Homework 01 - STAT440

Joseph Sepich (jps6444)

8/31/2020

1. What is the best way to contact the instructor?

The best way to contact the instructor is through the Canvas messaging/email system and **not** via Penn State email.

2. Describe a resource listed in the syllabus (do not copy and paste).

One of the resources listed in the syllabus is Counseling and Psychological Services as UP otherwise known as CAPS. CAPS is a place for students to go regarding their mental health and will provide students with counseling or other services to help combat mental health issues that may effect their academic progress, social development, or emotional wellbeing. Sometimes CAPS even brings dogs during finals week to pet!

3. When do you have to complete the assigned readings per lecture?

The assigned readings must be completed **before each lecture**, so that you are prepared to participate in the lecture.

4. With the exception of the first homework, when are homeworks due?

The homework in STAT 440 will be due weekly on Fridays at 17:00 EST. (The first homework is due Monday Aug 31)

5. What is the general policy regarding collaboration?

Collaboration is encouraged in this class and will be enabled through the use of Canvas discussions or Piazza. This will allow for us to bounce ideas off each other and ask questions about the homework or general concepts in the class. The important aspect of this policy is that any work submitted must be your own as plagiarism is, was, and will always be unacceptable.

6. What are the different elements of the class that will be graded, and what is their percentage contribution to the total grade?

As denoted in the syllabus, **participation** will count towards 10% of the grade. Each of the other three elements, **homework**, **exams**, and the **final project** will count for 30% of the final grade.

7. What are the two elements of the final project, and what format can they be delivered?

The final project will involve both a presentation in class and a final report in the style of an academic paper. The final report can also be delivered in an alternative format such as a comic or movie; however any alternative mediums must be pre-approved.

8. In your own words, describe two etiquette principles for zoom meetings.

Two important Zoom etiquette principles are to remain engaged and to attend all classes on time. Just as if we were meeting in person, zoom lectures will start at the scheduled time, so you must be on time and remember that early is on time, and on time is late. Remaining engaged, the other principle, is vital for having productive lectures and a productive class. This principle includes not opening other stuff on the computer during lecture or having a “hot mic” playing music.

9. Have you read the syllabus in its entirety and understood it?

I, Joseph Sepich, have read the syllabus in its entirety and understood it.

10. Do you agree to abide by the principles listed in the syllabus?

I, Joseph Sepich, agree to abide by the principles listed in the syllabus.

11. Basic definitions.

Let X be an exponential random variable with rate parameter $\lambda > 0$.

(a) What is the range of X, its PDF, and its CDF?

The PDF of X is $f(x) = \lambda e^{-\lambda x}$ and the CDF is $F(x) = 1 - e^{-\lambda x}$. The support of X is $[0, \infty)$.

(b) What is the n^{th} moment of X?

The moment generating function for an exponential variable is

$$M_X(t) = \frac{\lambda}{\lambda - t}$$

The n^{th} moment is represented as

$$M_X^{(n)}(0) = \frac{n!}{\lambda^n}$$

This value is derived from the taking the derivative multiple times.

(c) What are the mean and variance of X?

Mean: $\mu = \frac{1}{\lambda}$

Variance:

$$\sigma^2 = \frac{1}{\lambda^2}$$

(d) Let $\epsilon > 0$. What should a be so that $(P(X > a) = \epsilon$?

For this let us use the CDF we defined above: $F(x) = 1 - e^{-\lambda x}$. We must determine for what value $1 - F(a) = \epsilon$.

$$\begin{aligned} 1 - (1 - e^{-\lambda a}) &= \epsilon \\ e^{-\lambda a} &= \epsilon \\ -\lambda a &= ln(\epsilon) \\ a &= -\frac{ln(\epsilon)}{\lambda} \end{aligned}$$

a must be $-\frac{ln(\epsilon)}{\lambda}$ for the probability of $X > a$ to be ϵ.

12. Transformations of random variables.

Let $\{X_i\}_{i=1}^N$ be i.i.d. exponential random variables with parameter λ. Let $(Y, Z) = T(X_1, X_2)$, where $T : D \to R$ is the transformation.

$$T(x, x') := (x + x', x - x')$$

(a) What is the joint PDF of $\{X_i\}_{i=1}^N$?

The PDF of a single exponential random variable as denoted above is $f(x) = \lambda e^{-\lambda x}$. When we multiply n of these independent, identically distributed random variables we then get the joint PDF of

$$f(x) = \lambda^n e^{-\lambda \Sigma_{i=0}^n X_i}$$

(b) What is the domain D of T?

The domain of the T is $[0, \infty)$, since it any value that a R.V. X can take. The domain could also be denoted as R^2 as it is 2-dimensional.

(c) What is the inverse transformation of T^{-1}, and its Jacobian determinant?

The inverse transformation of $T^{-1} := (\frac{1}{2}(Y+Z), \frac{1}{2}(Y-Z))$. This gives use the Jacobian determinant:

$$\frac{d(0.5(Y+Z))}{Y}\frac{d(0.5(Y-Z))}{Z} - \frac{d(0.5(Y+Z))}{Z}\frac{d(0.5(Y-Z))}{Y} = -0.25 - (0.25) = -0.5$$

(d) What is the range R of T?

The range R of T is the same as the domain R^2.

(e) What is the joint PDF of Y and Z? Provide the support also.

The joint pdf of Y is $\lambda(e^{-\lambda x} + e^{-\lambda x'})$. The joint pdf of Z is $\lambda(e^{-\lambda x} - e^{-\lambda x'})$. The support is positive real numbers.

(f) What are the marginal PDFs of Y and Z?

Marginal PDFs of Y (same for either x, just replace the x with x' and the term):

$$\int \lambda(e^{-\lambda x} + e^{-\lambda x'})dx = -e^{-\lambda x} + \lambda x e^{-\lambda x'}$$

Marginal PDFs of Z (only difference is the sign)

$$\int \lambda(e^{-\lambda x} - e^{-\lambda x'})dx = -e^{-\lambda x} - \lambda x e^{-\lambda x'}$$

13. Important Theorems

State the following.

(a) The Central Limit Theorem

The Central Limit Theorem states that the sampling distribution of the sample means approximates that of a normal distribution as the sample size becomes larger. Generally a sample size of 30 is considered large enough for this to be true. Paraphrased this theorem means that the average of your sample means will become normally distributed around the population mean.

(b) The Weak Law of Large Numbers

The Weak Law of Large Numbers states that the average of the results from a large number of trials should approach the expected value, assuming independent and identically distributed random variables that are being sampled.

14. MLE

Suppose that $X_1, X_2, ..., X_n$ are i.i.d. random variables with PDF:

$$f_x(x) = \frac{x}{a} exp\{\frac{-x^2}{2a}\}, x \geq 0$$

(a) Find the maximum likelihood estimate of the parameter a.

Recall that the likelihood of a parameter is related to the joint pdf:

$$L(a) = \frac{\Pi x_i}{a^n} exp\{\Sigma_{i=1}^n - x_i^2/2a\}$$

The maximum likelihood estimate is the point in the function that would cause a max. Let's derivate w.r.t a.

$$L'(a) = \frac{-n\Pi x_i}{a^n a} exp\{\frac{1}{2a}\Sigma_{i=1}^n - x_i^2\} + \frac{\Pi x_i}{a^n}\frac{-1}{2a^2} exp\{\frac{1}{2a}\Sigma_{i=1}^n - x_i^2\}$$

Setting this equal to zero to maximize we can remove the exp term (that would divide out):

$$\begin{aligned} 0 &= \frac{-n\Pi x_i}{a^n a} + \frac{\Pi x_i}{a^n}\frac{-1}{2a^2} \\ \frac{n\Pi x_i}{a^n a} &= \frac{\Pi x_i}{a^n}\frac{-1}{2a^2} \\ \frac{n}{a} &= \frac{-1}{2a^2} \\ a &= \frac{-1}{2n} \end{aligned}$$

(b) Find the Fisher Information of $X_1, X_2, ..., X_n$ and use it to estimate a 95% confidence interval on the MLE of a.

The loglikelihood is $l(a) = ln(\frac{\Pi x_i}{a^n}) + \frac{1}{2a}\Sigma_{i=1}^n - x_i^2$. We can use this to find the Fisher Information:

$$I(a) = -E[\frac{2n^2 - n}{\Pi x_i} a^{2n-2} + \frac{1}{a^3}\Sigma_{i=1}^n - x_i^2]$$

This makes our 95% confidence interval:

$$a + - \frac{1}{\sqrt{-E[\frac{2n^2-n}{\Pi x_i} a^{2n-2} + \frac{1}{a^3}\Sigma_{i=1}^n - x_i^2]}}$$

15. Good 'ol Bayes.

A lab blood test is 95% effective in detecting a certain disease when the disease is present. However, if a healthy individual is tested, there is a 1% chance that the test result will imply that this individual has the disease. If it is known that 5% of the population have the disease, what is the probability that a person has the disease, given that the test says they do?

Problem Constraints:

- Events:
  - Has disease: D
  - Positive Test: P
- $P(P|D) = 0.95$
- $P(P|D') = 0.01$
- $P(D) = 0.05$

Determine $P(D|P)$. First using Bayes theorem we can write:

$$P(D|P) = \frac{P(P|D)P(D)}{P(P)}$$

Using the law of total probability we can find $P(P)$, the probability of a positive test.

$$P(P) = P(P|D)P(D) + P(P|D')P(D') = 0.95 * 0.05 + 0.01 * 0.95 = 0.057$$

Plugging this result back into our original equation.

$$P(D|P) = \frac{P(P|D)P(D)}{P(P)} = \frac{0.95 * 0.05}{0.057} = 0.833$$

Therefore we know that the probability that a person has the disease, given that the test says they do, is **83.3%**.