Fall 2025 AMACSS STAB57 Midterm Review Seminar

Sohil Chhabra and Xinyue Zhang

Nov 3, 2025

Recap of Week 2 & Week 3

- Estimating population parameter from sample observations
  - Method of Moment (MOM) Estimator:
    - Express lower order population moment(s) in terms of parameter(s)
    - Invert the expression (s) to express the parameter (s) in terms of the population moment (s)
    - Replace the population moment using sample moment
  - Maximum likelihood Estimation
    - Define log-likelihood function, $l(\theta) = \ln(L(\theta))$
    - Solve the equation, $\frac{dl(\theta)}{d\theta} = 0$ for θ . Assume θ' is the solution
    - Check $\left.\frac{d^2 l(\theta)}{d\theta^2}\right|_{\theta = \theta'} < 0$
- Properties of an Estimator
  - Sampling distribution of $\bar{X}$
  - Measuring quality of an estimator
  - Independence of Normal Variables

MOM Question 1

Let $X_1, X_2, X_3, \ldots\ldots, X_n \overset{\text{iid}}{\sim} \text{Uniform(a,b)}$. Find the Method of Moments Estimator for $a \; and \; b$

MOM Question 2

Let $X_1, X_2, X_3, \ldots\ldots, X_n$ be a random sample from Normal(μ, σ^2) distribution, with both parameters unknown. Find the Maximum Likelihood Estimators for μ, σ^2

MOM Question 3

Let $X_1, X_2, X_3, \ldots\ldots, X_n$ be a random sample from a population with pdf $f(x|\theta) = \frac{1}{\theta} x^{\frac{1-\theta}{\theta}}$, $0 < x < 1$, $0 < \theta < \infty$

Important Notes about Maximum Likelihood Estimator

- MLE is not unique
- MLE may not exist
- **Likelihood may not always be differential** - common identifier when range of X (random variable) involves θ (parameter) **
- **Invariance property of MLE :** Let $\hat{\theta}$ be the MLE of θ and $\mathrm{g}(\theta)$ be any 1-1 function of θ defined on Ω**,** then $g(\hat{\theta})$ is the MLE of $g(\theta)$**

** examples later

MLE Question 1

Let $X_1, X_2, X_3, \ldots\ldots, X_n$ be a random sample of size n from a distribution with probability density function

$$f(x, \alpha) = \alpha^{-2} x e^{-\frac{x}{\alpha}}, x > 0, \alpha > 0$$

a) Obtain the maximum likelihood estimator of α

b) Find the approximate distribution of the MLE that you derived in part(a). Provide the name and parameters of the distribution.

c) Given $g(x) = 2 + x^2$. Find the MLE of g(x) with appropriate justification.

MLE Question 2

- Let $X_1, X_2, X_3, \ldots\ldots, X_n$ be a random sample from a population with pdf $f(x|\theta) = \frac{2\,\theta^2}{x^3}, \theta \leq \mathrm{y} < \infty$

Find the maximum likelihood estimator of θ

Properties of an Estimator

Sampling distribution of $\overline{X}$

- $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ Then $\bar{X} \overset{\text{iid}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$ **, doesn't matter whether n is small or large.**
- **However, even when the random variables of the sample don't follow Normal distribution, by CLT, when** $n \to \infty, \overline{X_n} \xrightarrow{D} N\left(\mu, \frac{\sigma^2}{n}\right)$

Measuring quality of an estimator

- **Mean squared error,** $MSE(T) = E\left[(T-\theta)^2\right] = var[T] + (E(T)-\theta)^2 = \text{var}[\text{T}] + (\text{Bias}[\text{T}])^2$
- **Bias:** $Bias[T] = E[T] - \theta$
  - **When Bias = 0, T is unbiased estimator of** θ
  - **Unbiased estimator > biased estimator**
  - **Biased estimator with lower variance > biased estimator with higher variance -> Minimze MSE**
- **Sample variance,** $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ **- unbiased estimator of** σ^2 **(Good practice to prove it)**

- **cov[U,V] = 0 if and only if U and V are independent when U and V are different linear combinations of** $X_i's$ $given$ $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$

Question 1

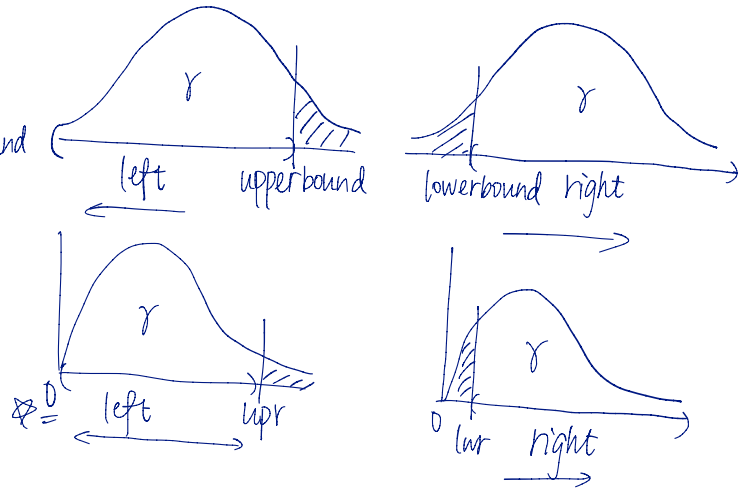
Let $Y_1, Y_2, \ldots . Y_n$ be a random sample from a population with mean μ and variance σ^2. Consider the estimator of $\mu' = \frac{X_1}{4} + \frac{X_2 + \cdots + X_n}{2(n-2)} + X_n/4$. Calculate the mean squared error for each estimator

Question 2

- Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$. Define $U = \sum_{i=1}^{n} X_i, \quad V = X_k - \bar{X}$. Show that Cov(U, V) = 0 for any k, and conclude that U and V are independent.

Recap of Confidence Interval

- Calculation of CIs
  - CI for parameters of Normal dist:
    - CI for μ (σ^2 known / unknown)
    - CI for σ^2
  - CI for population proportion (θ)
  - One-sided CIs:
    - Left-sided CI: upper bound
    - Right-sided CI: lower bound
- Relevant Definitions
  - Width of CI = upper bound − lower bound
  - Margin of Error (ME) = $\frac{1}{2}$ width
- Interpretation of CIs



a. Here, $n=30$, $\sum(x_i-\bar{x})^2=2450 \Rightarrow (n-1)s^2=2450$

For 90% CI for σ^2, $\chi^2_{0.05(df=29)}=17.70837$

$\chi^2_{0.95(df=29)}=42.55697$

90% CI for σ^2: $\left(\frac{(n-1)s^2}{\chi^2_{0.95(df=29)}}, \frac{(n-1)s^2}{\chi^2_{0.05(df=29)}}\right) \Rightarrow \left(\frac{2450}{42.55697}, \frac{2450}{17.70837}\right)$

$\Rightarrow (57.56988, 138.3527)$

CI & Test Hypothesis Question 1

An instructor is interested in the variation (σ^2) among the final scores for a particular course. She takes a random sample of 30 students and records their final scores (x_1, x_2, …, x_{30}). Assume all the observations are independently sampled from a $N(\mu, \sigma^2)$ distribution.

You are told that for this particular sample $\Sigma(x_i - \bar{x})^2 = 2450$

a. Construct a 90% CI for σ^2

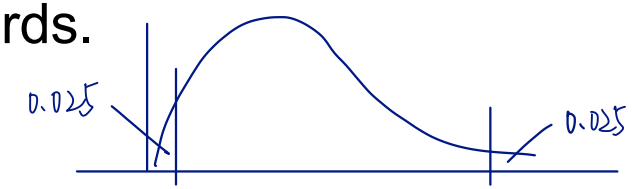
b. By constructing the rejection regions for $\alpha = 0.05$, test the hypothesis H_0: $\sigma^2 = 100$ vs. H_a: $\sigma^2 \neq 100$. Write down your conclusion in plain words.

b. Given, $\alpha = 0.05$

$H_0: \sigma^2 = 100$
$H_a: \sigma^2 \neq 100$
} two-sided test

test stat: $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$

$= \frac{2450}{100} = 24.5$



Rejection Region:
$(0, \chi^2_{0.025(df=29)}) \cup (\chi^2_{0.975(df=29)}, \infty)$

$= (0, 16.04767) \cup (45.72229, \infty)$

Since, test stat $\notin$ RR at $\alpha = 0.05$, we fail to reject H_0.

Conclusion: We do not have any evidence against the statement that the pop^n variance is 100.

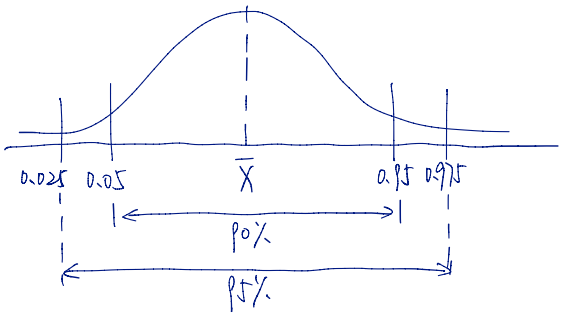
CI Question 2

Suppose we have a sample size of 50 from a $N(\mu, \sigma^2)$ where μ, σ^2 are unknown. Given that the 95% CI for μ is calculated to be (8.234, 10.484), calculate the 90% CI for μ.

$n = 50 \Rightarrow n-1 = 49$

Recall the formula of CI for μ (σ^2 unknown): $\bar{X} \pm t_{\frac{1+x}{2}(df=n-1)} \frac{S}{\sqrt{n}}$

Sample data not given $\Rightarrow$ cannot calculate $\bar{X}$ and S directly



$$\bar{X} = \frac{8.234 + 10.484}{2} = 9.359$$

$$ME(0.95\text{-CI}) = 10.484 - 9.359 = 1.125$$

$$= t_{\frac{1+0.95}{2}(df=49)} \frac{S}{\sqrt{n}}, \quad t_{0.975(df=49)} = 2.009575$$

$$\Rightarrow \frac{S}{\sqrt{n}} = \frac{ME}{t} = \frac{1.125}{2.009575} = 0.559782$$

$$ME(0.9\text{-CI}) = t_{\frac{1+0.9}{2}(df=49)} \frac{S}{\sqrt{n}}, \quad t_{0.95(df=49)} = 1.676551$$

$$= 1.676551 \times 0.559782 = 0.938587$$

$\Rightarrow$ 90%-CI for μ $= \bar{X} \pm ME(0.9\text{-CI}) = 9.359 \pm 0.938587$

$= (8.420433, 10.297587)$

CI Question 3

Suppose you have a population following a Normal distribution of mean μ and variance σ^2, and you wish to estimate an unknown μ. After drawing a sample of 5 (X_1, X_2, X_3, X_4, X_5), you are given 2 possible estimators for μ:

$$T_1 = (X_1 + X_2 + X_3 + X_4 + X_5)/5$$

$$T_2 = 0.1X_1 + 0.2X_2 + 0.4X_3 + 0.2X_4 + 0.1X_5$$

Without explicitly deriving the bounds, determine which estimator will result in a wider confidence interval and explain why.

Recall: width of CI ↑ if < confidence level γ ↑
standard deviation (either σ or s) ↑

Intuition: the estimator with higher standard deviation ⇒ wider CI

$\mathrm{Var}(T_1) = \mathrm{Var}\left(\frac{1}{5}(X_1 + \ldots + X_5)\right) = \left(\frac{1}{5}\right)^2 \mathrm{Var}(X_1) + \ldots + \left(\frac{1}{5}\right)^2 \mathrm{Var}(X_5) = \frac{5}{25}\sigma^2 = 0.2\sigma^2$

$\mathrm{Var}(T_2) = \mathrm{Var}(0.1X_1 + 0.2X_2 + 0.4X_3 + 0.2X_4 + 0.1X_5) = 0.1^2\mathrm{Var}(X_1) + 0.2^2\mathrm{Var}(X_2) + 0.4^2\mathrm{Var}(X_3) + 0.2^2\mathrm{Var}(X_4) + 0.1^2\mathrm{Var}(X_5)$

$= (0.01 + 0.04 + 0.16 + 0.04 + 0.01)\sigma^2 = 0.26\sigma^2$

$\Rightarrow SE(T_1) = \sqrt{\mathrm{Var}(T_1)} = 0.4472136\sigma$, $SE(T_2) = \sqrt{\mathrm{Var}(T_2)} = 0.5099020\sigma$ $\Rightarrow SE(T_2) > SE(T_1) \Rightarrow T_2$ has wider CI.

CI Question 4

Suppose an advertising company wants to estimate the proportion of people who have heard about their product through an advertisement. They want their 90% confidence interval (two-sided) to have a margin of error of ±0.07. They decide to randomly call some phone numbers and ask the question, “Have you heard about product XYZ through an advertisement?”

Assume that each person gives one independent response, every person contacted will answer the question (there are no missing values), and no person is contacted more than once.

In order to achieve the desired margin of error, at least how many people should be contacted (i.e., n = ?)?

CI for proportion: $\bar{X} \pm z_{\frac{1+0.9}{2}} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \Rightarrow ME = z_{\frac{1+0.9}{2}} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} = 0.07$

$\Rightarrow n = \bar{X}(1-\bar{X}) \left(\frac{z_{0.95}}{0.07}\right)^2$, $z_{0.95} = 1.644854$

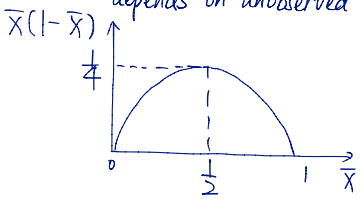
depends on unobserved $\bar{X}$, note that:

$f(\bar{X}) = \bar{X}(1-\bar{X})$

$\bar{X}$ is the proportion

$\Rightarrow 0 \le \bar{X} \le 1$

$\Rightarrow 0 \le f(\bar{X}) \le \frac{1}{4}$



we want the smallest n that can guarantee $\forall$ possible values of $\bar{X}$, $ME = 0.07$

use the conservative approach, set $f(\bar{X}) = \frac{1}{4}$

$\Rightarrow n = \frac{1}{4} \times 1082.151978 \ge \bar{X}(1-\bar{X}) \times 1082.151978$

$\Rightarrow n = 138.037994$ (n int, round up)

$n = 139$ ensures that $ME \le 0.07$

∴ At least 139 people should be contacted

$\Rightarrow$ smallest n we need is the max. value of RHS

↓ obtained when $\bar{X}(1-\bar{X}) = \frac{1}{4}$

$\Rightarrow$ at least $n = 138.238$ achieve the desired ME

test stat: $N(\mu, \sigma^2)$

μ (σ^2 known): $T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(\mu_0, \sigma^2)$

μ (σ^2 unknown): $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{(df=n-1)}$

σ^2: $T = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{(df=n-1)}$

① Null & Alternative hyp.

H_0

② T i) $H_0: \mu = \mu_0$ (σ^2 known)

ii) $H_0: \mu = \mu_0$ (σ^2 unknown)

iii) $H_0: \sigma^2 = \sigma_0^2$

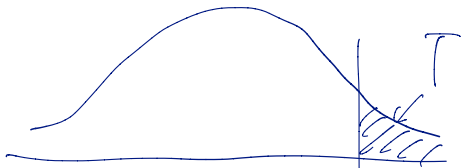
③ Rejection region → t values

Recap of Test Hypothesis

④ Determine whether T is in RR or not

⑤ reject H_0 / fail to reject

- Approaches to test hypotheses
  - Critical value approach (by calculating test stat)
  - p-value approach → p-value: smallest α that rejects H_0
    if p-value $< \alpha$ (predefined cutoff), then reject H_0
    requires CDF of a standard Normal dist/t dist.
- One-sided test
  - H_a: $\theta > \theta_0$ → right rejection region
  - H_a: $\theta < \theta_0$ → left rejection region
  - Different p-values for left / right rejection region
- Type-1, Type-2 error & Power of a test



$H_0: \theta = \theta_0$ $1-\alpha$

$H_a: \theta > \theta_0$

reject H_0, given H_0 true — fail to reject H_0, given H_0 false

$P[\text{Type-1 error}] = \alpha$

$P[\text{Type-2 error}] = \beta$

$\text{Power} = 1 - \beta$

Test Hypothesis Question 2

one-sided test

A company claims that the average amount of time customers spend on their website per visit is **more** than 8 minutes. To test their claim, they take a random sample of website visits and record the time spent on each visit (in minutes). The numbers are given below.

7.6, 6.0, 9.2, 8.4, 7.6, 7.2, 5.2, 6.1, 6.5, 6.3, 10.0, 8.2, 9.4, 8.3, 9.4

For this sample, $\Sigma x_i = 115.4$ and $\Sigma(x_i - \bar{x})^2 = 29.8$.

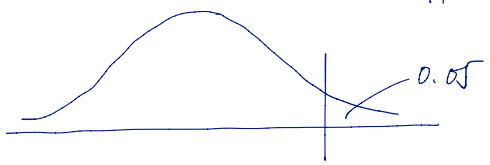
At 5% level of significance, by stating the null and alternative hypothesis, **test** their hypothesis and write a complete **conclusion**. [*hint:* take $\mu = 8$ as H_0]

Null hypothesis $H_0: \mu = 8$ vs. Alternative hypothesis $H_a: \mu > 8$

$n = 15$, $\bar{x} = \frac{1}{n}\Sigma x_i = \frac{1}{15} \times 115.4 = 7.69333$, $s^2 = \frac{1}{n-1}\Sigma(x_i - \bar{x})^2 = \frac{1}{14} \times 29.8 = 2.128571$

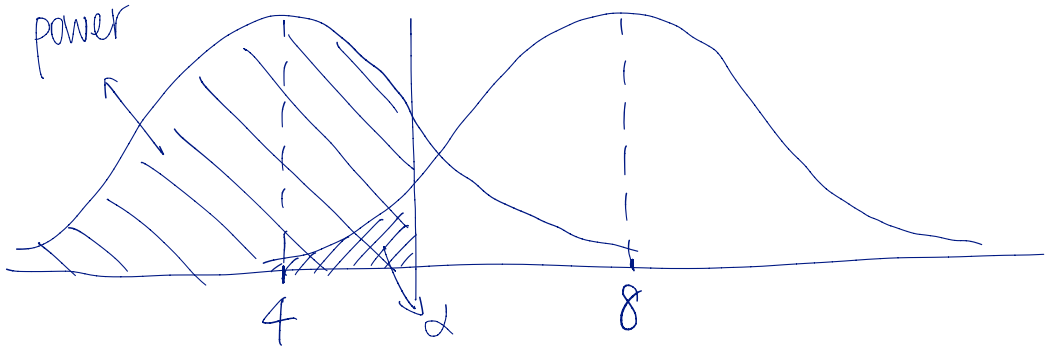
test stat: $\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1)$

$= \frac{7.69333 - 8}{\sqrt{2.128571/15}} = -0.81409$



Rejection Region:
$(t_{0.95(df=14)}, \infty) = (1.76131, \infty)$
Since, test stat $\notin$ RR at $\alpha = 0.05$, we fail to reject H_0.

Conclusion: We do not have sufficient evidence against the statement that the pop^n mean is 8.



[*unrelated to previous part*]

Suppose you want to test H_0: $\mu = 8$ vs. H_a: $\mu = 4$ based on 16 observations at 5% level of significance. The population standard deviation can be assumed to be 1.5. By deriving the rejection rule in terms of $\bar{X}$, calculate the power of this test. Leave your final answer in terms of *pnorm()* or the CDF of standard normal [i.e., $\Phi()$].

↗ $4<8$

$(-\infty, z_{0.05})$

Rejection Rule: ↗ RR for test against $\mu<8$

use critical value approach

reject H_0 if $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{0.05} \mid H_0$ is true

$\Rightarrow \bar{X} < \mu + z_{0.05}\,\sigma/\sqrt{n} \mid \mu = 8$

$\Rightarrow \bar{X} < 8 - 1.644854 \times 1.5/\sqrt{16}$

$\Rightarrow \bar{X} < 7.383$

Power:

↗ only one of H_0 & H_a is true, H_0 is false $\Rightarrow$ H_a is true $\Rightarrow \mu = 4$

$P[\text{reject } H_0 \mid H_0 \text{ is false}]$

$= P[\bar{X} < 7.383 \mid \mu = 4]$ → standardize $\bar{X}$ to use pnorm() / CDF Φ

$= P\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{7.383-4}{1.5/\sqrt{16}}\right]$

$\Rightarrow P[z < 9.021] = \Phi(9.021)$ or $pnorm(9.021)$

Acknowledgements

We would like to express our sincere gratitude to

Professor Shahriar Shams

for his continuous support, mentorship, and guidance

throughout the preparation and delivery of this seminar.

We also extend our heartfelt thanks to

Issey Sone

for his valuable feedback, insightful suggestions, and encouragement, which greatly enhanced the quality of our presentation.

Presented by: Sohil Chhabra and Xinyue Zhang

on behalf of the **Association of Mathematics & Computer Science Students (AMACSS)**