SPRING 2019

CE 311S : PROBABILITY AND STATISTICS

Discussion session

M 1:00 – 2:00 PM

PRIYADARSHAN PATIL
Teaching Assistant, The University of Texas at Austin

Welcome back

Hope y’all had a great spring break!

Discussion overview

- Administrative stuff
- Question for the week
- Review of special continuous distributions
- Example questions for special distributions (work in groups of 2-3 on these)
- Introduction to joint random variables

Administrative Stuff

- Homework 3 due Friday 11:59 PM
- Online assignment 4 due next Friday 11:59PM

Week 7: Question

I ask all college students what their average class size is, and compute the mean. It turns out to be 56.

I go to the administration, and ask what the average class size is. They respond 31.

Can both these statements be true together? (Same semester, same college, same circumstances, etc.)

Week 7: Question

Please turn in your answer (Yes/No) with your wager and name.

Current standings: Link

Week 7: Answer

- Yes, both statements are true together.
- Large classes are oversampled, thus, their pmf values are skewed higher.
- In general, if the class size is x, it will be overrepresented in the sample by a factor of x.

Week 7: Answer

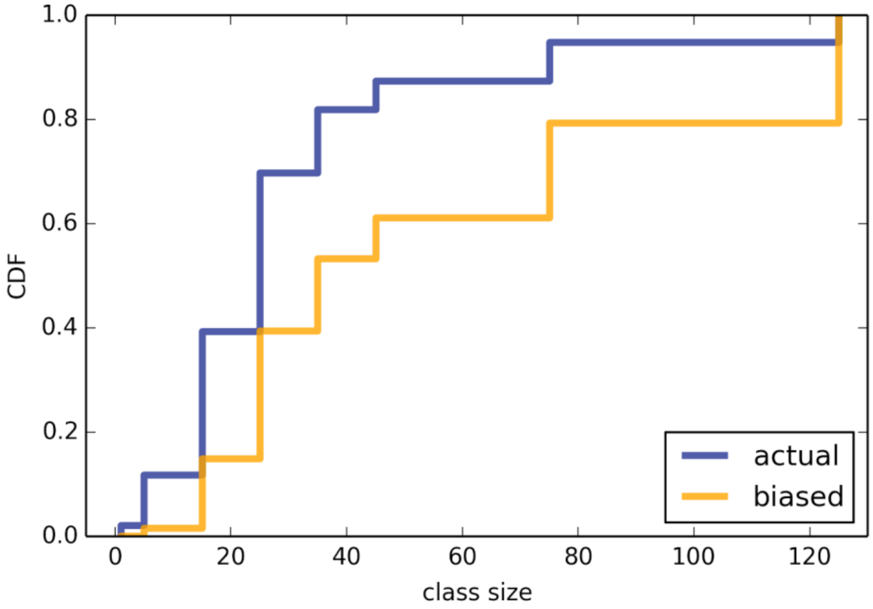

Figure 1: Undergraduate class sizes at Purdue University, 2013-14 academic year: actual distribution and biased view as seen by students.

Week 7: Answer

- The same effect applies to passenger planes. Airlines complain that they are losing money because so many flights are nearly empty. At the same time passengers complain that flying is miserable because planes are too full. They could both be right. When a flight is nearly empty, only a few passengers enjoy the extra space. But when a flight is full, many passengers feel the crunch.
- Also observed in many other places

Any questions so far?

- About this specific problem, course material covered so far, etc.

Distribution visualization

- Distribution visualization: http://students.brown.edu/seeing-theory/probability-distributions/index.html#section2

Name of special distribution (X)	Properties	Range (R_X) and Parameters	PDF (it is not probability X=x) $f(x)$	CDF $F(x) = P(X \leq x)$	Expected value $E(X)$	Variance $Var(X)$
Uniform distribution	○ If all possible values of X are equally likely to occur ○ And, all values fall in the range $[a, b]$	R_X: $[a, b]$ Parameters: a, b	$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & otherwise \end{cases}$	$f(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x \geq b \end{cases}$	$(b+a)/2$	$\frac{(b-a)^2}{12}$
Exponential distribution	○ X represents the waiting time between Poisson distribution events ○ X has property of being memoryless which means $P(X \geq x + x_0 \mid X \geq x_0) = P(X \geq x)$	R_X: $[0, \infty)$ Parameters: λ	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases}$	$F(x) = 1 = e^{\{-\lambda x\}}$	$1/\lambda$	$\frac{1}{\lambda^2}$
Normal distribution	○ When the pdf of X has a bell curve shape	R_X: $(-\infty, \infty)$ Parameters: μ, σ	$f(x) = \frac{1}{\sqrt{\{2\pi\}}\sigma} e^{-(x-\mu)^2/2\sigma^2}$	Use z-table or R	μ	σ^2
Gamma distribution	○ General case of Exponential distribution	R_X: $[0, \infty)$ Parameters: α, λ	$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases}$	Use R	α/λ	$\frac{\alpha}{\lambda^2}$

We will learn about two more continuous distributions later in the course: t-distribution and chi-squared distribution

$\Gamma(.)$ is the Gamma function with properties: (a) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ (b) $\Gamma(n) = (n-1)!$ if n is positive integer (c) $\Gamma(1/2) = \sqrt{\pi}$

Red River Showdown

- You are part of the research team for analyzing the Red river rivalry
- You learn from the historical analysis of the past games against OU that:
  - Longhorns score touchdowns at **rate of 1 touchdown every 15 min of game time**
- Assuming that the scoring of goals is a Poisson process, What is the probability that there will be **no touchdowns in the first quarter**?
- Ans: **0.368**

Red River Showdown

- Watching the 2017 rivalry game, you observe that Longhorns are playing slow and have scored **no touchdown in the first 9 minutes**.
- What is the probability they will **score a touchdown in the next 6 minutes** before the quarter whistle is blown? (Hint: Exponential distribution is memoryless)
- Ans: **0.3297**
- Find the **mean and median** of time taken by longhorns to score a touchdown
    - Mean = **15 minutes**
    - Median= **10.397 minutes**

Red River Showdown

- Discouraged by the poor performance by the first half, you start analyzing how much will Longhorns lose by against OU
- You define $X = OU\ score - Longhorns\ score$
- Even though X is discrete in nature, using central limit theorem the historic data tells you that on an average X is **distributed normally** with **mean of -1** and **standard deviation of 8**.
- What is the probability that Longhorns will not lose this game?
- Ans: **0.550**
- What is the probability that the scores of two teams will differ by less than 3?
- Ans: **0.290**

Thank you

Any Questions?