STAT305: Introduction to Inference **Fall 2021-22**

Final: instructions + questions

Instructions

Read this right away! We will distribute the exam questions at 15:35 PT

1. Once you have read this, write on a piece of paper labelled "Question 0" the following, to be submitted on Gradescope by the time limit (18:00 PT):
   (a) Your zoom id.
   (b) Write down: "I have read all the instructions, in particular, I will not communicate with any other humans (except the 305 instructors) during the exam." Add your legal signature.
2. I have taken a selfie picture of myself holding the first page of my written submission and have submitted this picture to Gradescope. Both you and the paper should be visible.
3. Failure to do both (1) and (2) above will result in marks being deduced.
4. Due to last minute technical problems with Webwork, we are not able to use Webwork. Instead, the questions are provided in a PDF (zoom chat window). Submit your answers, both numerical and derivations, to Gradescope.
5. I understand that if there is a technical issue with Gradescope, it is still my responsibility to transmit my written answers within the time limit by email including all the following email addresses: bouchard@stat.ubc.ca, jiayang.yin@stat.ubc.ca, li.zha@stat.ubc.ca, jzhan@cheos.ubc.ca, grace.yin@stat.ubc.ca
6. Failure to comply to the exam rules, including those above and those shared before the exam on piazza, will lead to serious consequences ranging from mark deduction to failing the exam and academic misconduct investigation.
7. Justify all your answers.
8. Timeline (note: for student with accommodations from the Accessibility Office, use the information in parenthesis below to make the adjustments):
   - 15:30 PT: You receive these instructions but without the exam questions. Read the instructions. If you have time, you may get started on the tasks in (1) to test Gradescope submission. If you encounter any problem you can then notify us early and still start your exam in time.
   - 15:35 PT (5 min after start of exam): You receive the exam questions via the Zoom chat window.
   - 17:40 PT (20 min before end of exam): You are required to stop writing and start the Gradescope submission process
   - 18:01 PT (1 min after end of exam): 5% penalty for late submissions between 18:01 and 18:04.
   - 18:05 PT: submission site closes and we stop accepting exam answers.

Latex document compiled on: Tuesday 21st December, 2021, 08:00

1 Discrete prior

Let $X = 22$ be a realization drawn from a Binomial$(43, \pi)$ distribution. Suppose you have a discrete prior distribution on the random variable Π defined via its probability mass function $f(\pi)$ as follows:

$$f(\pi) = \begin{cases} 0.4 & \pi = 0.3 \\ 0.6 & \pi = 0.7 \end{cases} \tag{1}$$

1. Derive the posterior distribution $P(\Pi = 0.3|X = 22)$.
   [8 mark(s)]
2. Evaluate $P(\Pi = 0.3|X = 22)$.
   [2 mark(s)]

Question total: [10 mark(s)]

2 Maximum Likelihood and information

Let Y_i be exponentially distributed with rate parameter $\lambda_i = x_i^7 \exp(\beta)$, where x_i are known, positive constants, and β is an unknown parameter. You can use without proof that the joint loglikelihood of this model is given by:

$$\log f(y_1, \ldots, y_n; \beta) = n\beta + \sum_{i=1}^{n} \left(7 \log x_i - \exp(\beta) x_i^7 y_i\right).$$

Consider a dataset with $(x_1, x_2) = (1.5, 0.7)$ and $(y_1, y_2) = (0.0001124, 0.081625)$.

1. Show that the MLE is given by
   $$\hat{\beta} = \log n - \log \sum_{i=1}^{n} x_i^7 y_i,$$
   and explain if the above expression is an estimate or an estimator. For the next questions, you can use without justification that for the dataset shown, $\hat{\beta} = 5.44420$.
   [5 mark(s)]
2. Derive the observed information matrix and write down its numerical value.
   [6 mark(s)]
3. Compute a 80% confidence interval based on the observed information matrix.
   [4 mark(s)]
4. Derive the Fisher information matrix and write down its numerical value.
   [3 mark(s)]

Question total: [18 mark(s)]

3 Non identically distributed observations

Let Y_i for $i = 1, 2, \ldots, n$ be n independently distributed Poisson($i\mu$) (note this is not a typo!) random variables.

1. Write down the log likelihood function. You may denote constants with C, C', and so on. **[4 mark(s)]**
2. Derive the maximum likelihood estimator $\tilde{\mu}$ of μ. Justify each step. Provide the estimate for the dataset $(y_1, y_2, y_3) = (3, 8, 5)$.

   [4 mark(s)]
3. Let $\tilde{\mu}_a = \frac{\sum_{i=1}^{n} Y_i}{n^2}$. If $n = 9, \mu = 6$, compute the bias of $\tilde{\mu}_a$, i.e., bias$(\tilde{\mu}_a) = E(\tilde{\mu}_a) - \mu$.

   [5 mark(s)]
4. Is $\tilde{\mu}_a$ asymptotically unbiased? Justify your answer.

   [4 mark(s)]

Question total: [17 mark(s)]

4 Most powerful test

Suppose the random variable X is a discrete random variable with four possible values, $x \in \{1, 2, 3, 4\}$ and with a probability mass function $f(x|\pi)$ shown below. We are interested in testing:

$$\mathrm{H}_0 : \pi = \pi_0 \quad \text{vs } \mathrm{H}_A : \pi = \pi_A$$

We are given the following probability mass functions under π_0 and π_A:

x	1	2	3	4
$f(x\|\pi_A)$	0.41	0.47	0.04	0.08
$f(x\|\pi_0)$	0.52	0.42	0.03	0.03

Assume we have only one sample from the random variable X, and therefore, do not use results based on asymptotic approximations.

1. What is the **most powerful test** for testing H_0 vs H_A and significance level $\alpha = 0.03$? Clearly state the decision rule in your written answer and the result that shows that it is the most powerful test.

   [4 mark(s)]
2. What is the power of the test from part 1?

   [2 mark(s)]

Question total: [6 mark(s)]

5 MGFs

1. Let V denote a random variable with a moment generating function given by $M_V(t) = \exp(2(e^t - 1))$. What is the moment generating function of $2V + 4$ evaluated at one?

[3 mark(s)]

2. Let $X \sim \text{Geom1}(0.5)$ (using same notation as course pack, i.e. the support of X is $1, 2, 3, \dots$) and $Y = aX + b$. Suppose that $M_Y(t) = e^{5t}/(2 - e^{2t})$ is the moment generating function of Y. Determine what are the values for a and b.

[2 mark(s)]

Question total: [5 mark(s)]

6 Posterior distribution

Consider a Bayesian model where $Y_i|Z \sim \text{Uniform}(0, Z)$. We assume the following prior density for the unknown parameter Z, which you can assume without providing a proof that it is a valid probability density for any $t > 0$ and $\beta > 0$:

$$f_Z(z|t, \beta) = \frac{\beta t^\beta}{z^{\beta+1}} \mathbf{1}_{z \geq t}, \tag{2}$$

where $\mathbf{1}_{\dots}$ denotes the indicator function. The following facts can be used without proof to save you many computations: the mean of the above density is $\beta t/(\beta - 1)$ when $\beta > 1$, and its cumulative distribution function is $F_Z(z|t, \beta) = (1 - (t/z)^\beta)\mathbf{1}_{z \geq t}$.

1. You observe a dataset $(y_1, y_2, y_3, y_4) = (16.2, 0.9, 0.6, 8.0)$. Suppose you use the prior shown in Equation (2) with $t = 1$ and $\beta = 5.2$. Derive the posterior density of Z by using a conjugacy argument. Hint: from the dataset, you will only need to know the maximum value, $y^* = \max y_i$.

   Report the parameters of the posterior distribution and the posterior mean below.

[7 mark(s)]

2. Suppose, for a different data set, the parameters of the posterior distributions are $t = 18$ and $\beta = 12.2$. Derive a 95% credible interval of the form $[y^*, R]$. Report the value of R. **[5 mark(s)]**

Question total: [12 mark(s)]

7 Many Gaussian distributed random variables

Let $X_1, \dots, X_6$ be $n = 6$ identically and independently Gaussian distributed random variables with mean $\mu = 1.6$ and variance $\sigma^2 = 4.5$. Show how to use a result from Chapter 2 to compute the probability

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i^2 - 2\mu n\bar{X} \leq 20\right), \tag{3}$$

where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$.

[5 mark(s)]

Question total: [5 mark(s)]