DS 102 Discussion 4

Wednesday, February 16, 2022

1. **Rejection Sampling**

   Recall that in Bayesian inference, we want to understand the posterior distribution, $p(\theta|X)$. Using Bayes' Theorem, we have that

   $$p(\theta|X) \propto p(X|\theta)p(\theta)$$

   where $p(X|\theta)$ and $p(\theta)$ are the likelihood and prior, respectively. Last week, we looked at solving for the exact posterior distribution via conjugate priors. However, this property only applies to special pairs of likelihood and prior distributions. To find the posterior distribution for any likelihood and prior pair, we turn to approximate inference, which involves repeatedly drawing samples from the posterior distribution.

   In this problem, we explore the properties of *Rejection Sampling*, a method which can draw samples from a specified target distribution. The algorithm is defined as follows:

Algorithm 1 Rejection Sampling

Require: Target Distribution $f(x)$, Proposal Distribution $g(x)$
Ensure: $f(x) \leq Mg(x)$ for all x **and** the support of g *includes* the support of f
  repeat
    Draw sample X_i from $g(x)$
    Compute the ratio $R = \frac{f(X_i)}{Mg(X_i)}$
    Draw sample U from Uniform$[0, 1]$
  until $U \leq R$
  return X_i

   (a) *Setting a Proposal Distribution*

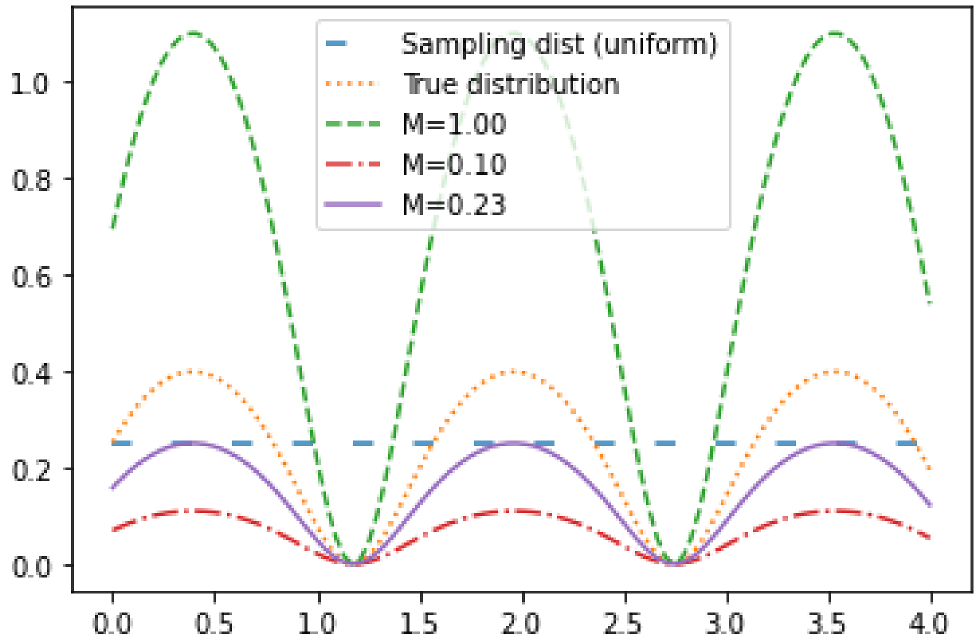
   We want to sample from the following, scaled unnormalized target distribution (where M is a scalar):

   $$f(\theta) = Mlog\left(2 + sin(4\theta)\right)$$

   for $\theta \in [0, 4]$. Find a valid proposal distribution that can be used in a rejection sampler for f.

(b) *Understanding the Scaling Factor*

Examine the following plot of the target distribution, for varying values of M.



What is special about the target distribution when $M \approx 0.23$? How is this related to the acceptance probability of a particular sample X_i?

(c) *Deriving the Acceptance Probability*

Now, let's consider a more general case, where f is some unnormalized target distribution and g is some proposal distribution. Under this scheme, what is the probability that we accept a sample, i.e. the probability $\mathbb{P}\left(U \leq \frac{f(X_i)}{Mg(X_i)}\right)$? What is the largest probability we can get by changing M?

2. **Gibbs Sampling for Gamma-Poisson model**

When the dimension of the parameters is large, sampling from the posterior over *all* the parameters θ is often difficult. The main insight behind Gibbs sampling is that it can be much easier to sample the posterior over just a *single* parameter, $\mathbb{P}(\theta_i \mid X, \theta_{-i})$ (where we use the index $-i$ to mean all indices except for i). Gibbs sampling then iterates through each parameter θ_i and samples from $\mathbb{P}(\theta_i \mid X, \theta_{-i})$. This loop is repeated, each time conditioning on the newly sampled values. Iterating through each parameter θ_i and sampling from $\mathbb{P}(\theta_i \mid X, \theta_{-i})$ is not the same thing as sampling from $\mathbb{P}(\theta \mid X)$. However, the good news is that given enough iterations, the former converges to the latter.

In this problem, we'll use a Bayesian hierarchical model to model the number of failures, X_i, for each of n power plant pumps[1]. Consider the following Gamma-Poisson model,

$$\begin{aligned} \beta &\sim \text{Gamma}(m, \alpha) \\ \theta_i \mid \beta &\sim \text{Gamma}(k, \beta), \quad i = 1, \ldots, n \\ X_i \mid \theta_i &\sim \text{Poisson}(\theta_i), \quad i = 1, \ldots, n, \end{aligned}$$

where the θ_i are independent of each other and represent the rate of failures for each power plant pump. The β and θ_i are unknown parameters, and m, α, and k are fixed and known.

We'd like to infer the parameters β and the θ from the data X. That is, we'd like to sample from the posterior distribution $\mathbb{P}(\beta, \theta \mid X)$ using Gibbs sampling. This entails deriving the posterior of each parameter, conditioned on the data and all the other parameters.

(a) *Drawing the Graphical Model*

Draw a graphical model which best represents the specified Gamma-Poisson model.

[1] E I George, U E Makov, and A F M Smith. Conjugate likelihood distributions. Scandinavian Journal of Statistics, 20:147–156, 1993.

(b) *Finding the Conditional Distribution of β*

Let's start with β. Derive $\mathbb{P}(\beta \mid \theta_1, \ldots, \theta_n, X_1, \ldots, X_n)$.

(c) *Finding the Conditional Distribution of θ_i*

Next, we'll look at each θ_i. Derive $\mathbb{P}(\theta_i \mid \beta, \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n, X_1, \ldots, X_n)$

(d) *Intuition for Gibbs Updates*

Interpret the derivations in parts (b) and (c) in words.

(e) *Implementing a Gibbs Sampler in Pseudocode*

Using the posterior distributions you derived in the last two parts, write out the algorithm for the Gibbs Sampler.

3. **Review of Markov Chains**

Oh no! The Wi-Fi is down at campus libraries once again. Frustrated by your lack of productivity, you leave to study somewhere else. The next day, you want to figure out whether you should study at a campus library. To make this decision, you model the Wi-Fi connection at the library as a Markov chain with two states: **1** for online and **0** for offline.

(a) *Drawing a Markov Chain and its Transition Matrix*

Your friend, Sarah, from IT staff tells you the following information:

- The chance of IT staff fixing the Wi-Fi when it is offline is 0.9
- The chance of the Wi-Fi failing after being online the previous day is 0.2

Using this information, draw a two-state Markov chain and write down its transition matrix.

(b) *Finding a Steady-State Distribution*

Looking at your Markov chain, you notice that it is has a finite state space, it is irreducible[2], and is aperiodic[3]. This means your Markov chain converges to some steady-state distribution $\vec{\pi}$. In other words, you can find the expected long-run proportion of time the Wi-Fi will fail! Write out and solve a system of equations to find the steady-state distribution of this chain.

[2]Check: Every state can be reached by all other states, so this chain is irreducible.

[3]Check: Each state can be reached in odd and even time steps, so this chain is aperiodic.