Gateway Quiz 2021W1_305_final

Jump to Problem: [1 2 3 4 5 6 7 **]**

Problem 1.

(1 point) Before beginning the exam, please read and affirm the following:

Enter your zoom username exactly as entered (write 'test' in the chat window if you are not sure and want to see your username). If it is not found in the recording you will get a grade of zero. Answer with your initials.

I swear or affirm that I will not interact with other humans in any way during the duration of the exam (e.g. I will not send or read messages, I will not discuss, etc). Answer with your initials (do not use the zoom account, use your real name, for example, if your real name is John Smith, write JS).

I understand that I need to submit my derivations within 20 minutes of the end of the exam or will have grades taken off afterward on a per-minute basis. Answer with your initials.

I confirm that I will write my numerical answers both on webwork and on the document that I will submit on gradescope. Answer with your initials.

I confirm that I have saved this page (e.g. using print > save as PDF) to ensure I can continue in the unlikely event that webwork crashes. Answer with your initials.

I confirm that I will click, during the allotted time but ONLY once when you are finished, the button 'Grade test'. Answer with your initials.

I confirm that I will use 'Preview answers' to save my progress (answers can still be modified after clicking 'preview answers'). Answer with your initials.

Note: *You can earn partial credit on this problem.*

preview answers

Problem 2.

(1 point) Suppose we conducted a poll in which we asked the following questions: first, their education background (A, secondary only, B undergraduate, C, graduate, and D, others) and second, their employment (either, “Full-time”, “Part-time”, and “Unemployed”). The null hypothesis is that the education background is independent of that person’s employment classification. The data are tabulated as:

	A	B	C	D
Full time	30	60	67	91
Part time	41	76	41	73
Total	119	190	148	217

Hint: do not forget the category "Unemployed"!

Part 1)
The Pearson statistic is 20.978. Calculate the first term in the sum used to compute this statistic (the term corresponding to ”Full-time” and ”A”). **[5 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below. Round your answer to 3 decimal places.

Part 2)
What is the degree of freedom of the test? **[4 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below.

Part 3)
What is the p-value? Give your answer to 3 significant digits. **[3 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below.

Note: *You can earn partial credit on this problem.*

preview answers

Problem 3.

(1 point) Let $X_1, X_2, \ldots, X_n$ denote a random iid sample, each from the probability density function:

$$f(x;\theta) = \theta x^{\theta-1}\mathbf{1}[0 \leq x < 1],$$

where $\theta > 0$ is an unknown parameter and $\mathbf{1}[\texttt{expression}]$ denotes the indicator function, i.e. a function which is one if the `expression` is true and zero otherwise.

At a given significance level 0.03 we want to test:

$$H_0 : \theta = 0.15 \quad \text{versus} \quad H_1 : \theta = 0.9.$$

Part 1)
Consider a test statistic of the form:

$$T(x_1, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^{n}\log(x_i),$$

with the associated rejection region:

$$R = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) > k^*\},$$

where k^* is such that the test has significance level 0.03. Is there as test with a higher power for this problem and fixed n? Justify your answer. **[4 marks]**

Write your derivations and answer on paper submitted to Gradescope.

Part 2)
Consider the following dataset consisting of 3 data points: $0.6, 0.2, 0.9$. Let $Y_i = \log X_i$. You can use without proof that $E[Y_i] = -1/\theta$ and $\text{Var}[Y_i] = 1/\theta^2$. Using this and the central limit theorem, obtain an approximate value for k^* such that the test has significance level 0.03. **[4 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below.
Round your answer to 3 decimal places.

Part 3)
In the same setting as above, what is the p-value? **[2 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below.
Round your answer to 4 decimal places.

Note: *You can earn partial credit on this problem.*

preview answers

Problem 4.

(1 point) **Part 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 $0.7V + 0.9$ evaluated at one? **[3 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below.
Round your answer to the nearest integer.

Part 2)
Let $X \sim \text{Pois}(4)$ and $Y = f(X)$ for some linear function f. Suppose that

$$M_Y(t) = \exp(4e^{3t} + 5t - 4)$$

is the MGF of Y. Determine what the function $f(x)$ is. Input $f'(0)$ below. **[2 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below.
Round your answer to 3 decimal places.

Note: *You can earn partial credit on this problem.*

preview answers

Problem 5.

(1 point) Let Y_i be exponentially distributed with rate parameter $\lambda_i = x_i exp(3\beta)$, where x_i are known constants, and β is an unknown parameter. You can use without proof that the joint log likelihood of this model is given by:

$$\log f(y_1, \ldots, y_n; \beta) = 3n\beta + \sum_{i=1}^{n} (\log x_i - \exp(3\beta)x_i y_i) .$$

Consider a dataset with $(x_1, x_2) = (5, 6)$ and $(y_1, y_2) = (2, 7)$.

Part 1)
Show that the MLE is given by

$$\hat{\beta} = \frac{1}{3}\log \frac{n}{\sum_i x_i y_i}$$

and explain if the above expression is an estimate or an estimator. For the data shown, $\hat{\beta} = -1.086$. **[4 marks]**

Write your derivations and answer on paper submitted to Gradescope.

This expression is an:
- **A.** Estimator
- **B.** Estimate

Part 2)
Derive the observed information and input its value below. **[2 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below.
Round your answer to 3 decimal places.

Part 3)
Compute a 60% confidence interval based on the observed information. Is it exact or approximate? **[3 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical values below.
Round your answers to 3 decimal places.

Lower bound:

Upper bound:

This interval is:

- A. Exact
- B. Approximate

Part 4)
Derive the Fisher information and input its value below. Would an interval based on the Fisher information be exact or approximate? **[3 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below. Round your answer to 3 decimal places.

An interval based on the Fisher information matrix would be:

- **A.** Approximate
- **B.** Exact

Part 5)
We now take $x_i = 1$ for all i. Consider a hypothesis test for $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$. For a large dataset, what is the approximate distribution of the following test statistic:

$$T = 2n \times 3\hat{\beta} + 2(1 - \exp(3\hat{\beta})) \sum_{i=1}^{n} Y_i.$$

Justify your answer. **[3 marks]**

Write your derivations and answer on paper submitted to Gradescope.

Note: *You can earn partial credit on this problem.*

preview answers

Problem 6.

(1 point) In this problem, denote the i-th sample by y_i, modelled as an iid realization of the random variable $Y_i \sim \text{Uniform}[-\theta, 0)$, with unknown $\theta > 0$. Consider the dataset of size $n = 3$ given by $y = (y_1, y_2, y_3) = (-3.5, -3.8, -2.2)$. Denote the maximum value in y by m^* and the minimum, by m_*. Denote the associated random maximum and minimum by M^* and M_*.

Part 1)

Show that the joint likelihood is given by **[4 marks]**

$$f_Y(y|\theta) = \theta^{-n}\mathbf{1}[-\theta \le m_*]\mathbf{1}[m^* < 0].$$

Write your derivations and answer on paper submitted to Gradescope.

Part 2)

Derive the MLE $\tilde{\theta}$ for θ. Give the numerical value for the estimate of $\hat{\theta}$ below. (Note: if you are stuck in parts 2 or 3, you can skip to part 4.) **[4 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below. Round your answer to 3 decimal places.

Part 3)

What is the bias of $\tilde{\theta}$, if any? Provide a numerical value at $\theta = 4.3$. Hint: you can use the fact that $-M_*/\theta \sim \text{Beta}(3, 1)$ without proof. **[3 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below. Round your answer to 3 decimal places.

Part 4)

We now consider a Bayesian approach to the problem of estimating θ, which we now view as the realization of the random variable Θ. We consider the following family of priors, denoted $\mathscr{P}(t, \beta)$, defined for $t > 0, \beta > 0$:

$$f_\Theta(\theta; t, \beta) = \beta t^\beta \theta^{-\beta-1}\mathbf{1}[\theta \ge t].$$

In the following, you can use without proof that $E(\Theta) = \beta t/(\beta - 1)$ for $\beta > 1$ and $E(\Theta) = \infty$ for $\beta \le 1$. Suppose we use a prior where we set $t = 2.15$ and $\beta = 8$. What is the posterior distribution? Evaluate it at 4.8. **[4 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below. Round your answer to 3 decimal places.

Part 5)
Write down an expression for the estimator $E[\Theta \mid Y]$. Compute the numerical value of the estimate for the dataset and prior described in the previous parts of this question. **[2 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical value below. Round your answer to 3 decimal places.

Part 6)
We will now compare the MLE $\tilde{\theta}$ and the posterior mean $E[\Theta \mid Y]$. As the sample size n goes to infinity, compute the limit, in distribution, of

$$E[\Theta|Y] - \tilde{\theta}.$$

If we picked $t > \theta$, explain why the answer would be different. **[2 marks]**

Write your derivations and answer on paper submitted to Gradescope.

Note: *You can earn partial credit on this problem.*

preview answers

Problem 7.

(1 point) We will study the following Bayesian model: $Y_1, Y_2, \ldots, Y_n$ are such that $Y_i \mid M \sim Poisson(M)$. We assume the following prior density on M:

$$f_M(\mu) = \frac{1}{2}\text{Gamma}(\mu; \text{shape } \nu_1 = 7, \text{rate}_1 = 1) + \frac{1}{2}\text{Gamma}(\mu; \text{shape } \nu_2 = 3, \text{rate}_2 = 1).$$

Part 1)
Identify the conjugate family which will allow you to compute the exact form of the posterior density for this Bayesian model. **[2 marks]**

Write your derivations and answer on paper submitted to Gradescope.

Part 2)
Provide update rules on the prior parameters of the conjugate family to obtain a posterior density based on a dataset $y_1, \ldots, y_n$. Provide numerical values for the updated prior parameters for the following dataset of size $n = 2 : (5, 2)$. **[2 marks]**

Write your derivations and answer on paper submitted to Gradescope, enter the numerical values below.

ν_1

v_2

rate$_1$

rate$_2$

Note: *You can earn partial credit on this problem.*

preview answers

Preview Test Grade Test