Module 7: Introduction to Gibbs Sampling

Rebecca C. Steorts

Agenda

In this lab, we will deriving conditional distributions, code a Gibbs sampler, and analyze the output of the Gibbs sampler.

Problem Statement

Consider the following Exponential model for observation(s) $\mathbf{x} = (\mathbf{x_1}, \ldots, \mathbf{x_n})$.[1]:

$$p(x|a, b) = ab \exp(-abx) I(x > 0),$$

where the x_i are assumed to be iid for $i = 1, \ldots n$. and suppose the prior is

$$p(a, b) = \exp(-a - b) I(a, b > 0).$$

You want to sample from the posterior $p(a, b|x_{1:n})$. You may assume that

$$a = 0.25, b = 0.25$$

when coding up your Gibbs sampler.

[1] Please note that in the attached data there are 40 observations, which can be found in data-exponential.csv.

Tasks

1. Find the conditional distributions needed for implementing a Gibbs sampler.
2. Code up your own Gibbs sampler using part (1).
3. Run the Gibbs sampler, providing convergence diagnostics.
4. Plot a histogram or a density estimate of the estimated posterior using (2) and (3).
5. How do you know that your estimated posterior in (3) is reliable?

Task 1:

Consider the following Exponential model for observation(s) $x = (x_1, \ldots, x_n)$.[2]:

$$p(x|a, b) = ab \exp(-abx) I(x > 0)$$

and suppose the prior is

$$p(a, b) = \exp(-a - b) I(a, b > 0).$$

You want to sample from the posterior $p(a, b|x)$.

[2] Please note that in the attached data there are 40 observations, which can be found in data-exponential.csv.

Task 1: Conditional distributions

$$
\begin{aligned}
p(\boldsymbol{x}|a, b) &= \prod_{i=1}^{n} p(x_i|a, b) \\
&= \prod_{i=1}^{n} ab \exp(-abx_i) \\
&= (ab)^n \exp\left(-ab \sum_{i=1}^{n} x_i\right).
\end{aligned}
$$

The function is symmetric for a and b, so we only need to derive $p(a|\boldsymbol{x}, b)$.

Task 1: Conditional distributions

This conditional distribution satisfies

$$\begin{aligned}
p(a|\boldsymbol{x}, b) &\propto_a p(a, b, \boldsymbol{x}) \\
&= p(\boldsymbol{x}|a, b)p(a, b) \\
&= (ab)^n \exp\left(-ab\sum_{i=1}^{n} x_i\right) \times \exp(-a-b)I(a, b > 0) \\
&\underset{a}{\propto} {\color{red}a^n} \exp(-abn\bar{x} - a)\mathbb{1}(a > 0) \\
&= {\color{red}a^{n+1-1}} \exp(-(bn\bar{x} + 1)a)\mathbb{1}(a > 0) \\
&\underset{a}{\propto} \text{Gamma}(a \mid n+1,\ bn\bar{x} + 1).
\end{aligned}$$

Therefore, $p(a|b, x) = \text{Gamma}(a \mid n+1,\ bn\bar{x} + 1)$ and by symmetry, $p(b|a, x) = \text{Gamma}(b \mid n+1,\ an\bar{x} + 1)$.

Task 2: Gibbs sampling code

```
knitr::opts_chunk$set(cache=TRUE)
library(MASS)
data <- read.csv("data-exponential.csv", header = FALSE)
```

Task 2: Gibbs sampling code

```
########################################
# This function is a Gibbs sampler
#
# Args
#   start.a: initial value for a
#   start.b: initial value for b
#   n.sims: number of iterations to run
#   data:  observed data, should be in a
            # data frame with one column
#
# Returns:
#   A two column matrix with samples
     #   for a in first column and
# samples for b in second column
########################################
```

Task 2: Gibbs sampling code

```r
sampleGibbs <- function(start.a, start.b, n.sims, data){
  # get sum, which is sufficient statistic
  x <- sum(data)
  # get n
  n <- nrow(data)
  # create empty matrix, allocate memory for efficiency
  res <- matrix(NA, nrow = n.sims, ncol = 2)
  res[1,] <- c(start.a,start.b)
  for (i in 2:n.sims){
    # sample the values
    res[i,1] <- rgamma(1, shape = n+1,
                       rate = res[i-1,2]*x+1)
    res[i,2] <- rgamma(1, shape = n+1,
                       rate = res[i,1]*x+1)
  }
  return(res)
}
```

Task 3: Run the Gibbs sampler

```
# run Gibbs sampler
n.sims <- 10000
# return the result (res)
res <- sampleGibbs(.25,.25,n.sims,data)
head(res)
```

```
##          [,1]      [,2]
## [1,] 0.250000 0.2500000
## [2,] 1.949952 0.2783548
## [3,] 1.687067 0.2665395
## [4,] 1.640766 0.3148827
## [5,] 1.776835 0.2938712
## [6,] 1.039362 0.4910279
```

Task 4

Plot a histogram or a density estimate of the estimated posterior using tasks (2) and (3).

Finish this for homework.

Task 5

How do you know that your estimated posterior in task (3) is reliable?

Finish for homework.