The Multi Stage Gibbs Sampling

Rebecca C. Steorts
Bayesian Methods and Modern Statistics: STA 360/601

Module 8

The generalization to more than two variables is straightforward.

We cycle through the variables, sampling each from its conditional distributional given all the rest.

Three Stage GS

Assume three random variables, with joint pmf or pdf: $p(x, y, z)$..

Set x, y, and z to some values (x_o, y_o, z_o).

Sample $x|y, z$, then $y|x, z$, then $z|x, y$, then $x|y, z$, and so on.
More precisely,

0. Set (x_0, y_0, z_0) to some starting value.
1. Sample $x_1 \sim p(x|y_0, z_0)$.
   Sample $y_1 \sim p(y|x_1, z_0)$.
   Sample $z_1 \sim p(z|x_1, y_1)$.
2. Sample $x_2 \sim p(x|y_1, z_1)$.
   Sample $y_2 \sim p(y|x_2, z_1)$.
   Sample $z_2 \sim p(z|x_2, y_2)$.

   $\vdots$

Multistage GS

Assume d random variables, with joint pmf or pdf $p(v^1, \ldots, v^d)$.

At each iteration $(1, \ldots, M)$ of the algorithm, we sample from

$$\begin{gathered} v^1 \mid v^2, v^3, \ldots, v^d \\ v^2 \mid v^1, v^3, \ldots, v^d \\ \vdots \\ v^d \mid v^1, v^2, \ldots, v^{d-1} \end{gathered}$$

always using the most recent values of all the other variables.

The conditional distribution of a variable given all of the others is referred to as the *full conditional* in this context, and for brevity denoted $v^i | \cdots$.

Example: Censored data

In many real-world data sets, some of the data is either missing altogether or is partially obscured.

One way in which data can be partially obscured is by *censoring*, which means that we know a data point lies in some particular interval, but we don't get to observe it exactly.

Medical data censoring

6 patients participate in a cancer trial, however, patients 1, 2 and 4 leave the trial early. Then we know when they leave the study, but we don't know information about them as the trial continues.

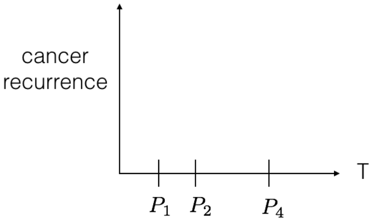

Figure 1: Example of censoring for medical data.

This is a certain type of missing data.

Heart Disease (Censoring) Example

- Researchers are studying the length of life (lifetime) following a particular medical intervention, such as a new surgical treatment for heart disease.
- The study consists of 12 patients.
- The number of years before death for each is

$$3.4, 2.9, 1.2+, 1.4, 3.2, 1.8, 4.6, 1.7+, 2.0+, 1.4+, 2.8, 0.6+$$

where $x+$ indicates that the patient was alive after x years, but the researchers lost contact with the patient at that point.

Consider the following model:

$$X_i = \begin{cases} Z_i & \text{if } Z_i \le c_i \\ * & \text{if } Z_i > c_i. \end{cases} \tag{1}$$

$$Z_1, \ldots, Z_n | \theta \overset{\text{iid}}{\sim} \text{Gamma}(r, \theta) \tag{2}$$

$$\boldsymbol{\theta} \sim \text{Gamma}(a, b) \tag{3}$$

where a, b, and r are known, and $*$ is a special value to indicate that censoring has occurred. The interpretation is:

- X_i is the observation
  - if the lifetime is less than c_i then we get to observe it $(X_i = Z_i)$,
  - otherwise all we know is the lifetime is greater than c_i $(X_i = *)$.
- θ is the parameter of interest—the rate parameter for the lifetime distribution.
- Z_i is the lifetime for patient i, however, this is not directly observed.
- c_i is the censoring time for patient i, which is fixed, but known only if censoring occurs.

Gibbs saves us again!

Straightforward approaches that are in closed form don't seem to work (think about these on your own). Instead we turn to GS.

To sample from $p(\theta, z_{1:n} | x_{1:n})$, we cycle through each of the full conditional distributions,

$$\begin{aligned}
\theta &\mid z_{1:n}, x_{1:n} \\
z_1 &\mid \theta, z_{2:n}, x_{1:n} \\
z_2 &\mid \theta, z_1, z_{3:n}, x_{1:n} \\
&\vdots \\
z_n &\mid \theta, z_{1:n-1}, x_{1:n}
\end{aligned}$$

sampling from each in turn, always conditioning on the most recent values of the other variables.

Recall

$$X_i = \begin{cases} Z_i & \text{if } Z_i \le c_i \\ * & \text{if } Z_i > c_i. \end{cases}$$

$$Z_1, \ldots, Z_n | \theta \overset{\text{iid}}{\sim} \text{Gamma}(r, \theta)$$

$$\boldsymbol{\theta} \sim \text{Gamma}(a, b)$$

The full conditionals are easy to calculate. Let's start with $\theta | \cdots$

- Since $\theta \perp x_{1:n} \mid z_{1:n}$ (i.e., θ is conditionally independent of $x_{1:n}$ given $z_{1:n}$),

$$p(\theta | \cdots) = p(\theta | z_{1:n}, x_{1:n}) = p(\theta | z_{1:n}) \tag{4}$$

$$= \text{Gamma}\left(\theta \,\middle|\, a + nr,\, b + \textstyle\sum_{i=1}^{n} z_i\right) \tag{5}$$

using the fact that the prior on θ is conjugate.

Now let's move to z? What happens here? This is the start of **Homework 5.**

1. Find the full conditional for $(z_i \mid \cdots)$.
2. Code up your own multi-stage GS in R. Be sure to use efficient functions.
3. Use the censored data

$$3.4, 2.9, 1.2+, 1.4, 3.2, 1.8, 4.6, 1.7+, 2.0+, 1.4+, 2.8, 0.6+$$

   and replicate such plots with explanations as in the Toy Example from Module 7. Specifically, give (a) give traceplots of all unknown paramaters from the G.S. (b) a running average plot, (c) the estimated density of $\theta \mid \cdots$ and $z_9 \mid \cdots$. Be sure to give brief explanations of your results.