DS 102 Discussion 4

Wednesday, Sep 22, 2021

1. **Graphical Models**

   Last lecture, we were introduced to *Graphical Models*, which are flexible diagrams to express the relationships between random variables. An important special case of graphical models is the Bayesian hierarchical model, which generally may look like the figure below:

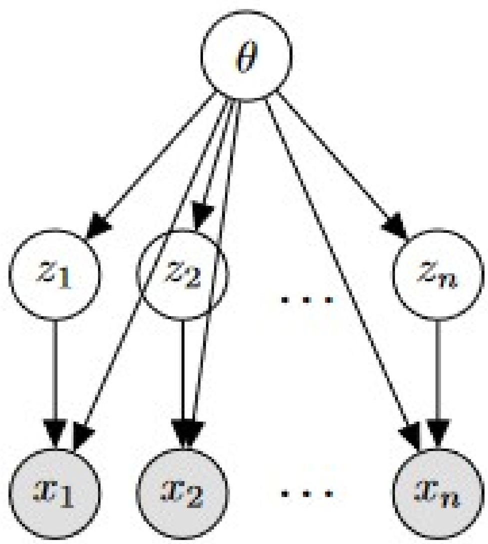


Figure 1: **Bayesian hierarchical model with hyperparameter θ, latent variables z_i, and observed variables x_i**

   In a Bayesian hierarchical model, observations are independent given the latent variables, and each observed variable depends only on its corresponding latent variable and the hyperparameters. As a result, Bayesian hierarchical models are always depicted as *directed acyclic graphs* (DAGs).

   In the following subparts, we will create our own graphical model and explore its properties.

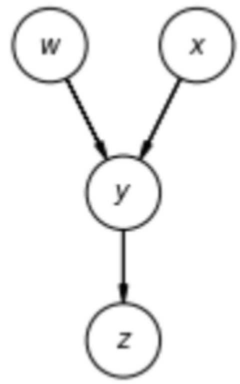
(a) *Formulating a Graphical Model*

Suppose you are a farmer who wants to model the upcoming crop harvest. You are interested in the following variables:

- w is the amount of pesticide used
- x is the amount of total rainfall for the season
- y is the number of bugs found in the field
- z is the total crop yield

Draw a graphical model to illustrate the relationships between these variables.

(b) *Identifying Independence and Conditional Independence*
Consider the following graphical model:



Which of the following statements are true about the graphical model above?

1. $x \perp\!\!\!\perp w$
2. $w \perp\!\!\!\perp x \mid y$
3. $w \perp\!\!\!\perp z \mid y$

2. **Rejection Sampling**

Recall that in Bayesian inference, we want to understand the distribution of 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 will 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
  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

The rejection sampling method generates sampling values from a target distribution X with arbitrary probability density function $f(x)$ by using a proposal distribution Y with probability density $g(x)$. The idea is that one can generate a sample value from X by instead sampling from Y and accepting the sample from Y with probability $f(x)/(Mg(x))$, repeating the draws from Y until a value is accepted. M here is a constant, finite bound on the likelihood ratio $f(x)/g(x)$, satisfying $1 < M < \infty$ over the support of X; in other words, M must satisfy $f(x) \leq Mg(x)$ for all values of x. Note that this requires that the support of Y must include the support of X—in other words, $g(x) > 0$ whenever $f(x) > 0$.

The validation of this method is the envelope principle: when simulating the pair $(x, v = u \cdot Mg(x))$, one produces a uniform simulation over the subgraph of $Mg(x)$. Accepting only pairs such that $u < f(x)/(Mg(x))$ then produces pairs (x, v) uniformly distributed over the subgraph of $f(x)$ and thus, marginally, a simulation from $f(x)$.

(a) *Setting a Proposal Distribution*

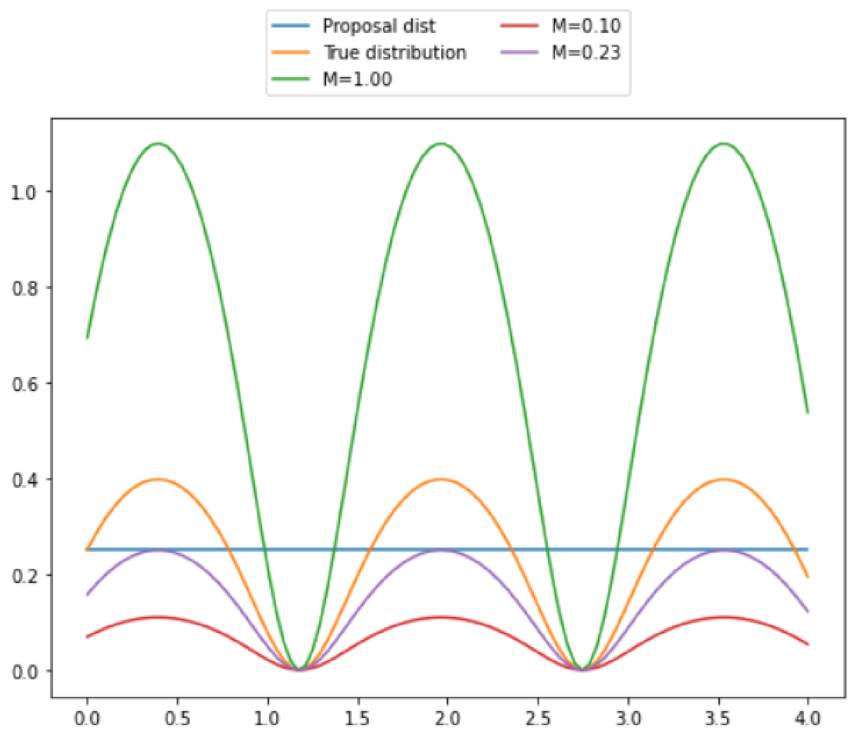
We want to sample from the following, scaled unnormalized target distribution:

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

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?

3. **Review of Markov Chains**

Oh no! The Wi-Fi is down at Moffitt Library 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 go to the library once again. 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 MC 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[1], and is aperiodic[2]. 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.

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

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