SECTION 8

Slides by Claris Winston

ADMINISTRIVIA

- Hw 6 was due yesterday
- Hw 7 due on August 18th
- Two part final exam next week on 08/20 and 08/21, both at 12:00 in HRC 155 (the lecture room)

Content Review

Likelihood

- **Likelihood**: Let $x_1, \dots x_n$ be iid realizations from probability mass function $p_X(\mathrm{x}\,; \theta)$ (if X discrete) or density $f_X(\mathrm{x}\,; \theta)$ (if X continuous), where θ is a parameter (or a vector of parameters). We define the likelihood function to be the probability of seeing the data (in the discrete case). If X is discrete:

$$L(x_1, \dots, x_n\,; \theta) = \prod_{i=1}^{n} p_X(x_i\,; \theta)$$

If X is continuous:

$$L(x_1, \dots, x_n\,; \theta) = \prod_{i=1}^{n} f_X(x_i\,; \theta)$$

MAXIMUM LIKELIHOOD ESTIMATOR (MLE)

- **Likelihood**: Let $x_1, \dots x_n$ be iid realizations from probability mass function $p_X(\mathrm{x}\,; \theta)$ (if X discrete) or density $f_X(\mathrm{x}\,; \theta)$ (if X continuous), where θ is a parameter (or a vector of parameters). We define the likelihood function to be the probability of seeing the data (in the discrete case). If X is discrete:

$$L(x_1, \dots, x_n\,; \theta) = \prod_{i=1}^{n} p_X(x_i\,; \theta)$$

If X is continuous:

$$L(x_1, \dots, x_n\,; \theta) = \prod_{i=1}^{n} f_X(x_i\,; \theta)$$

MLE: Find value $\hat{\theta}$ which maximizes L.

UNBIASED ESTIMATOR

- **Bias**: The bias of an estimator $\hat{\theta}$ for a true parameter θ is defined as

$$\text{Bias}(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}(X_1, \ldots, X_n)] - \theta.$$

Here, $\hat{\theta}$ is treated as a *random variable*, computed from i.i.d. random variables $X_1, \ldots .X_n$.

- **Unbiased Estimator**: An estimator $\hat{\theta}$ of θ is unbiased iff $\text{Bias}(\hat{\theta}, \theta) = 0$, or equivalently $\mathbb{E}[\hat{\theta}(X_1, \ldots, X_n)] = \theta$.

TASKS

1D

d) **True or False**: We get a different answer if we maximize log-likelihood rather than likelihood, but because it is close enough and easier to compute, we use it for our estimate of θ.

1D

d) **True or False**: We get a different answer if we maximize log-likelihood rather than likelihood, but because it is close enough and easier to compute, we use it for our estimate of θ.

False! The maximal parameters are the same!

1E

e) **True or False:** [Notation question] $\hat{\theta}$ is the true parameter and θ is our estimate.

1E

e) **True or False:** [Notation question] $\hat{\theta}$ is the true parameter and θ is our estimate.

False

1F

f) **True or False**: An estimator is unbiased if $\text{Bias}(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}(X_1, \ldots, X_n)] - \theta = 0$ or equivalently $\mathbb{E}[\hat{\theta}(X_1, \ldots, X_n)] = \theta$

1F

f) **True or False**: An estimator is unbiased if $\text{Bias}(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}(X_1, \ldots, X_n)] - \theta = 0$ or equivalently $\mathbb{E}[\hat{\theta}(X_1, \ldots, X_n)] = \theta$

True

1G

g) You flip a coin 10 times and observe HHHTHHTHHH (8 heads, 2 tails). What is the MLE of θ, where θ is the true probability of seeing tails?

- ○ $\hat{\theta} = .2$
- ○ $\hat{\theta} = .25$
- ○ $\hat{\theta} = .8$
- ○ $\hat{\theta} = .3$

1H

h) True or false: the Union Bound always gives a result in $[0, 1]$.

1H

h) True or false: the Union Bound always gives a result in [0, 1].

False. Consider X and Y, which are independent indicator random variables. Suppose $p_X(x) = \begin{cases} 0.75 & x = 0 \\ 0.25 & x = 1 \end{cases}$ and $p_Y(y) = \begin{cases} 0.75 & y = 0 \\ 0.25 & y = 1 \end{cases}$. Then we may apply the Union Bound to place a bound on $P(X = 0 \cup Y = 0)$:

$$P(X = 0 \cup Y = 0) \leq P(X = 0) + P(Y = 0) = 0.75 + 0.75 = 1.5.$$

In these cases, the Union Bound tells us very little, since the probability of any event occurring is at most 1.

2B

b) True or false: Markov's Inequality always gives a non-negative result.

2B

b) True or false: Markov's Inequality always gives a non-negative result.

True. Markov's Inequality is

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}$$

as long as X is a non-negative random variable and $\alpha > 0$. Since X is a non-negative random variable, $\mathbb{E}[X] \geq 0$, so $\frac{\mathbb{E}[X]}{\alpha} \geq 0$.

2c

c) True or false: Chebyshev's Inequality can best be described as giving an upper bound on the distribution's right tail.

2c

c) True or false: Chebyshev's Inequality can best be described as giving an upper bound on the distribution's right tail.

False. Chebyshev's Inequality gives an upper bound on the sum of the probabilities of the left and right tails of the distribution.

15 - TAIL BOUNDS

Suppose $X \sim \mathsf{Binomial}(6, 0.4)$. We will bound $\mathbb{P}(X \geq 4)$ using the tail bounds we've learned, and compare this to the true result.

a) Give an upper bound for this probability using Markov's inequality. Why can we use Markov's inequality?

15 - TAIL BOUNDS

Suppose $X \sim \text{Binomial}(6, 0.4)$. We will bound $\mathbb{P}(X \geq 4)$ using the tail bounds we've learned, and compare this to the true result.

b) Give an upper bound for this probability using Chebyshev's inequality. You may have to rearrange algebraically and it may result in a weaker bound.

15 - Tail Bounds

Suppose $X \sim \mathsf{Binomial}(6, 0.4)$. We will bound $\mathbb{P}(X \geq 4)$ using the tail bounds we've learned, and compare this to the true result.

c) Give an upper bound for this probability using the Chernoff bound.

17 - Robbie's Late!

Suppose the probability Robbie is late to teaching lecture on a given day is at most 0.01. Do not make any independence assumptions.

a) Use a Union Bound to bound the probability that Robbie is late at least once over a 30-lecture quarter.

17 - Robbie's Late!

Suppose the probability Robbie is late to teaching lecture on a given day is at most 0.01. Do not make any independence assumptions.

b) Use a Union Bound to bound the probability that Robbie is **never** late over a 30-lecture quarter.

17 - Robbie's Late!

Suppose the probability Robbie is late to teaching lecture on a given day is at most 0.01. Do not make any independence assumptions.

c) Use a Union Bound to bound the probability that Robbie is late at least once over a 120-lecture quarter.

8 - A Red Poisson

Suppose that $x_1, \ldots, x_n$ are i.i.d. samples from a Poisson(θ) random variable, where θ is unknown. Find the MLE for θ.

8 - A Red Poisson

Poisson PMF:

$$p_X(x) = e^{-\theta} \cdot \frac{\theta^x}{x!}$$

Suppose that $x_1, \ldots, x_n$ are i.i.d. samples from a Poisson(θ) random variable, where θ is unknown. Find the MLE for θ.

8 - A Red Poisson

Poisson PMF:

$$p_X(x) = e^{-\theta} \cdot \frac{\theta^x}{x!}$$

Suppose that $x_1, \ldots, x_n$ are i.i.d. samples from a Poisson(θ) random variable, where θ is unknown. Find the MLE for θ.

Step 1: Write down the likelihood function L, then the log-likelihood function $\ln L$

8 - A Red Poisson

Suppose that $x_1, \ldots, x_n$ are i.i.d. samples from a Poisson(θ) random variable, where θ is unknown. Find the MLE for θ.

Step 1: Write down the likelihood function L, then the log-likelihood function $\ln L$

Step 2: Compute the derivative $\frac{\partial}{\partial \theta} \ln L$

8 - A Red Poisson

Suppose that $x_1, \ldots, x_n$ are i.i.d. samples from a Poisson(θ) random variable, where θ is unknown. Find the MLE for θ.

Step 1: Write down the likelihood function L, then the log-likelihood function $\ln L$

Step 2: Compute the derivative $\frac{\partial}{\partial \theta} \ln L$

Step 3: Find $\widehat{\theta}$ which zeroizes the derivative

8 - A Red Poisson

Suppose that $x_1, \dots, x_n$ are i.i.d. samples from a Poisson(θ) random variable, where θ is unknown. Find the MLE for θ.

Step 1: Write down the likelihood function L, then the log-likelihood function $\ln L$

Step 2: Compute the derivative $\frac{\partial}{\partial\theta} \ln L$

Step 3: Find $\hat{\theta}$ which zeroizes the derivative

Step 4 (Omitted for 312): Check second derivatives and boundaries.

11 - BIRD WATCHING

You are an ornithologist studying a rare species of birds in a nature reserve. Over a period of 50 days, you record the number of sightings of this bird (you see $x_1, x_2, ..., x_{50}$ birds on each day). Your research has shown that the number of sightings of this species depends on the average number of monkeys in the reserve, θ_1, and the average number of eagles in the reserve, θ_2. After years of studying this rare species in other environments, you've found that the number of birds observed on a particular day follows the following distribution:

$$p_X(k) = \frac{1}{k!}(\theta_1^k \cdot e^{-\theta_1} \cdot \theta_2^k \cdot e^{-3\theta_2})$$

Find the MLE for θ_1 and θ_2 (i.e., find $\hat{\theta_1}$ and $\hat{\theta_2}$).

a) What is the likelihood function?

b) What is the log-likelihood function?

c) We want to find values of θ_1 and θ_2 that maximize the likelihood function. To do this, we will take the partial derivative with respect to each of these parameters and solve for the values that make them both zero. First, take the partial derivative of the likelihood function with respect to θ_1.

d) Now, take the partial derivative with respect to θ_2.

e) Set both these partial derivatives to 0, and solve for $\hat{\theta_1}$ and $\hat{\theta_2}$.

12 - A Biased Estimator

In class, we showed that the maximum likelihood estimate of the variance θ_2 of a normal distribution (when both the true mean μ and true variance σ^2 are unknown) is what's called the *population variance*. That is

$$\hat{\theta}_2 = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\theta}_1)^2\right)$$

where $\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the MLE of the mean. Is $\hat{\theta}_2$ unbiased?

Final Review

The puzzle game *Connections* shows the user 16 words, which should be grouped into 4 groups of 4. There is exactly one correct answer (i.e., each of the 16 words belongs in exactly one group of 4). The user selects 4 items they think form a group. If a correct group of 4 is selected, that entire group is removed from the puzzle; if it is not exactly a group of 4, then all items remain in the puzzle.

You get two groups immediately, so 8 items (2 groups of 4) remain. You wish to know the expected number of **incorrect** guesses required until you guess all the groups correctly.

(a) If you do the puzzle while **focused**, you will identify 5 words, of which 4 form a group. You will guess a new set of 4 among the 5 words until you find the group (with each not-yet-guessed set equally likely). What is the expected number of incorrect guesses until you get the group of 4, assuming you are focused? [4 points] (Hint: our answer doesn't need combinations)

Final Review

The puzzle game *Connections* shows the user 16 words, which should be grouped into 4 groups of 4. There is exactly one correct answer (i.e., each of the 16 words belongs in exactly one group of 4). The user selects 4 items they think form a group. If a correct group of 4 is selected, that entire group is removed from the puzzle; if it is not exactly a group of 4, then all items remain in the puzzle.

You get two groups immediately, so 8 items (2 groups of 4) remain. You wish to know the expected number of **incorrect** guesses required until you guess all the groups correctly.

(b) If you do the puzzle while **distracted**, you will choose a set of 4 uniformly at random from among the 8 remaining. Since you are distracted, each guess is independent (and therefore you might repeat a previous guess). What is the expected number of incorrect guesses, assuming you are distracted? [4 points]

Final Review

The puzzle game *Connections* shows the user 16 words, which should be grouped into 4 groups of 4. There is exactly one correct answer (i.e., each of the 16 words belongs in exactly one group of 4). The user selects 4 items they think form a group. If a correct group of 4 is selected, that entire group is removed from the puzzle; if it is not exactly a group of 4, then all items remain in the puzzle.

You get two groups immediately, so 8 items (2 groups of 4) remain. You wish to know the expected number of **incorrect** guesses required until you guess all the groups correctly.

(c) Let a be your answer from (a) and b be your answer from (b), and let p be the probability you are focused (all other times you are distracted). Give a formula for the average number of **incorrect** guesses in terms of a, b, p. [2 points]

(d) Next you play the *Wordle* (a word-guessing game), with starting word ADIEU.

When the correct answer is **vowel-heavy**, you do well: half the time you get a score of 4 or better; the other half the time you get a score of 5 or worse.

When the correct answer is not **vowel-heavy**, you don't do as well: $1/10$ of the time you get a score of 4 or better; the other $9/10$ of the time you get a score of 5 or worse.

You know that $1/3$ of words are vowel-heavy.

- Let V be the event "the correct answer is vowel-heavy."
- Let $\overline{V}$ be the event "the correct answer is not vowel-heavy."
- Let E be the event you got a score of 4 or better.

(i) You wish to find the probability that the correct answer is vowel heavy, assuming that you got a score of 4 or better. Fill in the blank below for the expression you're seeking, using the notation above. [2 points]

$$\mathbb{P}(\underline{\hspace{5cm}})$$

(ii) Compute the expression from the last part (i.e., find the probability the correct answer is vowel-heavy, assuming you got a score of 4 or better). [4 points]