Midterm Revision 2

In Son Zeng

University of Michigan

insonz@umich.edu

November 19, 2018

Overview

1. First Section
   - Test Reminders
   - Videotaped Review and Solution

2. Second Section
   - Sample Questions

Maximum Likelihood Estimator

Step of deriving the Maximum Likelihood Estimator

1. Given $X_1, X_2, ...X_n$ are i.i.d. random variables, we first find the joint PDF (likelihood function), which is $f(x_1, x_2, ..., x_n; \lambda) = \prod_{i=1}^{n} f(x_i; \lambda)$.
2. Take log for the likelihood function we obtain the log-likelihood function: $log\Big(f(x_1, x_2, ..., x_n; \lambda)\Big)$
3. Then, we take the first derivative for the log-likelihood function with respect to the parameter of interest, and set to 0 (Maximizing): $\frac{\partial ln(f(x_1, x_2, ..., x_n; \lambda}{\partial \lambda} = 0$. The solution of the equation is the candidate MLE $\hat{\lambda}$.
4. We take the second derivative $\frac{\partial^2 ln(f(x_1, x_2, ..., x_n; \lambda}{\partial \lambda^2}$ and check whether it is less than 0. If yes, then the MLE is verified.

Invariance Principle of Maximum Likelihood Estimator

This principal did not appear in the homework and Professor's exam review. But during the class he said we do need to know the Invariance Principle of MLE. Check the lecture notes and class note 13 for detail.

1. Given $X_1, X_2, ...X_n$ are i.i.d. random variables. Suppose we have found that the MLE for λ is $\hat{\lambda}$, then for any one-to-one transformation, say $\lambda \to g(\lambda)$, we can instantly recognize that the MLE of $g(\lambda)$ is $g(\hat{\lambda})$.
2. Examples: For normal distribution $N(\mu, \sigma^2)$, we have found during the class that $\hat{\sigma}^2 = \frac{n-1}{n}S^2$, then since the domain of $\sigma^2 \in [0, \infty)$ and $\sigma \in [0, \infty)$, so $\hat{\sigma} = \sqrt{\frac{n-1}{n}S^2}$. (Here $g(x) = \sqrt{x}$)
3. Examples: Since the MLE for μ is $\bar{X}$, then by one-to-one transformation, the MLE for μ^5 is $\bar{X}^5$, the MLE for $\mu^{\frac{1}{3}}$ is $\bar{X}^{\frac{1}{3}}$. With the same logic, what is the MLE for μ^3 in the question?

Test Reminders

Let us be careful also about the following points during the exam 1:

1. You need to specify the probability correctly. For example, "What is the probability that the reaction time for a randomly selected person is greater than 1.00 second?" Then you have to answer $P(X > 1.00)$. If the question becomes greater than or equal to 1.00 second or at least 1.00 second, then you should answer $P(X \geq 1.00)$ to receive full credit.
2. To derive the standard deviation for the summation of independent random variables, say $X_1 + X_2 + \ldots\ldots + X_n$, please compute the variance first by formula $Var(X_1 + X_2 + \ldots\ldots + X_n) = Var(X_1) + \ldots\ldots + Var(X_n)$. Then, you can take the square root to obtain the standard deviation. Do not directly add the standard deviations.

Test Reminders

Steps of constructing confidence interval for population mean:

1. The first step is the define the parameter correctly! The parameter should be the **population/true mean** of the subject mentioned in the question.
2. The second step is to check **randomness** and Underlying population distribution (approximately) normal (**UPDN**) of the sample. First, we check whether the sample is random. If it is told, great! Just move on! If not, we can explain (most of the time) by your own reasoning why you think the collect sample is random or not.
3. Then, we check normality. If you are told UPDN, just proceed to next step. If not, then we check whether the sample size is large enough $n \geq 30$ to employ the CLT, which claims that the sampling distribution of the sample mean of the measurements is approximately normal. If the sample size is small $n < 30$, then we rely on the robustness of the t procedures against violations of normality.

Test Reminders

Steps of constructing confidence interval for population mean:

1. If σ is given, skip this part. Otherwise, compute $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$, which is the square root of the sample variance.
2. **Computation** Find the t-score or z-score by referencing the table according to the specified significance level. For two-tails (keyword: **between**,$\pm$) , we find $t_{\frac{\alpha}{2}}$ or $z_{\frac{\alpha}{2}}$; for one-tail (keyword: **no greater than, no less than**), we derive the confidence upper bound (CUB) or confidence lower bound (CLB) by finding t_α or z_α
3. **Conclusion** We are approximately $100(1-\alpha)\%$ confident that the population mean of .....what question say........... is (between /no greater than/no smaller than) ...the confidence interval....

Test Reminders

Steps of performing hypothesis testing for population mean:

1. The first step is the define the parameter correctly! The parameter should be the **population/true mean** of the subject.
2. The second step is the set up null hypothesis and alternative hypothesis. For two-tails, $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ and for one-tail, we either
3. The second step is to check **randomness** and Underlying population distribution (approximately) normal (**UPDN**) of the sample.

Test Reminders

Steps of performing hypothesis testing for population mean:

1. We specify the significance level α, the values are usually 0.05 or 0.01.
2. We calculate the t-score with df $n-1$ by $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$
3. We check the t-table for the corresponding p-value given t-score you derived. If the question is asking two-tails, you need to multiply the p-value by 2. (No multiplication for one-tail)
4. Conclusion: With a p-value lower/greater than the significant level (such as 0.05, 0.01), we reject/fail to reject H_0. There is/is not sufficient evidence to suggest that the population mean of ...what question say... is (different from/greater than/smaller than) ...the number....

Useful Links

Let us watch some videos to revise the assumptions for various statistical concepts before the exam:

1. Coming soon!

Extra Questions

Question 1

Let $X_1, X_2, ..., X_n$ be independent identically distributed random variables with common probability density function:

$$f(x_i;\theta) = (\theta+1)x_i^{\theta} \ 0 \leq \theta \leq 1$$

and 0, otherwise. We typically know that the expectation and variance of $X_1, ...X_n$ are identical.

a) First, compute $E(X_i)$ and $SD(X_i)$ for $i = 1, 2, ..., n$.

b) Find the Maximum likelihood estimator of θ

c) (Optional) Is the Maximum likelihood estimator of θ unbiased? Explain by showing the steps.

d) Find the Maximum likelihood estimator of θ^2.

Extra Questions

Question 1 Solution

a) First, compute $E(X_i)$ and $SD(X_i)$ for $i = 1, 2, ..., n$.
Sol: $E(X_i) = \int_0^1 x_i(\theta+1)x_i^\theta dx_i = \frac{\theta+1}{\theta+2}$, $E(X_i^2) = \int_0^1 x_i^2(\theta+1)x_i^\theta dx_i = \frac{\theta+1}{\theta+3}$,
then $SD(X_i) = \sqrt{E(X_i^2) - [E(X_i)]^2} = \sqrt{\frac{\theta+1}{\theta+3} - [\frac{\theta+1}{\theta+2}]^2}$

b) Find the Maximum likelihood estimator of θ
Sol: $f(x_1, ..., x_n; \theta) = (\theta+1)^n \cdot x_1^\theta \cdot ... \cdot x_n^\theta$, 0 otherwise. Take the log we get $logf(x_1, ..., x_n; \theta) = nlog(\theta+1) + \theta(x_1 + ... + x_n)$.
Take the first derivative we get $\frac{n}{\theta+1} + \sum_{i=1}^n logx_i = 0 \rightarrow \hat{\theta} = \frac{-n-\sum_{i=1}^n logx_i}{\sum_{i=1}^n logx_i}$.
To verify, we take the second derivative to get $-\frac{n}{(\theta+1)^2} < 0$ so that the MLE for θ is $\hat{\theta} = -\frac{n}{\sum_{i=1}^n logx_i} - 1$

Extra Questions

Question 1 Solution

c) (Optional) Is the Maximum likelihood estimator of θ unbiased? Explain by showing the steps.
Come to my office hour because it is more challenging.

d) Find the Maximum likelihood estimator of θ^2.
By the Invariance Principle of MLE, judging from the domain $0 \leq \theta \leq 1$, the transformation is one-to-one. Therefore, the MLE for θ^2 is just the square of the MLE of θ. Hence, $\hat{\theta}^2 = \frac{(-n - \sum_{i=1}^{n} log x_i)^2}{(\sum_{i=1}^{n} log x_i)^2}$

Extra Questions

Question 2

Recall that if $X_1, X_2, ..., X_n$ are independent and normally distributed random variables with mean μ and variance σ^2, that is, $X_i \sim N(\mu, \sigma^2)$.

a) Find the Maximum likelihood estimator for μ.
b) Find the Maximum likelihood estimator for μ^3.
c) Find the Maximum likelihood estimator for σ^2.
d) Let the Maximum likelihood estimator for σ^2 is $\hat{\sigma}^2$, find $E(\hat{\sigma}^2)$.
e) Is the Maximum likelihood estimator for σ^2 unbiased? If not, then what should be the bias?

Extra Questions

Question 2 Solution:

a) Find the Maximum likelihood estimator for μ.
c) Find the Maximum likelihood estimator for σ^2.
Sol: From the lecture notes, we have
$f(x_1, ..., x_n; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot exp\Big(- \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \Big)$. You can derive like the notes that $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$

b) Find the Maximum likelihood estimator for μ^3.
Sol: By the Invariance Principle, we take $g(x) = x^3$, so $\hat{\mu}^3 = (\bar{X})^3$

d) Let the Maximum likelihood estimator for σ^2 is $\hat{\sigma}^2$, find $E(\hat{\sigma}^2)$.
e) Is the Maximum likelihood estimator for σ^2 unbiased? If not, then what should be the bias? (Hint: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$)
Sol: $E(\hat{\sigma}^2) = E(\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2) = E(\frac{n-1}{n} S^2) = \frac{n-1}{n} \sigma^2$. Hence, the MLE of σ^2 is biased. The bias is $E(\hat{\sigma}^2) - \sigma^2 = -\frac{1}{n} \sigma^2$.

Extra Questions

Question 3

Suppose the nutrition label of apple cider says that apple cider contains an average concentration of 90 grams of sugar per liter, with standard deviation of 10 grams of sugar per liter. Now we want to check whether the claim is true. Therefore, we bought 50 liters of apple cider and found that 5050g sugar are contained in the 50 liters of apple cider in total.

Based on this information, could we perform a hypothesis testing to check whether the claim in the nutrition label is true, at significance level $\alpha = 0.05$? If yes, please perform the hypothesis testing based on the steps taught in class. If not , please provide reason detailing why we could not perform such a hypothesis testing.

Extra Questions

Question 3 Solution

1. Let μ be the population mean concentration of sugar contained in a liter of apple cider.
2. The null hypothesis is $H_0 : \mu = \mu_0 = 90$ grams versus $H_1 \neq 90$ grams.
3. Check randomness: the sample is not given random, but we believe that each liter of apple cider will not affect the other samples (independent).
4. Check UPDN: UPDN is not given. We check the sample size: $n = 50 \geq 30$, so we can use CLT and claim that the sample is approximately normal. We still use the t-test for the calculation given the degree of freedom $n - 1 = 49$.

Extra Questions

Question 3 Solution

1. We specify the significance level $\alpha = 0.05$. We also have known sample mean $\bar{X} = \frac{5050}{50} = 101$ grams/liter
2. The t-statistics is: t $= \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{101-90}{10/\sqrt{50}} = 7.778$
3. By $1 - pt(7.778, 49)$ in R (Check the table during the test), and multiply the p-value by 2, we get the p-value 4.2×10^{-10}
4. Conclusion: Since the p-value is lower than 0.05, we reject the null hypothesis. There is sufficient evidence to suggest that the population mean concentration of sugar contained in a liter of apple cider is different from 90 grams/liter.

Extra Questions

Question 4

Suppose the nutrition label of apple cider says that apple cider contains an average concentration of 90 grams of sugar per liter, but the standard deviation is not given. Now we bought 10 liters of apple ciders instead, and found the concentrations of sugar for each liter of apple cider as follows: 95.3, 101.2, 92.3, 90.1, 96.4, 99.2, 110.3, 103.4, 91.2, 89.9.

Based on this information, could we perform a hypothesis testing to check whether the claim in the nutrition label is true, at significance level $\alpha = 0.05$? If yes, please perform the hypothesis testing based on the steps taught in class. If not , please provide reason detailing why we could not perform such a hypothesis testing.

Extra Questions

Question 4 Solution

1. Again, let μ be the population mean concentration of sugar contained in a liter of apple cider.
2. The null hypothesis is $H_0 : \mu = \mu_0 = 90$ grams versus $H_1 \neq 90$ grams.
3. Check randomness: the sample is not given random, but we believe that each liter of apple cider will not affect the other samples (independent).
4. Check UPDN: UPDN is not given. We check the sample size: $n = 10 \leq 30$, so we cannot use CLT and claim that the sample is approximately normal. We need to rely on the robustness of t-test for the calculation given the degree of freedom $n - 1 = 9$.

Extra Questions

Question 4 Solution

1. We specify the significance level $\alpha = 0.05$. We also have known sample mean $\bar{X} = \frac{95.3+...+89.9}{10} = 96.93$ grams/liter and sample standard deviation $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} = 7.00$
2. The t-statistics is: $t = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{96.93-90}{7.00/\sqrt{10}} = 3.13$
3. By $1 - pt(3.13, 9)$ in R (Check the table during the test), and multiply the p-value by 2, we get the p-value 0.0121.
4. Conclusion: Since the p-value is lower than 0.05, we reject the null hypothesis. There is sufficient evidence to suggest that the population mean concentration of sugar contained in a liter of apple cider is different from 90 grams/liter.

Extra Questions

Question 5

Many people are concerned that the average grade for STATS 250 in Winter 2018 semester was significantly lower than the average grade in previous semesters. Particularly, the grade distributions (assume that the first column is the Winter 2018 data, and the second column is the sample mean over 30 semesters) are as follows:

Grade	Observed Counts	Expected Counts	X^2 Contribution
As	468.52	684.76	68.28631579
Bs	720.8	728.008	0.071366337
Cs	468.52	275.706	134.8437778
Ds	108.12	75.684	13.90114286
Es	36.04	36.04	0
Total	1802	1800.2	**217.1026028**

Extra Questions

Question 5 (Continue)

a) If the standard deviation of the number of students getting A's is 220, what is the 99 percent confidence interval of the true mean of the student getting the A's? (Since we have data from 30 semesters, the sample size should be 30 here)

b) What is the p-value of the 468.52 students getting A's in the Winter 2018 semester? What is the interpretation of such a p-value?

c) If the standard deviation of the number of students getting B's is 130, what is the 95 percent confidence interval of the true mean of the student getting the B's?

d) What is the p-value of the 720.8 students getting B's in the Winter 2018 semester? What is the interpretation of such a p-value?

e) Could we use the average grade of STATS 250 to make inference to the average grade of all classes in the University of Michigan? Why or why not?

Extra Questions

Question 5 Solution:

a) If the standard deviation of the number of students getting A's is 220, what is the 99 percent confidence interval of the true mean of the student getting the A's? (Since we have data from 30 semesters, the sample size should be 30 here)
b) What is the p-value of the 468.52 students getting A's in the Winter 2018 semester? What is the interpretation of such a p-value?

1. Let μ be the population mean number of students getting A's.
2. Check randomness: the sample is not given random, but we believe that each semester grade distribution will not affect the other samples (independent).
3. Check UPDN: UPDN is not given. We check the sample size: $n = 30 \geq 30$, so we can use CLT and claim that the sample is approximately normal. We still use the t-test for the calculation given the degree of freedom $n - 1 = 29$.

Extra Questions

Question 5 Solution:

1. From the technology $qt(0.995, 29) = 2.756$, we have the confidence interval
$\bar{x} \pm t_{\frac{\alpha}{2}, n-1} = 684.76 \pm 2.756 \cdot \frac{220}{\sqrt{30}} = 684.76 \pm 110.71 = (574.05, 795.47)$
2. Conclusion: Hence, we are approximately 99% confident that the population mean number of students getting A's is between 574.05 and 795.47 (approximately between 574 to 795 students).

Extra Questions

Question 5 Solution:

1. With the same setting, $t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{468.52-684.76}{220/\sqrt{30}} = -5.384$.
2. From technology, the corresponding p-value is 4.37×10^{-6} (lower-tail), which means that the probability of observing a semester with 468.52 or even less students getting A's in STATS 250 is 4.37×10^{-6}.
3. Alternatively, we can use two-tail, which gives p-value 8.74×10^{-6}, which means that the probability of observing a semester as extreme as, or more extreme than, 468.52 students getting A's in STATS 250 is 8.74×10^{-6}.

Extra Questions

Question 5 Solution:

c) If the standard deviation of the number of students getting B's is 130, what is the 95 percent confidence interval of the true mean of the student getting the B's?
d) What is the p-value of the 720.8 students getting B's in the Winter 2018 semester? What is the interpretation of such a p-value?

1. Let μ be the population mean number of students getting B's.
2. Check randomness: the sample is not given random, but we believe that each semester grade distribution will not affect the other samples (independent).
3. Check UPDN: UPDN is not given. We check the sample size: $n = 30 \geq 30$, so we can use CLT and claim that the sample is approximately normal. We still use the t-test for the calculation given the degree of freedom $n - 1 = 29$.

Extra Questions

Question 5 Solution:

1. From the technology $qt(0.975, 29) = 2.045$, we have the confidence interval $\bar{x} \pm t_{\frac{\alpha}{2},n-1} = 728.008 \pm 2.045 \cdot \frac{130}{\sqrt{30}} = 728.008 \pm 48.537 = (679.47, 776.55)$
2. Conclusion: Hence, we are approximately 95% confident that the population mean number of students getting B's is between 679.47 and 776.55 (approximately between 679 to 777 students).

Extra Questions

Question 5 Solution:

1. With the same setting, $t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{720.8-728.008}{130/\sqrt{30}} = -0.304$.
2. From technology, the corresponding p-value is 0.382 (lower-tail), which means that the probability of observing a semester with 728.008 or even less students getting B's in STATS 250 is 0.382.
3. Alternatively, we can use two-tail, which gives p-value 0.764, which means that the probability of observing a semester as extreme as, or more extreme than, 728.008 students getting B's in STATS 250 is 0.764.

Extra Questions

Question 5 Solution:

e) Could we use the average grade of STATS 250 to make inference to the average grade of all classes in the University of Michigan? Why or why not?

1. The students studying STATS 250 only represents a subset of population, which is all students in University of Michigan. Therefore, we cannot use the average grade of STATS 250 to make inference to the population mean grade of all class in the University of Michigan.

Extra Questions

Question 6

More questions are coming soon for CLT and distributions.
Good luck I tried my really best...

Go Wolverines!

Wolverines.png

The End