STAT 510: Exam 02

David Dalpiaz

Monday, May 10

General Directions

The exam begins on the next page. You have one hour and three hours to compete the exam. This **includes** submitting the exam to **Gradescope**. Be sure to indicate which questions appear on which pages of your submission.

The following resources may be used during this exam:

- Pens, pencils, erasers, and blank scratch paper.
- A laptop for logging into the CBTF Scheduler and obtaining your exam.
- A phone or tablet for running Zoom for proctoring. This may also be used to take photos of a paper exam at the end of the exam.
- A calculator.
- A student-created single-sheet formula sheet is allowed. (You may use both sides.)
  - You are not expected to know densities, means, variances, etc for well-known distributions. If they are needed, they will be provided.
- Symbolab, WolframAlpha, or Mathematica for assistance with calculations.
- Any websites necessary to create and upload relevant files to Gradescope.
  - Once these have been accessed, you should not change any answer or add any work to your exam.
- R and RStudio.

Please include as much supporting work as is reasonably possible. Do include any code used for any portion of the exam. If you use an external tool such as a calculator or WolframAlpha, please indicate where you have done so. **Mysterious or unsupported answers will not receive full credit**. A correct answer, unsupported by calculations, explanation, or algebraic work will receive no credit; an incorrect answer supported by substantially correct calculations and explanations might still receive partial credit.

Organize your work, in a reasonably neat and coherent way, in the space provided. Work scattered all over the page without a clear ordering will receive very little credit.

If you engage in an act of academic dishonesty, you become liable to severe disciplinary action. Such acts include cheating; falsification or invention of information or citation in an academic endeavor; helping or attempting to help others commit academic infractions; plagiarism; offering bribes, favors, or threats; academic interference; computer related infractions; and failure to comply with research regulations. **Do not discuss the exam with anyone until the instructor informs the entire course that it is OK to do so.**

Good Luck!

Question 01 (Exponential Likelihood)

Consider $X_1, \ldots, X_n \sim \text{Exp}(\lambda)$. That is,

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0.$$

(a) [2 points] Show that the MLE for λ is $\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$.

(b) [2 points] Find $I_n(\lambda)$ and $\text{se}\left(\hat{\lambda}\right)$.

(c) [1 point] Give an approximate 95% confidence interval for λ.

(d) [3 points] Find the MLE and an approximate 95% confidence interval for $2 \log \lambda$.

(e) [3 points] Suppose that in addition to the exponential likelihood above, we use a Gamma(α, β) prior. Derive the posterior distribution for λ. (Note that Gamma is a conjugate prior in this case. You simply need to show which Gamma distribution.) Use the following density for a Gamma(α, β) distribution:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\beta x}.$$

Note that with this parameterization, the mean is α/β. (This fact might be useful later.)

Question 02 (Geometric Likelihood)

Consider $X_1, \ldots, X_n \sim \text{Geometric}(p)$. That is,

$$f(x) = (1 - p)^{x - 1} \cdot p, \quad x = 1, 2, 3, \ldots, \quad 0 < p < 1.$$

(a) [2 points] Show that the MLE for p is $\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i}$.

(b) [2 points] Find $I_n(p)$ and $\text{se}\,(\hat{p})$.

(c) [2 points] Find the test statistic for the Wald test of

$$H_0 : p = 0.5 \quad \text{versus} \quad H_1 : p \neq 0.5.$$

(d) [3 points] Perform a simulation study to estimate the power function, $\beta(p)$ for $0 < p < 1$ when $\alpha = 0.05$. Use $n = 50$. Use a "reasonably" large number of simulations for each value of p. (That is, use enough values of p and enough simulations for each such that the resulting curve is reasonably smooth.) Plot the resulting curve.

Question 03 (Poisson Likelihood)

Consider $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$. That is,

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 1, 2, 3, \ldots, \quad \lambda > 0.$$

(a) [2 points] Show that the MLE for λ is $\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$.

(b) [2 points] Find $I_n(\lambda)$ and $\text{se}\left(\hat{\lambda}\right)$.

(c) [3 points] Recall that if we add a $\text{Gamma}(\alpha, \beta)$ prior for λ, the posterior distribution is given by

$$\lambda \mid X_1, \ldots X_n \sim \text{Gamma}\left(\alpha + \sum_{i=1}^{n} x_i,\ \beta + n\right).$$

Note that we are assuming the same parameterization as Question 01. Also note that the second parameter here is considered the "rate" parameter.

```
set.seed(42)
some_data = rpois(n = 25, lambda = 4)
```

Using the above data and a $\text{Gamma}(\alpha = 2, \beta = 5)$ prior, calculate three point estimates for λ.

- Using only the prior.
- Using only the data.
- Using both the prior and the data, that is, using the posterior.

(d) [3 points] Using the same data, prior, and scenarios as the previous part, calculate three 95% interval estimates.

Question 04 (Likelihood Ratio Test)

Assume

$$X_1, \ldots, X_{n_x} \sim \text{Exp}(\lambda_x) \quad \text{and} \quad Y_1, \ldots, Y_{n_y} \sim \text{Poisson}(\lambda_y).$$

Then, consider testing

$$H_0 : \lambda_x = \lambda_y \quad \text{versus} \quad \lambda_x \neq \lambda_y.$$

(a) [3 points] Find the MLE under the null. That is, find the MLE of λ assuming that $\lambda = \lambda_x = \lambda_y$.

(b) [2 points] Derive the LRT statistic for the above test. You do not need to do any algebraic simplifications.

(c) [1 point] State the asymptotic distribution of your statistic in the previous part.

(d) [3 points] Perform a simulation study to verify that this is a level α test for any α. To do so, repeatedly do the following:

- Generate data according to the data generating process under the null. For this question use $\lambda_x = \lambda_y = 2$.
- Calculate the test statistic for the generated sample.
- Calculate and store the (large sample) p-value.

Plot a histogram of these p-values. If done correctly, this histogram should indicate that the p-values are roughly uniform.