

# STA360/601 Final Exam

## Instructions

- Write your name, NetID, and signature below.
- If you need extra space for any problem, continue on the back of the page.

## 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

1. \_\_\_\_\_

2. \_\_\_\_\_

3. \_\_\_\_\_

4. \_\_\_\_\_

5. \_\_\_\_\_

6. \_\_\_\_\_

7. \_\_\_\_\_

8. \_\_\_\_\_

Overall: \_\_\_\_\_

## List of common distributions

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

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

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

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

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

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

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

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

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

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

$$\mathcal{N}(x|\mu, C) = \frac{1}{(2\pi)^{d/2}|C|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T C^{-1}(x - \mu)\right) \text{ for } \mu \in \mathbb{R}^d, C \in \mathbb{R}^{d \times d} \text{ symmetric positive definite.}$$

## Exponential family form

$$p(x|\theta) = \exp(\varphi(\theta)^T t(x) - \kappa(\theta)) h(x)$$

## List of special functions

$$\text{Beta function: } B(a, b) = \int_0^1 t^{a-1} (1 - t)^{b-1} dt \text{ for } a, b > 0$$

$$\text{Gamma function: } \Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \text{ for } x > 0$$

1. (14 points) Graphical models

(a) (4 points) Draw the directed graphical model (DGM) for a Markov chain. Also draw the associated moral graph.

(b) (4 points) A hidden Markov model is a distribution on  $Z_1, \dots, Z_n, X_1, \dots, X_n$  that respects the DGM shown below. Write the factorization of  $p(z_1, \dots, z_n, x_1, \dots, x_n)$  implied by this DGM.

(c) (6 points) Consider a distribution respecting the DGM in part 1b. For each question, circle either Yes, No, or Indeterminate. Indeterminate means that the answer cannot be determined from the information given. (You do not need to justify your answer.)

- |                                    |     |    |               |
|------------------------------------|-----|----|---------------|
| i. Is $X_1 \perp X_3 \mid Z_2$ ?   | Yes | No | Indeterminate |
| ii. Is $X_1 \perp X_3$ ?           | Yes | No | Indeterminate |
| iii. Is $X_1 \perp X_3 \mid X_2$ ? | Yes | No | Indeterminate |

2. (10 points) (Semi-conjugate priors)

Given  $S \in \mathbb{R}^{d \times d}$  symmetric positive definite, and  $\nu > d - 1$ , the Wishart distribution with inverse scale  $S$  and  $\nu$  degrees of freedom has density

$$W_d(X | S^{-1}, \nu) = \frac{|S|^{\nu/2} |X|^{(\nu-d-1)/2} \exp(-\frac{1}{2} \text{tr}(SX))}{2^{\nu d/2} \Gamma_d(\nu/2)}$$

for  $X \in \mathbb{R}^{d \times d}$  symmetric positive definite. Here,  $\Gamma_d(\nu/2)$  is the multivariate gamma function (its definition is unimportant for this problem), and  $\text{tr}$  is the trace, i.e.,  $\text{tr}(A) = \sum_{i=1}^d A_{ii}$ .

Show that the Wishart distribution is a (semi-)conjugate prior for  $S$ . That is, if  $X_1, \dots, X_n | S \stackrel{\text{iid}}{\sim} W_d(S^{-1}, \nu)$  and  $S \sim W_d(S_0^{-1}, \nu_0)$ , then  $p(S | X_{1:n})$  is a Wishart distribution. (Show your work.) (Hint:  $\text{tr}(SX) = \text{tr}(XS)$ .)

3. (10 points) (Multivariate normal)

- (a) (5 points) How can you transform a random vector  $X \sim \mathcal{N}(\mu, C)$  into a  $\mathcal{N}(0, I)$ -distributed random vector? (Hint: Any symmetric positive definite matrix  $C$  can be factored as  $C = U\Lambda U^T$  where  $U$  is orthogonal and  $\Lambda$  is diagonal with positive diagonal entries. Use this and the affine transformation property.)
- (b) (5 points) Suppose  $X \sim \mathcal{N}(a, C)$  and  $Y \sim \mathcal{N}(b, D)$  independently, with means  $a, b \in \mathbb{R}^d$  and covariance matrices  $C, D \in \mathbb{R}^{d \times d}$  respectively. What is the distribution of  $X + Y$ ? Justify your answer. (Hint: A fairly easy way to do this is to note that  $\begin{pmatrix} X \\ Y \end{pmatrix}$  is multivariate normal, and use the affine transformation property.)

(Extra paper for question 3.)

4. (14 points) (Linear regression, Variable selection)

Vibration analysis is often used to monitor the condition of large complicated factory equipment. To monitor a particular machine, you have a sensor that precisely measures the position of the machine over time. Suppose that at times  $x_1, \dots, x_n \in \mathbb{R}$ , you obtain position measurements  $y_1, \dots, y_n \in \mathbb{R}$ . Since the movement is periodic, you model the mean position at time  $x_i$  as

$$\sum_{k=0}^m c_k \cos(2\pi k x_i)$$

for some known  $m$  and some unknown  $c_0, c_1, \dots, c_m \in \mathbb{R}$ .

- (a) (5 points) Assuming Gaussian noise, write down the likelihood of a linear regression model for the position measurements. Assume a common variance  $\sigma^2$  for all times.
- (b) (4 points) Write down a (semi-)conjugate prior for the regression coefficients.
- (c) (5 points) Each coefficient  $c_k$  represents the amount of vibration at a particular frequency. To analyze the machine, you want to know which frequencies it is vibrating at.
  - i. In words, describe how variable selection can be used to address this.
  - ii. You want to do Gibbs sampling for variable selection. Assume you have already derived an expression for the marginal likelihood  $p(y|x, z)$ , where  $y = y_{1:n}$ ,  $x = x_{1:n}$ , and  $z = z_{1:m}$  with  $z_k \in \{0, 1\}$  being the indicator variable for whether  $c_k$  is included. Assuming a uniform prior on  $z$ , derive an expression for  $\mathbb{P}(Z_k = 1 \mid z_{-k}, y, x)$ , where  $z_{-k}$  denotes the variables in  $z$  other than  $z_k$ .

(Extra paper for question 4.)



5. (14 points) (Bayesian hypothesis testing)

Suppose you are conducting an experiment on the effects of exercise on cognitive performance. To assess performance, you measure the time required to solve a creative thinking task. You divide subjects into a control group of  $m$  subjects and a test group of  $n$  subjects. The test group exercises for 30 minutes before performing the task, while the control group does not exercise beforehand. You measure the time required by each individual to solve the task, yielding data  $x_1, \dots, x_m$  and  $y_1, \dots, y_n$ . You want to consider the evidence for the two hypotheses,  $H_0$ : no difference between groups, and  $H_1$ : the groups are different.

- (a) (5 points) Modeling the data as Exponential, write down a model to address this using Bayesian hypothesis testing. Use conjugate priors, and assume the two hypotheses have equal prior probability.
- (b) (5 points) Give a closed-form expression for the Bayes factor  $B_{01}$  in favor of  $H_0$  over  $H_1$ .
- (c) (4 points) Give an expression for the posterior probability of  $H_1$  in terms of the Bayes factor  $B_{01}$ .

(Extra paper for question 5.)

6. (14 points) (Hierarchical models, Gibbs sampling)

In another experiment on cognitive performance, you have  $n$  subjects play the computer game Tetris, and you count how many “tetrises” they get, resulting in a nonnegative integer score. Each subject plays the game 5 times, yielding scores  $x_{i1}, \dots, x_{i5} \in \{0, 1, 2, \dots\}$  for subject  $i$ .

- (a) (5 points) Modeling the scores for each subject as Poisson, write down a hierarchical model for this data, using semi-conjugate priors. For the top level in your model, put a prior on only one of the hyperparameters (one for which there is a nice semi-conjugate prior).
- (b) (4 points) Draw the directed graphical model (DGM) for your model.
- (c) (5 points) Derive closed form expressions for all the full conditionals required to do Gibbs sampling. (Show the derivations.)

(Extra paper for question 6.)

7. (10 points) (Markov chains)

Find a transition matrix  $T$  on three states (say,  $\{1, 2, 3\}$ ) which gives rise to a Markov chain that is irreducible and has stationary distribution  $\pi = (1/4, 1/2, 1/4)$ , BUT is periodic (i.e., not aperiodic). (For full credit, you must explain why it has each of the properties stated above.)

8. (14 points) (Metropolis–Hastings MCMC)

Consider a target distribution  $\pi(x)$  on a discrete space, and a transition matrix  $T$ .

- (a) (4 points) What does it mean for  $\pi$  and  $T$  to have detailed balance? (Give the mathematical definition.)
- (b) (5 points) Suppose  $T$  is the transition matrix corresponding to a Metropolis–Hastings move with proposal distribution  $q_x(x^*)$ . Write down an expression for  $T_{xy}$  when  $x \neq y$ . (Note: A Metropolis–Hastings move refers to a single propose–accept/reject iteration.)
- (c) (5 points) Using your answers from the first two parts, show that Metropolis–Hastings moves always have detailed balance. Assume that  $\pi(x) > 0$  and  $q_x(x^*) > 0$  for all  $x, x^*$ . (Hint: You don’t need an explicit expression for  $T_{xx}$  since the case of  $x = y$  is trivial.)