Homework 6 Fall 2023 Due Date posted on canvas

Name: Ivan Leung **Student ID:** 67485846 **Section:** C

Instructions

Please read the following instructions carefully:

1. Please show all notation for probability statements.
2. Box your final answers.
3. Please verify that your scans are legible.
4. Please assign pages the the questions when submitting to gradescope.
5. This assignment is due via gradescope on the due date.

1. Suppose $X_1, \ldots, X_n \overset{iid}{\sim}$ Exponential(λ) is a set of n observations drawn independently from an Exponential distribution with rate parameter λ.

$\lambda e^{-\lambda x}$

(a) Write out the likelihood function.

$$L(\theta) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

$$= \boxed{\lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}}$$

(b) Write out the log-likelihood function.

$$l(\lambda) = \ln\left(\lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}\right)$$

$$= \ln(\lambda^n) + \ln\left(e^{-\lambda \sum_{i=1}^{n} x_i}\right)$$

$$= \boxed{n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i}$$

(c) Find the score function by taking the partial derivative of the log-likelihood function.

$$\frac{d\, l(\lambda)}{d\lambda} = \frac{d}{d\lambda}\left(n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i\right)$$

$$= \boxed{\frac{n}{\lambda} - \sum_{i=1}^{n} x_i}$$

(d) Set the score function equal to zero and solve for the parameter λ.

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\frac{n}{\lambda} = \sum_{i=1}^{n} x_i$$

$$\boxed{\frac{n}{\sum_{i=1}^{n} x_i}} = \lambda$$

(e) Take the second partial derivative of the score function.

$$\frac{d^2}{d\lambda^2} \text{score} < 0$$

$$\frac{d^2}{d\lambda^2}\left(\frac{n}{\lambda} - \sum_{i=1}^{n} x_i\right) < 0$$

$$\boxed{-\frac{n}{\lambda^2} < 0}$$

(f) Check to make sure this value is negative to ensure that the log-likelihood function is concave down.

since n and λ^2 are always positive, $-\frac{n}{\lambda^2}$ is always negative,

$\boxed{\text{It concaves down}}$

(g) You want to estimate the average time it takes for your roommate to cook dinner because it seems like they are ALWAYS in the kitchen. On a sample of 9 days, your roommate spent the following number of minutes cooking: (67, 76, 88, 90, 49, 98, 75, 100, 62). Assuming that the time it takes for them to cook dinner follows an Exponential Distribution with the same rate parameter λ, what is the Maximum Likelihood Estimator for the mean time that it takes for your roommate to cook dinner?

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{9}{67+76+88+90+49+98+75+100+62}$$

$$= \frac{9}{705} = 0.01276596$$

$$X \sim Exp\left(\lambda = \frac{1}{0.01276596} = 78.26 \text{ minutes}\right)$$

$$\boxed{78.26 \text{ minutes}}$$

2. Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$ is a set of n observations drawn independently from a Poisson distribution.

   (a) Write out the likelihood function.

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

$$= \boxed{e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^{n} x_i} \cdot \prod_{i=1}^{n} \left(\frac{1}{x_i!}\right)}$$

   (b) Write out the log-likelihood function.

$$\ell(\lambda) = \ln\left(e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^{n} x_i} \cdot \prod_{i=1}^{n} \left(\frac{1}{x_i!}\right)\right)$$

$$= \ln(e^{-n\lambda}) + \ln(\lambda^{\sum_{i=1}^{n} x_i}) + \ln\left(\prod_{i=1}^{n} \frac{1}{x_i!}\right)$$

$$= \boxed{-n\lambda + \sum_{i=1}^{n} x_i \ln(\lambda) - \sum_{i=1}^{n} \ln(x_i!)}$$

   (c) Find the score function by taking the partial derivative of the log-likelihood function.

$$\frac{d}{d\lambda} \ell(\lambda) = \frac{d}{d\lambda}\left(-n\lambda + \sum_{i=1}^{n} x_i \ln(\lambda) - \sum_{i=1}^{n} \ln(x_i!)\right)$$

$$= -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} - 0$$

$$= \boxed{-n + \frac{\sum_{i=1}^{n} x_i}{\lambda}}$$

(d) Set the score function equal to zero and solve for the parameter λ.

$$-n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

$$\frac{\sum_{i=1}^{n} x_i}{\lambda} = n$$

$$\sum_{i=1}^{n} x_i = n\lambda$$

$$\boxed{\frac{\sum_{i=1}^{n} x_i}{n} = \hat{\lambda}}$$

(e) Take the second partial derivative of the score function.

$$\frac{d^2}{d\lambda^2} \ell(\lambda) = \frac{d^2}{d\lambda^2} \left(-n + \frac{\sum_{i=1}^{n} x_i}{\lambda}\right)$$

$$= \boxed{-\frac{\sum_{i=1}^{n} x_i}{\lambda^2}}$$

(f) Check to make sure this value is negative to ensure that the log-likelihood function is concave down.

$\sum_{i=1}^{n} x_i$ is always positive.

square of any number is always positive.

$-\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0$, $\therefore$ $\boxed{\text{It concaves down.}}$

(g) You want to estimate the number of times that you see someone you know at UTC at lunch time. You go to UTC 12 times over the first 3 weeks of the quarter and you count the number of times you bump into a friend each time. The data is the following $(2, 6, 1, 8, 4, 6, 1, 9, 3, 6, 4, 8)$. Compute the Maximum Likelihood Estimator for the rate of friends that you see each lunch time at UTC, using the formula that you derived above.

$$\lambda = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{(2+6+1+8+4+6+1+9+3+6+4+8)}{12}$$

$$= \frac{58}{12}$$

$$= \boxed{4.83}$$

3. Suppose you are looking to improve your test scores and want to join a tutoring program. You have ultimately narrowed your choices to Program A and Program B. You want to join the program that has the higher test scores on average. You discover that out of a sample of 100 of Program A's students, they had a mean test score of 87 with a sample standard deviation of 4.8. For Program B, out of a sample of 120 students, they had a mean test score of 86 with a sample standard deviation of 2.2. You want to know if there is a difference in average test scores between the two programs.

   (a) What is the parameter of interest in this study?

Program A	Program B
$\bar{x}_1 = 87$	$\bar{x}_2 = 86$
$S_1 = 4.8$	$S_2 = 2.2$
$n_1 = 100$	$n_2 = 120$

μ_1 : population averge test score in program A
μ_2 : population averge test score in program B

   (b) What assumptions, if any, do you have to make about the parameter in (a) for the expected value of the difference of the sample means to be 0?

We assume that Program A and B have the same mean scores.
If $\mu_1 - \mu_2 = 0$,
then $\mu_1 = \mu_2$.

   (c) What is the standard error of the difference between the average test scores of the two programs? Round to 4 decimal places.

$$SE_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{(4.8)^2}{100} + \frac{(2.2)^2}{120}} = \boxed{0.5203}$$

   (d) Under the assumption that $\mu_A = \mu_B$, what is the probability that the sample difference in test scores between Program A and Program B is less than 1?

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE} = \frac{(87-86) - 0}{0.5203} = \frac{1}{0.5203} = 1.9220$$

$P(\bar{x}_1 - \bar{x}_2 < 1)$
$= P(t < 1.9220)$
$=$ pt(t, df)
$=$ pt(1.9220, 99)
$= \boxed{0.9713}$

(e) Construct a 90% confidence interval for the difference in average test scores between Program A and Program B and provide a **conclusion** in context. Be sure to use the confidence interval information to conclude about the original hypothesis.

$t^* = qt(90\% = \frac{1-0.90}{2} = \frac{0.10}{2} = 0.05$, $df = n_1 - 1 = 99)$
$= -1.660391$

$PE \pm t \cdot SE_{\bar{x}_1 - \bar{x}_2} = (\bar{x}_1 - \bar{x}_2) \pm t^* \cdot SE_{\bar{x}_1 - \bar{x}_2}$
$= 1 \pm (-1.660391)(0.5203)$
$= \boxed{(0.1361, 1.8639)}$

We are 90% confident the difference in average test score between Program A and B is captured within the interval.

4. Your favorite soccer team previously scored an average of 2.8 goals per game. However, the team recently lost its best player and you think that the average number of goals has changed. You took a sample of 38 games played after the team lost its best player and found a sample standard deviation of 0.46 and a sample mean of 2.6. Let μ be the true mean goals scored by your favorite team.

   (a) Create a 95% confidence interval of the new true mean ~~GPA~~ goals using the sample as described above.

   $\mu = 2.8 \quad \bar{x} = 2.6 \quad s = 0.46 \quad n = 38$

   $t^* = qt(95\% = \frac{1-0.95}{2} = 0.025,\ df = n-1 = 37)$

   $= -2.026192$

   $PE \pm t^* \frac{s}{\sqrt{n}} = 2.6 \pm (-2.026192)(\frac{0.46}{\sqrt{38}}) = \boxed{(2.4488, 2.7512)}$

   (b) Use your confidence interval from part (a) to test your claim and interpret your confidence interval.

   Since $\mu = 2.8$, it is not captured within the interval, we reject the claim at the $\alpha = 0.05$ level.
   We are 95% confident the true mean goals scored by your favorite team is between 2.4488 and 2.7512.

   (c) Your friend took a different sample of 16 games and found the average number of goals scored to be 2.4 with a standard deviation of 0.61, do they get a different conclusion to you when they construct their 95% confidence interval?

   $\mu = 2.8 \quad \bar{x} = 2.4 \quad s = 0.61 \quad n = 16$

   $t^* = qt(95\% = \frac{1-0.95}{2} = 0.025,\ df = n-1 = 15)$

   $= -2.13145$

   $PE \pm t^* \frac{s}{\sqrt{n}} = 2.4 \pm (-2.13145)(\frac{0.61}{\sqrt{16}}) = (2.0750, 2.7250)$

   Since $\mu = 2.8$, it is not captured within the interval, we reject the claim at the $\alpha = 0.05$ level.
   We are 95% confident the true mean goals scored by your favorite team is between 2.0750 and 2.7250.

Have seen		Haven't seen	
$\hat{p}_1$	0.81	$\hat{p}_0$	0.72
n_1	50	n_0	100

5. Your favorite movie director has just released a sequel to one of their movies. They claim that viewers will have the same likelihood of enjoying the new movie regardless of whether or not they have seen the original movie. To test this claim, you've surveyed a random sample of 50 people who have seen the original movie and 100 people who have not.

The results of your survey show that 72% of the people who hadn't seen the original movie enjoyed the new movie while 81% of the people who had seen the original movie enjoyed the new one. Based on these findings, can we reject the director's claim that the new movie will be equally enjoyable regardless of whether or not you have seen the original?

(a) What is the parameter of interest in this study?

$\hat{p}_1$: the proportion of peope who liked the new movie and have seen the original.
$\hat{p}_0$: the proportion of peope who liked the new movie and haven't seen the original.
The parameter of interest is the difference between the proportion of people who liked the new movie and have seen the original and the proportion of people who liked the new movie and haven't seen the original which is $\hat{p}_1 - \hat{p}_0$.

(b) Let $\hat{p}_1$ be the proportion of the sample of people who had seen the original movie that enjoyed the new one and let $\hat{p}_0$ be the proportion of people who had not seen the original movie that enjoyed the new one. Calculate a 90% confidence interval for $p_1 - p_0$. Give your answer to 4 decimal places.

$z^* = \text{qnorm}(90\% = \frac{1-0.90}{2} = 0.05) = -1.644584$

$SE_{\hat{p}_1-\hat{p}_0} = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_0\hat{q}_0}{n_0}} = \sqrt{\frac{(0.81)(0.19)}{50} + \frac{(0.72)(0.28)}{100}} = 0.07137226$

$PE \pm ME = (\hat{p}_1 - \hat{p}_0) \pm z^* SE_{\hat{p}_1-\hat{p}_0} = (0.81 - 0.72) \pm (-1.644584)(0.07137226)$
$= \boxed{(-0.0274, 0.2074)}$

We are 90% confident the difference between the true proportion of people haven seen the original movie that enjoyed the new one and people haven't seen the original movie that enjoyed the new one is captured within the interval.

(c) Now calculate a 90% confidence interval for $p_0 - p_1$. What stays the same and what changes from what you obtained in part (b)?

$z^* = -1.644584$
$SE_{\hat{p}_1-\hat{p}_0} = 0.07137226$

$PE \pm ME = (\hat{p}_0 - \hat{p}_1) \pm z^* SE_{\hat{p}_0-\hat{p}_1} = (0.72 - 0.81) \pm (-1.644584)(0.07137226)$
$= \boxed{(-0.2074, 0.0274)}$
The lower and upper bounds are swayed with their sign flipped.

(d) Based on the results are you able to state with 90% confidence that you can refute the director's claim that the two groups are equally likely to enjoy the movie?

Since $\hat{p}_d = 0$, it captured within the interval, we failed to reject the claim at the $\alpha = 0.1$ level

6. For the multiple choice questions, circle the best answer.

(a) Let $\bar{X}$ denote the sample mean. If the size of a randomly selected sample from a population is decreased from n = 100 to n = 25, then the standard deviation of $\bar{X}$ will

A. increase by a factor of 4.
B. decrease by a factor of 4.
C. increase by a factor of 2.
D. remain the same.

(b) Assuming the population standard deviation σ_X ***is known***, which of the following statements is true about $\sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}}$, i.e. the standard deviation of the sample mean $\bar{X}$?

A. It is a random variable.
B. It varies each time a new sample is drawn.
C. All of the above.
D. It decreases as the sample size n increases.

(c) When the population standard deviation σ_X ***is not known***, we must *estimate* the standard deviation of $\bar{X}$ using the sample standard deviation s_x. Which of the following is true about the standard *error* of $\bar{X}$, i.e. $\hat{\sigma}_{\bar{X}} = \dfrac{s_x}{\sqrt{n}}$?

A. *In general,* it decreases as the sample size n increases.
B. All of the options.
C. It is a random variable.
D. It varies each time a new sample is drawn.

(d) In which of the following scenarios is a paired design not appropriate?

A. Comparison of two different brands of hand cream, where each person tests one brand on one hand and one brand on the other hand.
B. Comparison of two different brands of hand cream, where each person tests only one brand.
C. Comparison of a cream and a spray in the treatment of a rash, where each person places the cream on one arm, and the spray on the other.
D. Comparison of a cream and an ointment in the treatment of a rash, where each person applies the ointment to half of a rash and the cream to the other half.

(e) We can make statistical inferences about populations even if we do not know the true population distribution because ...

A. Regardless of the true population distribution, all we need for a test statistic is the sample means and their standard deviations (or estimates of the standard deviations), and we can get these from our samples.
B. A sample size $\geq$ 30 guarantees Normality of the sample mean regardless of the true population distribution.
C. Regardless of the true population distribution, the Central Limit Theorem guarantees Normality of the sample mean as sample size goes to infinity.
D. Random sampling guarantees a representative sample regardless of the true population distribution.

(f) A random sample of 450 adults is taken to estimate a population proportion among US adults. If the researchers wanted to decrease the width of the confidence interval, they could

A. decrease the size of the population.
B. increase the size of the population.
C. decrease the size of the sample.
D. increase the size of the sample.