STA360 Exam I Fall 2020

Instructions

- This exam is closed note and closed book.
- 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.**

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: ____________________

Score

(For TA use only — leave this section blank.)

1. ____________/20
2. ____________/10
3. ____________/10
4. ____________/10

Overall: ____________________/50

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. (20 points) Assume the data is $x_{1:n} = (x_1, \ldots, x_n)$ and θ is the unknown parameter of interest. Let x_{n+1} be a new data point. **Circle the most correct answer (or correct answers).**

   (a) (2.5 points) Assume that you have the likelihood $p(x_{1:n} \mid \theta)$, the prior $p(\theta)$, and marginal distribution $p(x_{1:n})$ available to you. The posterior distribution $p(\theta \mid x_{1:n})$ can be derived (written) as the following:

   i. $$p(\theta \mid x_{1:n}) = \frac{p(x_{1:n} \mid \theta)p(\theta)}{p(x_{1:n})}$$

   ii. $$p(\theta \mid x_{1:n}) = \frac{p(x_{1:n}, \theta)}{p(x_{1:n})}$$

   iii. $$p(\theta \mid x_{1:n}) = \frac{p(\theta, x_{1:n})}{p(x_{1:n})}$$

   iv. $$p(\theta \mid x_{1:n}) \propto p(x_{1:n} \mid \theta)p(\theta)$$

   (b) (2.5 points) Suppose that $X \sim \text{Binomial}(n, \theta)$ and $\theta \sim \text{Beta}(a, b)$, where $a, b > 0$ and known.

   a. The posterior distribution is $\theta \mid x \sim \text{Binomial}(n, \theta)$

   b. The posterior distribution is $\theta \mid x \sim \text{Beta}(a, b)$

   c. The posterior distribution is $\theta \mid x \sim \text{Binomial}(n + a, x + b)$

   d. The posterior distribution is $\theta \mid x \sim \text{Beta}(a + x, b + n - x)$

(c) (2.5 points) What is the minimal condition that must be satisfied in order to use an improper prior in Bayesian inference?

  i. The posterior distribution must be continuous.
  ii. The prior distribution must be symmetric.
  iii. The prior distribution must have finite mean and variance.
  iv. The posterior distribution must be proper.

(d) (2.5 points) Which of the following is not a meaningful prior elicitation approach?

  i. Expert opinion based upon the data
  ii. Prior centered weakly at the mean
  iii. Point mass at the mean
  iv. A flat non-informative (improper) prior

(e) (2.5 points) Assume $X_1, \ldots, X_n \mid \theta \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \lambda^{-1})$. Assume the precision $\lambda = 1/\sigma^2$ is known and fixed, and θ is given a $\mathcal{N}(\mu_0, \lambda_0^{-1})$ prior:

$$\boldsymbol{\theta} \sim \mathcal{N}(\mu_0, \lambda_0^{-1}).$$

Give the form of the posterior of $\theta \mid X_1, \ldots, X_n$ with the updated parameters. **(In the interest of time, it is recommended that you do not write out your derivation on your exam).**

(f) (5 points) Assume the **model in e**. When would it be reasonable to assume this model (unknown mean, fixed precision) versus putting a prior on the precision? Explain. If you did put a prior on the precision (λ), what choice would you make and why?

(g) (2.5 points) Why do we use conjugate priors? (Circle all correct answers.)

i. Conjugate priors always give us proper posterior distributions.

ii. Conjugate priors are priors of convenience.

iii. The resulting posterior inference does not depend on the choice of the prior parameters.

iv. Conjugate priors always explain the underlying data well.

2. (10 points) (Conjugacy, The Geometric-Beta Model)
   Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Geometric}(\theta)$ given θ. Consider a $\text{Beta}(a, b)$ prior on θ.

   (a) (2 points) Write down the likelihood $p(x_{1:n} \mid \theta)$.

   (b) (3 points) Derive the posterior distribution of $p(\theta \mid x_{1:n})$.

   (c) (5 points) Derive the marginal distribution $p(x_{1:n})$.

3. (10 points) (Decision theory)
Consider a decision problem in which the state is $\theta \in \mathbb{R}$, the observation is x, you must choose an action $\hat{\theta} \in \mathbb{R}$, and the loss function is

$$\ell(\theta, \hat{\theta}) = a\theta^2 + b\theta\hat{\theta} + c\hat{\theta}^2$$

for some known $a, b, c \in \mathbb{R}$ with $c > 0$. Suppose you have computed the posterior distribution and it is $p(\theta|x) = \mathcal{N}(\theta|M, L^{-1})$ for some M and L. What is the Bayes procedure (minimizing posterior expected loss)?

(Your answer must be an explicit expression in terms of a, b, c, M, and L. You must show your work to receive full credit.)

Extra space for problem 3.

4. (10 points) (Conceptual Component) Suppose you are working with an investigator to understand how massage therapy can alleviate exam stress. A lucky $n = 500$ college students will be selected and provided with free massage treatments the week before final exams. The outcome of interest is whether each student reports decrease in stress levels from a baseline to a follow-up interview in the middle of the exam week (which is measured as a binary outcome). Specifically, each student reports they had decrease in stress (x=1) or no decrease in stress (x=0).

   You may assume that the data is denoted by $x_{1:n}$, where $n = 500$. You may assume that there is an unknown parameter θ, which represents the probability that a student reports a decrease in stress levels from the baseline.

   (a) (1 point) Write out the likelihood function for the data $x_{1:n}$ described above. That is, write that $x_{1:n} \mid \theta \sim$ as what distribution? (You do not need to write the likelihood out in a formal equation).

   (b) (1 point) Explain why your likelihood function is appropriate in part (a).

   (c) (1 point) A pilot study of students in the investigator's lab found that 2 of 4 students reported decreased stress. How would one encode this information directly in a conjugate prior distribution? (Provide the prior distribution with specific values for the hyper-parameters of your prior).

   (d) (2 points) What form does the posterior distribution take (Normal, Gamma, etc.) when combining the likelihood with the prior? Provide the distribution and the updated parameters. (If you can state the updated parameters, you don't need to show any work).

   (e) (5 points) What benefits does your model specification have? What limitations does it have? (List one benefit and one limitation).

Extra space for problem 4.