Mathematical Statistics Recitation 9

TA: Sonia Reilly

April 3, 2026

Midterm Review

Sampling and Confidence Intervals (Lec 7-8)

Unbiased estimates:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

χ_n^2 is the distribution of $\sum_{i=1}^n Z_i^2$, where $Z_i \sim N(0,1)$ i.i.d.
t_n is the distribution of $\frac{Z}{\sqrt{U/n}}$, where $Z \sim N(0,1)$, $U \sim \chi_n^2$.

	X_i's $\sim N(\mu, \sigma^2)$	X_i's not normal
$n \leq 30$	$\frac{\sqrt{n}(X_n - \mu)}{S_n} \sim t_{n-1}$	no general theorem
$n \geq 30$	$\frac{\sqrt{n}(X_n - \mu)}{S_n} \sim t_{n-1} \approx N(0,1)$ (approx equivalent)	$\frac{\sqrt{n}(X_n - \mu)}{S_n} \sim N(0,1)$ (approx)

$1-\alpha$ confidence interval (CI) for μ: (use $t_{n-1,\alpha/2}$ instead if needed)

$$\bar{X}_n - z_{\alpha/2}\frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\alpha/2}\frac{S_n}{\sqrt{n}}$$

Two types of sampling from finite population: (difference only matters if $n \gtrsim 0.05N$)

1. With replacement: independent draws, so simpler math
2. Without replacement: non-independent draws, $\mathrm{Var}(\bar{X}_n) < \frac{\sigma^2}{n}$, $\mathbb{E}[S^2] \neq \sigma^2$

Parameter Estimation (Lec 9-12)

e.g., $X_1, \ldots, X_n \sim \mathrm{Exp}(\lambda)$. What is λ?

1. **method of moments (MOM)**

$$\text{solve } E_\theta[X^k] = \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad \text{for } k \text{ up to } \dim(\theta)$$

2. **maximum likelihood estimation (MLE)**

$$\hat{\theta}_n = \arg\max_{\theta} \log L(\theta) = \arg\max_{\theta} \sum_i \log p_\theta(X_i)$$

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right)$$

Fisher information:

$$I(\theta) = -E_\theta\left[\frac{\partial^2}{\partial\theta^2} \log p_\theta(X)\right]$$

Cramér–Rao: any unbiased estimator has variance $\geq \frac{1}{nI(\theta)}$ (MLE asymptotically lowest)

3. **Bayesian perspective**

θ viewed as random variable (as opposed to fixed but unknown)

Bayes law: posterior $\propto$ prior $\times$ likelihood

$$\pi(\theta|X_1, \ldots, X_n) \propto \pi(\theta)\, f_\theta(X_1, \ldots, X_n)$$

estimators, e.g. posterior mode, mean, etc. generally converge to $\hat{\theta}_{MLE}$ as $n \to \infty$

Hypothesis Testing (Lec 13-15)

1. **Simple vs Simple**: $H_0 : \theta = \theta_0,\ H_1 : \theta = \theta_1$

$$\mathrm{LR} = \frac{L(\theta_0)}{L(\theta_1)} = \frac{p_0(X)}{p_1(X)}, \qquad \text{reject } H_0 \text{ if } \mathrm{LR} < c$$

Truth \ Test	H_0	H_1
H_0	Correct, $1-\alpha$	Type I error, α “level”
H_1	Type II error, β	Correct, $1-\beta$ “power”

tradeoff between $\alpha = P(\text{reject } H_0|H_0)$ and $\beta = P(\text{not reject } H_0|H_1)$

Neyman–Pearson lemma: likelihood ratio test makes the optimal tradeoff

2. **Simple vs Composite**: $H_0 : \theta = \theta_0,\ H_1 : \theta \in \Theta_1$

p-value = prob. of same or more extreme data given H_0

= smallest level at which we would reject

3. **Composite vs Composite**: $H_0 : \theta \in \Theta_0,\ H_1 : \theta \in \Theta_1$

Generalized LR test:

$$\mathrm{LR} = \frac{\max_{\theta\in\Theta_0} L(\theta)}{\max_{\theta\in\Theta} L(\theta)}, \qquad \text{reject } H_0 \text{ if } \mathrm{LR} < c$$

$$-2\log\mathrm{LR} \xrightarrow{d} \chi^2_d \quad \text{under } H_0, \qquad d = \dim(\Theta) - \dim(\Theta_0)$$

Pearson χ^2 test (specific case of LR):

$X_1, \dots, X_m$ are counts of # individuals taking on each categorical value

H_0: prob. of each category is p_i, H_1: not p_i

$$-2 \log \mathrm{LR} \approx \sum_{j=1}^{m} \frac{(X_j - np_j)^2}{np_j} \xrightarrow{d} \chi^2_{m-1}$$

Short Midterm Practice Problems

1. True or False?

   (a) In order to find the maximum likelihood estimator, it suffices to find the estimator that achieves zero gradient/derivative of the likelihood function.

   False – likelihood function may have minima where gradient is also zero.

   (b) The MLE estimator $\hat{\theta}$ has distribution $N(\theta, \frac{1}{nI(\theta)})$, where $I(\theta)$ is the Fisher information.

   False

   $$\hat{\theta}_{\mathrm{MLE}} \xrightarrow{d} N\left(\theta, \frac{1}{nI(\theta)}\right), \quad \text{only true in large } n \text{ limit}$$

   (c) A biased estimator can achieve lower than $\frac{1}{nI(\theta)}$ variance.

   True – Cramer–Rao bound only applies to unbiased

   (d) The method of moments estimator is consistent if the moments are a continuously invertible function of θ.

   True – proved this in class (generally we only care about consistent estimators)

   (e) In the Bayesian setting, the posterior mode converges to the MLE in the limit as the number of data points $n \to \infty$ for any choice of prior.

   False – consider choosing

   $$\pi(\theta) = \begin{cases} 1 & \text{if } \theta = 5 \\ 0 & \text{otherwise} \end{cases}$$

   Then the posterior mean, median, mode are 5, regardless of the data.

   (f) There is a tradeoff between the power of a test and its significance level.

   False – there is a tradeoff between β and α, not $1 - \beta$ and α

   (g) If a test rejects the null hypothesis with probability 0.5 regardless of the data, then the significance level is 0.5.

   True –

   $$\alpha = P(\text{test rejects } H_0 \mid H_0) = 0.5$$

   (h) For a one-sided Generalized LR test, e.g. $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$, we reject if $LR < c$. For a two-sided Generalized LR test, e.g. $H_1 : \mu \neq \mu_0$, we reject if $LR < -c$ or $LR > c$.

   False – always LR $< c$. If the test is two-sided, LR $< c$ will often be equivalent to, e.g., $|\bar{X} - \mu_0| > \varepsilon$, but this is already accounted for in the definition of LR.

   (i) As the number of data points $n \to \infty$, we always expect to see $-2 \log LR$ converge in distribution to a chi-squared distribution.

   False – only under H_0

2. (a) Suppose we measure the heights of 10 students in this class. What is an unbiased estimate of the population variance if the students are sampled with replacement? What if they're sampled without replacement?

   (b) Suppose we want a CI for the average height. (You may assume that height is a normally distributed variable, with unknown mean and variance). What CDF table would we look in?

   Solution:

   (a) With replacement, S_n^2, without replacement, $\frac{N-1}{N}S_n^2$.

   (b) A table for the t-distribution. $(\bar{X}-\mu)/(S/\sqrt{n})$ follows a t_{n-1}-distribution, and n is not large enough here to approximate it as $N(0,1)$.

3. Suppose we want to test if a disease has any effect on sleep. We collect data on N independently selected people, recording whether they have the disease, and whether their average length of sleep is < 5 hours, $5-7$ hours, $7-9$ hours, or > 9 hours. Our null hypothesis is that the probability of sleep length falling into each category is the same whether one has the disease or not, and the alternative hypothesis is that the probability is not the same. If we use the generalized LR test with the fact that $-2\log LR \to \chi_d^2$, what is d?

   Solution:

$$d = \dim(\Theta) - \dim(\Theta_0)$$

$$\Theta = \left\{ \begin{array}{l} p_{<5}^{\text{with}},\ p_{5-7}^{\text{with}},\ p_{7-9}^{\text{with}},\ 1-p_{<5}^{\text{with}}-p_{5-7}^{\text{with}}-p_{7-9}^{\text{with}}, \\ p_{<5}^{\text{w/out}},\ p_{5-7}^{\text{w/out}},\ p_{7-9}^{\text{w/out}},\ 1-p_{<5}^{\text{w/out}}-p_{5-7}^{\text{w/out}}-p_{7-9}^{\text{w/out}} \end{array} \right\}$$

$$\Rightarrow \dim(\Theta) = 6$$

$$\Theta_0 = \{p_{<5},\ p_{5-7},\ p_{7-9},\ 1-p_{<5}-p_{5-7}-p_{7-9}\} \quad \text{(same prob. w/ or w/out disease)}$$

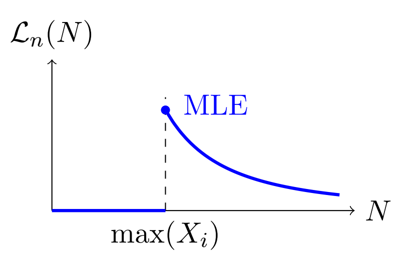
$$\Rightarrow \dim(\Theta_0) = 3$$

$$\Rightarrow d = 6-3 = 3$$

4. Suppose the data set $X_1, X_2, \ldots, X_n$ is drawn iid from Unif$([0,N])$. What is the likelihood function? Find the MLE of N.

   Solution:

$$L_n(N) = \begin{cases} \left(\frac{1}{N}\right)^n & N \geq \max\{X_1,\ldots,X_n\} \\ 0 & \text{otherwise} \end{cases}$$



$$\hat{N}_{\text{MLE}} = \arg\max_N L_n(N) = \max\{X_1,\ldots,X_n\}$$

From Practice Midterms:

1. [2025 Problem 1(c)] True or False? Let $X_1, \ldots, X_n \sim \text{Multi}(n, \bar{p})$ follow a multinomial distribution with probability vector $\bar{p} = (p_1, \ldots, p_m)$. Let $H_0 : \bar{p} \in \text{span}\{\bar{p}_0, \bar{p}_0'\}$ and $H_1 : \bar{p} \notin \text{span}\{\bar{p}_0, \bar{p}_0'\}$. Then $-2 \log LR \xrightarrow{d} \chi^2_{m-1}$ as $n \to \infty$.

   Solution: False. $\dim \Theta_0 = 1$ and $\dim(\Theta_1 \cup \Theta_0) = m - 1$, so $-2 \log LR \xrightarrow{d} \chi^2_{m-2}$.

2. [2025 Problem 5] Carol studies whether the "Big Three" (Djokovic, Federer, and Nadal) are equally dominant in their rivalries. She collects the following match data:

Rivalry	Number of matches	Number of wins
Djokovic–Federer	50	27
Djokovic–Nadal	60	31
Federer–Nadal	40	24

   Assume each row follows an independent binomial distribution: $X_{D-F} \sim B(n = 50, p_{D-F})$, $X_{D-N} \sim B(n = 60, p_{D-N})$, and $X_{F-N} \sim B(n = 40, p_{F-N})$, with unknown probabilities $(p_{D-F}, p_{D-N}, p_{F-N})$. Carol tests:

$$H_0 : p_{D-F} = p_{D-N} = p_{F-N} = 0.5 \text{ v.s. } H_1 : \text{not } H_0.$$

   Carry out the generalized likelihood ratio test, by writing down the expression for the test statistic $-2 \log LR$, and identifying its approximate null distribution. You may use without proof that the MLE of p for a binomial observation $X \sim B(n, p)$ is $\hat{p} = X/n$. No numerical computation is needed.

   Solution: Let $L(p_{D-F}, p_{D-N}, p_{F-N})$ be the likelihood function, expressed as

$$L(p_{D-F}, p_{D-N}, p_{F-N}) = \binom{50}{27} p_{D-F}^{27}(1-p_{D-F})^{23} \cdot \binom{60}{31} p_{D-N}^{31}(1-p_{D-N})^{29} \cdot \binom{40}{24} p_{F-N}^{24}(1-p_{F-N})^{16}.$$

   Under H_0, $(p_{D-F}, p_{D-N}, p_{F-N}) = (0.5, 0.5, 0.5)$. Under H_1, the MLE for $(p_{D-F}, p_{D-N}, p_{F-N})$ is $(\frac{27}{50}, \frac{31}{60}, \frac{24}{40})$. Therefore, the test statistic is

$$\begin{aligned}
-2 \log LR &= 2 \log L\left(\frac{27}{50}, \frac{31}{60}, \frac{24}{40}\right) - 2 \log L(0.5, 0.5, 0.5) \\
&= 2\left(27 \log \frac{27}{50} + 23 \log \frac{23}{50} + 31 \log \frac{31}{60} + 29 \log \frac{29}{60} + 24 \log \frac{24}{40} + 16 \log \frac{16}{40}\right) + 300 \log 2.
\end{aligned}$$

   As $\dim(H_0) = 0$ and $\dim(H_0 \cup H_1) = 3$, the approximate null distribution of $-2 \log LR$ is χ^2_3. (Numerical calculation shows that $-2 \log LR = 1.998$, with a p-value of 0.573.)

3. [2025 Problem 3] Based on n i.i.d. observations $X_1, \ldots, X_n$, and under an assumed statistical model, Alice computes that the log-likelihood function $\ell_n(\theta)$ is approximately

$$\frac{\ell_n(\theta)}{n} \approx 1 - 2(\theta - 3)^2.$$

   Based on this expression, write down the approximate value of the MLE $\hat{\theta}_n$, and the approximate distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$. (Hint: how is $\ell_n''(\theta)/n$ related to the Fisher information $I(\theta)$?)

Solution: Since the MLE maximizes the log-likelihood $\ell_n(\theta)$, we have $\hat{\theta}_n = 3$. Moreover, since

$$\frac{\ell_n''(\theta)}{n} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log f_\theta(X_i)}{\partial \theta^2} \xrightarrow{p} E\left[\frac{\partial^2 \log f_\theta(X)}{\partial \theta^2}\right] = -I(\theta)$$

by LLN, and the second derivative of our approximation is $\frac{d^2}{d\theta^2}(1 - 2(\theta - 3)^2) = -4$, we compute that $I(\theta) \approx 4$. Finally, since the asymptotic variance of the MLE is characterized by the inverse Fisher information $I(\theta)^{-1}$, the approximate distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ is $N(0, 1/4)$.