Bayesian Statistics, Assignment for Tuesday, Dec. 10

Our final subject is Gibbs sampling.

Reading

Our final subject before Exam 3 is Gibbs sampling, and I am not going to have you look at Donovan and Mickey Chapter 16, because they have obscured what is simple. Instead, I have written up my own version of the Gibbs sampling theory and will hand it out on Tuesday:

* https://brianhill.github.io/bayesian-statistics/resources/MonteCarloMethodsWhyDoTheyWork-III.nb.pdf

My reference for Gibbs sampling and how it is applied from Chapters 1 and 2 of *Markov Chain Monte Carlo in Practice,* by Gilks, Richardson, and Spiegelhalter. To keep all of our eyes on the final prize, I want to reiterate that we are trying to get ourselves to the point of understanding the methods used in the case study on Hepatitis B vaccination. I am working up my own version of the case study that will be ready on Tuesday, Dec. 17th:

* https://brianhill.github.io/bayesian-statistics/resources/MonteCarloMethodsCaseStudy.nb.pdf

In-Class Exercise Serving as Problem Set 18 — Illustrating Gibbs Sampling

Gibbs sampling only gets interesting when you are trying to sample multiple independent axes in the posterior. So our iPhone sales example, which only had one axis, isn't going to cut it as a Gibbs sampling example. I am therefore going to beef up the potato-rock example to have two axes.

We are going to one axis (or attribute) of the nodules be whether they are a potato or a rock. We are going to have another axis (or attribute) be whether they are small, medium, or large. The field contains some concentration of each of these six possibilities. The contents of the field is our prior. Here is the story problem with a lot of blanks for you to fill in.

Exercise 1 — The Priors

20000 nodules are dug up from the field. The field is 90% rocks and 10% potatoes.

Of the rocks, 20% are small, 40% are medium, and 40% are large. Also, of the potatoes 20% are small, 40% are medium, and 40% are large.

Turn those facts into a table of the priors:

$N_{\text{rocks total}} =$
$N_{\text{potatoes total}} =$

$N_{\text{rock small}} =$
$N_{\text{rock medium}} =$
$N_{\text{rock large}} =$
$N_{\text{potato small}} =$
$N_{\text{potato medium}} =$
$N_{\text{potato large}} =$

Exercise 2 — The Likelihoods and the Priors Make the Posteriors

The Likelihoods

The harvester is 83.33333% efficient at rejecting small rocks (that’s 5/6 rejected and 1/6 slip through). It is 87.5% efficient at rejecting medium rocks (that’s 7/8 rejected and 1/8 slip through). It is 95.83333% efficient at rejecting large rocks (that’s 23/24 rejected and 1/24 slip through).

The harvester is 75% efficient at keeping small potatoes (that’s 3/4). It is 75% efficient at keeping medium potatoes (that’s 3/4). It is 37.5% efficient at keeping large potatoes (that’s 3/8).

The Posteriors

The priors (the amounts in which they occur in the field), times the likelihoods (what the potato harvester does) makes the posteriors (what is sent to the kitchen):

$N_{\text{rock small}} =$
$N_{\text{rock medium}} =$
$N_{\text{rock large}} =$

$N_{\text{potato small}} =$
$N_{\text{potato medium}} =$
$N_{\text{potato large}} =$

The kitchen still gets more rocks than potatoes, but if you did all the numbers right, you’ll see that the harvester has drastically improved the mix. I get 60% rocks and 40% potatoes out of a total of 3000 nodules sent to the kitchen.

Exercise 3 — Normalizing The Posteriors

$p_{\text{rock small}} = 300/3000 = 1/10 = 0.10 = 10\%$
$p_{\text{rock medium}} =$
$p_{\text{rock large}} =$

$p_{\text{potato small}} =$
$p_{\text{potato medium}} =$
$p_{\text{potato large}} =$

Exercise 4 — Conditional Probabilities — Along Size Axis

Now we have to compute probabilities along each axis! Suppose the Gibbs sampling algorithm has just made a tally at potato-medium and the algorithm is doing a move along the size axis. What are:

$p_{\text{small}\,|\,\text{potato}} =$
$p_{\text{medium}\,|\,\text{potato}} =$
$p_{\text{large}\,|\,\text{potato}} =$

Exercise 5 — An Algorithm Step — Along Size Axis

The algorithm could now generate a random number between 0 and 1, and if it is a 0.00 to 0.25, it goes to potato-small, 0.25 to 0.75 it goes to potato-medium, and 0.75 to 1.00 it goes to potato-large. That is an implementation of the conditional probabilities you found. The number is 0.88. Where do you go?

______________________ . You make a tally there.

Exercise 6 — New Conditional Probabilities — Along Potato-Rock Axis

Then the algorithm does a move along the potato-rock axis. What are the conditional probabilities applicable to this move?

$p_{\text{rock}\,|\,\text{large}} =$
$p_{\text{potato}\,|\,\text{large}} =$

Exercise 7 — Another Algorithm Step — Along Potato-Rock Axis

Another random number is drawn. If it is between 0.00 and 0.50, you go to rock-large. If it is between 0.50 and 1.00, you go to potato-large. Let's say it is 0.66 so you go to:

______________________. You make a tally there.

Exercise 8 — New Conditional Probabilities — Along Size Axis

Now the algorithm tries a move along the size axis again. You have the same condition, and you are moving along the size axis again, so the same three probabilities as you found in Exercise 4 apply:

$p_{\text{small}\,|\,\text{potato}} =$
$p_{\text{medium}\,|\,\text{potato}} =$
$p_{\text{large}\,|\,\text{potato}} =$

Exercise 9 — Another Algorithm Step — Along Size Axis

You draw another random number, and it is 0.22. Where do you go?

____________________.   You make a tally there.

Exercise 10 — New Conditional Probabilities — Along Potato-Rock Axis

The algorithm generates its next move based on these conditional probabilities:

$p_{\text{rock}\,|\,\text{small}} =$
$p_{\text{potato}\,|\,\text{small}} =$

Exercise 11 — Another Algorithm Step — Along Potato-Rock Axis

You draw another random number, it is 0.55. You find yourself still at

____________________.   You make another tally there.

Exercise 12 — New Conditional Probabilities — Along Size Axis

The next move will be along the size axis and will be based on

$p_{\text{small}\,|\,\text{rock}} =$
$p_{\text{medium}\,|\,\text{rock}} =$
$p_{\text{large}\,|\,\text{rock}} =$

Exercise 13 — Repeat 1,000,000 Times :)

Repeat steps like 5-12 to get 999,995 more tallies. Use the representative sampling to draw some conclusions. Submit your paper. To appear: A.Da, H.Arper, H.Exi, J.Acob, J.Eremy, R.Ania, R.En, R.Uby, S.Asha, T.Ahm, “Village Feeding with Efficient Harvester,” *J. International Farming* **10** (2025) 1917.