BASKIN SCHOOL OF ENGINEERING

DEPARTMENT OF STATISTICS

2020 First Year Exam: Monday June 8th, 2020

INSTRUCTIONS

- The exam is from 1:00 PM to 5:00 PM. Submissions are accepted until 5:30 PM.
- This is a closed-book and closed-notes exam. You are expected to work alone on the exam without any outside resources.
- Please write all your answers on blank sheets of paper or blank notes. You do not need to rewrite the questions. On the top of each page, write your exam ID number and the problem number. Please start each of the five questions on a new page. You may use the same page for subparts of a question. Please show all work clearly.
- After finishing the exam, scan each page and assemble the pages into one document. A pdf file is highly preferred. Check that your scanned answers are legible and email your document to the graduate director at `juheelee@soe.ucsc.edu`.
- Please carefully read the statement below before you start the exam.

> As a student at UC Santa Cruz, I hold myself to a high standard of integrity, and by submitting this exam I reaffirm my pledge to act ethically by honoring the UC Santa Cruz Code of Student Conduct. I will also encourage other students to avoid academic misconduct.
> I acknowledge that the work I submit is my individual effort. I did not consult with or receive any unauthorized help from any person or other source. I also did not provide help to others. I may work with others or consult other resources only if the instructor gave specific instructions, and only to the extent allowed by the instructor.

Some formulas that may be useful:

- The density function of the beta distribution with parameters $\alpha > 0$ and $\beta > 0$ is given by

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1 \;.$$

- The density function of the half-Cauchy distribution is given by

$$f(x) = \frac{2}{\pi}\frac{1}{1+x^2}, \quad x > 0.$$

- The density function of the inverse-gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ is given by

$$f(x) = \exp\left(-\frac{\beta}{x}\right)\frac{\beta^\alpha}{x^{\alpha+1}\Gamma(\alpha)}, \quad x > 0.$$

- The density function of the gamma distribution with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$ (mean α/β, variance α/β^2) is given by

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\exp(-\beta x), \quad x > 0 \;.$$

Use the formulas above for the distributions in the upcoming questions.

Problem 1 (STAT 203):

A machine produces parts that are either defective or not. Let θ denote the proportion of defectives among all parts produced by this machine. Suppose that we observe a fixed number n of such parts and let Y be the number of defectives among the n parts observed. We assume that the parts are conditionally independent given θ. Suppose that the marginal distribution of θ follows the beta distribution with parameters α and β.

(i) (30%) Find the conditional distribution of $\theta \mid Y$.

(ii) (30%) If $\alpha = \beta = 1$, find the marginal distribution of Y.

(iii) (10%) Find the mean of Y.

(iv) (30%) Find the variance of Y.

Problem 2 (STAT 205B):

Consider a random sample $X_i, i = 1, \ldots, n$, from the $Unif(0, \theta)$ distribution, with unknown $\theta > 0$.

(i) (10%) Derive the likelihood of θ and find the MLE, $\hat{\theta}$.

(ii) (15%) Show that $\hat{\theta}$ is a function of a minimal sufficient statistic.

(iii) (20%) Find the distribution of $\hat{\theta}$.

(iv) (30%) Find a pivot $T(\theta, \hat{\theta})$ based on the CDF of $\hat{\theta}$ and use the equation

$$Pr(a \leq T(\theta, \hat{\theta}) \leq b) = 0.95, \ a, b \in [0, 1]$$

to find a family of 95% confidence intervals for $\hat{\theta}$.

(v) (25%) Find the value of a that produces the shortest interval in the confidence interval family in part (iv).

Problem 3 (STAT 206B):

Let the response y_i and the p dimensional predictor $\mathbf{x}_i = (x_{i,1}, ..., x_{i,p})'$ obey the linear regression framework

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i, \ \ \epsilon_i \overset{i.i.d.}{\sim} \mathrm{N}(0,1), \quad i = 1, \ldots, n, \tag{1}$$

where n is the number of data points and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ is a $p-$dimensional vector of unknown regression coefficients. The notation $'$ denotes the transpose of a matrix or a vector. Assume the following prior distributions on the parameters;

$$\beta_j \mid \lambda_j^2, \tau^2 \overset{ind}{\sim} \mathrm{N}(0, \lambda_j^2\tau^2), \ \ \lambda_j \overset{i.i.d.}{\sim} C^+(0,1), \ \ \tau \sim C^+(0,1), \tag{2}$$

where $C^+(0,1)$ is the half-Cauchy distribution on the positive real line. We further assume that λ_j in j and τ are mutually independent.

(i) (15%) Write down the full posterior distribution of the parameters.

(ii) (45%) *Fact:* $\alpha^2 \mid \eta \sim$ Inv.-Gamma$(1/2,\ 1/\eta)$, and $\eta \sim$ Inv.-Gamma$(1/2,\ 1)$ together imply $\alpha \sim C^+(0,1)$.

Use the above fact to design a data augmentation procedure by introducing latent variables so that all full conditional posterior distributions appear in familiar distributional forms. Write down the full conditional distributions of all parameters (including the latent variables).

(iii) (25%) Describe a Gibbs sampler for drawing samples from the posterior distribution of the parameters. Discuss choices of initial values.

(iv) (15%) Now assume that except for $\boldsymbol{\beta}$, all other parameters are fixed and known. Consider the loss function $L(\boldsymbol{\beta}, \mathbf{D}) = (\boldsymbol{\beta} - \mathbf{D})'(\boldsymbol{\beta} - \mathbf{D})$, where $\mathbf{D}$ is the decision rule. Derive the Bayes estimator for $\boldsymbol{\beta}$.

Problem 4 (STAT 207):

A probit regression entails associating a set of covariates to a binary response, using a link function based on the cumulative distribution of the standard normal random variable, denoted by $\Phi(\cdot)$. For i.i.d. samples $y_1, \ldots, y_n$ of a Bernoulli random variable with probability of success p_i, let $p_i = \Phi(\boldsymbol{x}_i'\boldsymbol{\beta})$. Here, $\boldsymbol{x}_i$ is a k-dimensional set of covariates for the ith observation, and $\boldsymbol{\beta}$ is a k-dimensional vector of coefficients. Thus, the likelihood for $\boldsymbol{\beta}$ can be obtained from the distribution

$$f(y_i|\boldsymbol{\beta}, \boldsymbol{x}) = \Phi(\boldsymbol{x}_i'\boldsymbol{\beta})^{y_i}(1 - \Phi(\boldsymbol{x}_i'\boldsymbol{\beta}))^{1-y_i}.$$

(i) (30%) Introduce the latent variables $z_1, \ldots, z_n$, $z_i \overset{ind}{\sim} \mathrm{N}(\boldsymbol{x}_i'\boldsymbol{\beta}, 1)$, such that

$$y_i = \begin{cases} 1 & \text{if} \quad z_i > 0 \\ 0 & \text{if} \quad z_i \leq 0 \end{cases}$$

to augment the model. Show that the augmented model is equivalent to the original model.

(ii) (35%) Assuming that $p(\boldsymbol{\beta}) \propto 1$, obtain and identify the full conditionals of each z_i and $\boldsymbol{\beta}$.

(iii) (35%) Suppose that the distribution of z_i is changed to $t_\alpha(\boldsymbol{x}_i'\boldsymbol{\beta}, 1)$. Use the representation

$$t_\alpha(\mu, \sigma^2) = \int_0^\infty \mathrm{N}(w \mid \mu, \sigma^2/\lambda)\mathrm{Ga}(\lambda \mid \alpha/2, \alpha/2)d\lambda,$$

where $\mathrm{Ga}(\lambda|a, b)$ is a gamma distribution with shape parameter a and rate parameter b such that $E(\lambda) = a/b$, to introduce an appropriate second set of latent variables. Obtain the full conditional distributions in this case.

Problem 5 (STAT 209):

Consider the general linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{Y}$ is an n-dimensional vector of observations, $\mathbf{X}$ is an $n \times p$ matrix of full rank, $\boldsymbol{\beta}$ is a p-dimensional unknown parameter vector, and $\boldsymbol{\epsilon}$ are n-dimensional random errors that have the multivariate normal distribution

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Gamma}).$$

You may assume that $\boldsymbol{\Gamma}$ is invertible.

(i) (25%) What is the distribution of $\boldsymbol{\Gamma}^{-1/2}\mathbf{Y}$, where $\boldsymbol{\Gamma}^{-1/2}$ is a matrix square root of $\boldsymbol{\Gamma}^{-1}$?

(ii) (25%) Derive an expression for the weighted least squares estimator of $\boldsymbol{\beta}$, which is denoted by $\hat{\boldsymbol{\beta}}$.

(iii) (25%) Derive the mean and variance of $\hat{\boldsymbol{\beta}}$.

(iv) (25%) Explain how to test the null hypothesis that $\beta_i = 0$ against the alternative that $\beta_i \neq 0$. Here, β_i denotes the ith component of $\boldsymbol{\beta}$.