail ends, closely resembles the respective marginal densities which can be obtained through direct computation.

In above illustration, we can get the marginal distribution by direct computation,

$$f(x) = \int_0^\infty f(x,\lambda)\, d\lambda$$

$$f(x) = \frac{\alpha\beta^\alpha}{(\beta + x)^{\alpha+1}}.\ x > 0$$

Which is, $Pareto(\alpha, \beta)$ density function. Similarly, we can find the density function for λ.

In effect, by taking very large random samples from the conditional posterior distributions, it appears as if the samples were taken from their respective marginal distributions.

Thus, the generated random variates can be used to study the properties of the distribution of interest. With more complicated models, sampling from the marginal distributions directly would have been impossible; but with Gibbs sampling, it can be simulated. We can formally state Gibbs sampling as:

The realization that as the number of iterations approaches infinity, the samples from the conditional posterior distributions converge to what the actual target distribution is that could not be sampled from directly.

In practical situation, the above computation for marginal density is not possible, in which case, Gibbs sampling can be used to get the sample from marginal density.

Next, consider the following,

$$f(x) = \int f(x|\lambda) f(\lambda) d\lambda$$

This states that the marginal distribution of X can be interpreted as the average of the conditional distribution of X given λ taken with respect to marginal distribution of λ.This fact is used to estimate actual $f(x)$ at point x.

$$f(x) \approx \frac{1}{n} \sum_{i=k}^{n} f(x|\lambda_i)$$

Concluding remarks:

1. The Gibbs sampler generates a markov chain of random variables which converge to the distrubtion of interest $f(x)$. Many of the popular approaches to extraxting information from the Gibbs sequence exploit this property by selecting some large value for k, after which we treat sample from conditional posterior density $f(x|\lambda^{(k)})$ same as that of sample from $f(x)$.
2. General startegy for choosing k is to monitor the convergence of some aspect of Gibbs sequence. But there is no foolproof method for choosing k.