STA360 Exam II Fall 2020

Instructions

- This exam is open/note and open book. It is open to all class resources. You can also use the internet; if you do please cite any resource used that help you (including your notes).
- Write your name, NetID, and signature below.
- **Only what is on the exam will be graded (or written work submitted as one pdf file).**
- **Show all work and back up all your results for full credit.**
- **You must label/assign pages when submitting to Gradescope to avoid losing points.**
- **The exam should be submitted via Gradescope. You will have 15 minutes to submit your exam to Gradescope after the exam.**
- **You are not to talk to another classmate, tutor, TA, or any person regarding help on the exam. You are not to seek help from posting from an online resource. Doing so may result in getting a 0 on the examination and failing the course. In addition, you should not post to Piazza during the exam period to avoid leaking information to your classmates that would give anyone an unfair advantage.**

Community Standard

To uphold the Duke Community Standard:

- I will not lie, cheat, or steal in my academic endeavors;
- I will conduct myself honorably in all my endeavors; and
- I will act if the Standard is compromised.

I have adhered to the Duke Community Standard in completing this exam.

Name: ______________________________

NetID: ______________________________

Signature: ______________________________

List of common distributions

$$\text{Geometric}(x|\theta) = \theta(1-\theta)^x\, \mathbb{1}(x \in \{0,1,2,\ldots\}) \text{ for } 0 < \theta < 1$$

$$\text{Bernoulli}(x|\theta) = \theta^x(1-\theta)^{1-x}\, \mathbb{1}(x \in \{0,1\}) \text{ for } 0 < \theta < 1$$

$$\text{Binomial}(x|n,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}\, \mathbb{1}(x \in \{0,1,\ldots,n\}) \text{ for } 0 < \theta < 1$$

$$\text{Poisson}(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}\, \mathbb{1}(x \in \{0,1,2,\ldots\}) \text{ for } \theta > 0$$

$$\text{Exp}(x|\theta) = \theta e^{-\theta x}\, \mathbb{1}(x > 0) \text{ for } \theta > 0$$

$$\text{Uniform}(x|a,b) = \frac{1}{b-a}\, \mathbb{1}(a < x < b) \text{ for } a < b$$

$$\text{Gamma}(x|a, b = \text{rate}) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}\, \mathbb{1}(x > 0) \text{ for } a, b > 0,$$

$$\text{Gamma}(x|a, b = \text{scale}) = \frac{1}{b^a\Gamma(a)} x^{a-1} e^{-x/b}\, \mathbb{1}(x > 0) \text{ for } a, b > 0,$$

$$\text{Pareto}(x|\alpha, c) = \frac{\alpha c^\alpha}{x^{\alpha+1}}\, \mathbb{1}(x > c) \text{ for } \alpha, c > 0$$

$$\text{Beta}(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}\, \mathbb{1}(0 < x < 1) \text{ for } a, b > 0$$

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\tfrac{1}{2\sigma^2}(x-\mu)^2\big) \text{ for } \mu \in \mathbb{R},\ \sigma^2 > 0$$

$$\mathcal{N}(x|\mu,\lambda^{-1}) = \sqrt{\frac{\lambda}{2\pi}} \exp\big(-\tfrac{1}{2}\lambda(x-\mu)^2\big) \text{ for } \mu \in \mathbb{R},\ \lambda > 0$$

1. (6 points) Free response.

   (a) (1 point) When performing Markov chain Monte carlo (such as Metropolis sampling or Gibbs sampling), the user can decide whether or not to use a burn-in. **Explain why or why not this statement is correct.**

   (b) (2 points) Write out Bayes' rule to express the posterior distribution. Which term in your expression of Bayes's rule creates an issue (**circle the term**) and motivates the use of Markov chain Monte Carlo? **Explain your reasoning/choice for full/partial credit.**

   (c) (1 point) Name a situation when we would use Gibbs sampling over importance or rejection sampling? **Explain your reasoning/choice for full/partial credit.**

   (d) (1 point) Name an algorithm or method for sampling from a truncated Normal distribution. **Explain your reasoning/choice for full/partial credit.**

   (e) (1 point) A high acceptance rate of a Metropolis sampler means that the resulting samples will be high quality. **Explain if this statement is correct or not in a few sentences.**

Extra space for problem 1.

2. (6 points) True/False. (There will be no partial credit since there are only two choices).

   (a) (1 point) Importance sampling is fairly efficient as the dimensionality of the problem increases.
   True or False.

   (b) (1 point) If a conjugate prior is available for a given sampling model, we should **always** adopt the conjugate prior for easy computation.
   True or False.

   (c) (1 point) A latent variable is one that can be measured from the data.
   True or False.

   (d) (1 point) If each of the full conditionals in a Gibbs sampler is proper, the joint distribution is necessarily proper.
   True or False.

   (e) (1 point) If you have samples from $p(\gamma|Y)$ and $p(\kappa|Y)$, then you necessarily also have samples from $p(\gamma, \kappa|Y)$.
   True or False.

   (f) (1 point) In Figure 1, there is a trace plot for the parameter μ. We can safely say that our sampler has definitely converged to the stationary distribution.
   True or False.

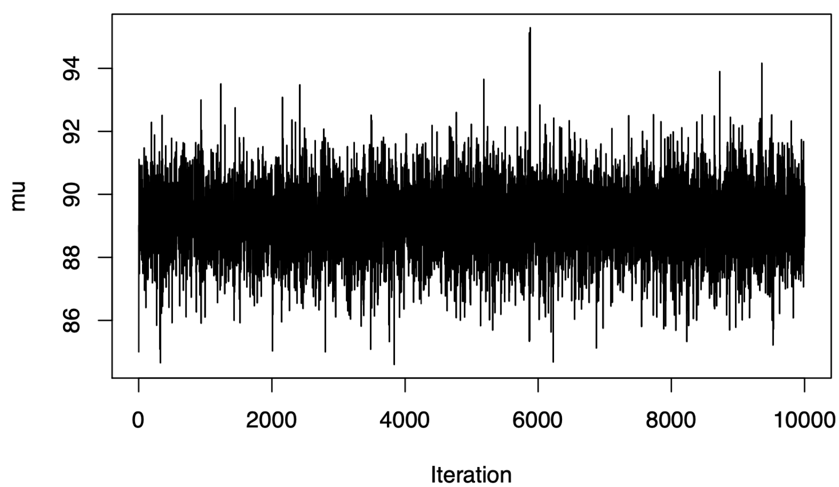

Figure 1: Traceplot for μ

3. (6 points) (The Two Stage-Gibbs Sampler) Consider the following model

$$y_i \mid \theta_i,\ \lambda \overset{ind}{\sim} \text{Poisson}(\theta_i \lambda) \tag{0.1}$$
$$\theta_i \sim \text{Gamma}(\nu/2, \nu/2) \tag{0.2}$$
$$\lambda \sim \text{Gamma}(c, d), \tag{0.3}$$

for $i = 1, \ldots, n$. **Use the Gamma density with the rate parameter in your derivation.**

(a) (4 points) Derive the conditional distributions $\theta_i \mid y_i, \lambda$ and $\lambda \mid y_i, \theta_i$.
Hint: Write the likelihood for $y_i \mid \theta_i,\ \lambda$. Use the likelihood to help you write out the two conditional distributions.

(b) (2 points) Explain how to use the conditional distributions in part (a) to perform Gibbs sampling to approximate the posterior distribution. **In describing your Gibbs sampling algorithm, outline your approach either via pseudocode or an outline of the procedure. Make sure to detail all the steps needed. (You DO NOT need to write pseudo-code as required in a COMPSI class; you DO NOT need to write any R code).**

Extra space for problem 3.

4. (6 points) (Gaussian Mixture models) Consider the following **Gaussian mixture model** for the data. Let Y_i be the glucose concentration (in mg/dL) of individual i, and let Z_i be an **unobserved, latent membership variable equaling 0 or 1.** Assume the following mixture model for observations $i = 1, \ldots, n$:

$$Y_i \mid Z_i, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2 = \begin{cases} N(\theta_1, \sigma_1^2), & \text{if } Z_i = 0, \\ N(\theta_2, \sigma_2^2), & \text{if } Z_i = 1, \end{cases} \quad i = 1, \cdots, n.$$

In addition, we will assume $Z_i|p \overset{i.i.d.}{\sim} \text{Bernoulli}(p)$, with independent priors:

$$p \sim \text{Beta}(a, b), \quad \theta_1, \theta_2 \overset{i.i.d.}{\sim} \text{Normal}(\mu, \tau^2), \quad \sigma_1^2, \sigma_2^2 \overset{i.i.d.}{\sim} \text{InverseGamma}(\nu/2, \nu\gamma^2/2).$$

(a) (0.5 point) Write out the likelihood of $p(y_{1:n} \mid Z_i, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$.

(b) (0.5 point) Given the likelihood (with the inclusion of the latent variable), explain why the posterior would be *easier* to sample from using Markov chain Monte Carlo, such as Gibbs sampling?

(c) (1 point) Using the latent variable formulation, derive the joint posterior distribution up to a normalizing constant.

(d) (4 points) Derive/provide **all** necessary conditional distributions needed to approximate the posterior $\theta \mid y_1 \ldots y_n$. That is, derive the following conditional distributions:

$$\theta_1 \mid y_i, \theta_2, \sigma_1^2, \sigma_2^2, p, z_i \tag{0.4}$$
$$\theta_2 \mid y_i, \theta_1, \sigma_1^2, \sigma_2^2, p, z_i \tag{0.5}$$
$$\sigma_1^2 \mid y_i, \theta_1, \theta_2, \sigma_2^2, p, z_i \tag{0.6}$$
$$\sigma_2^2 \mid y_i, \theta_1, \theta_2, \sigma_2^1, p, z_i \tag{0.7}$$
$$p \mid y_i, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, z_i \tag{0.8}$$
$$z_i \mid y_i, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p \tag{0.9}$$

Hint 1: Write $\tau_1 = 1/\sigma_1^2$ and $\tau_2 = 1/\sigma_2^2$. Then use a fact that you know about the Inverse Gamma and Gamma distributions.
Hint 2: If you can write the conditional distributions down directly, you do not need to re-derive them.

Extra space for problem 4.

Extra space for problem 4.