SPRING 2019

TEXAS
The University of Texas at Austin

CE 311S : PROBABILITY AND STATISTICS

Discussion session

M 1:00 – 2:00 PM

PRIYADARSHAN PATIL
Teaching Assistant, The University of Texas at Austin

Discussion overview

- Administrative stuff
- Question for the week
- Review of joint random variables
- Review of correlation and covariance
- (If time permits) LLN and CLT

Administrative Stuff

- Online assignment 5 due Friday 11:59PM
- Midterm 2 next Friday in class (9-10 AM)
- HW3 has been graded and solutions will be posted soon

Week 8: Question

Abraham is tasked with reviewing damaged planes coming back from sorties over Germany in the Second World War. He has to review the damage of the planes to see which areas must be protected even more.

Abraham finds that the fuselage and fuel system of returned planes are much more likely to be damaged by bullets or flak than the engines. What should he recommend to his superiors?

Week 8: Question

Please turn in your answer with your wager and name. Current standings: Link

Week 8: Answer

Abraham Wald, a member of the Statistical Research Group at the time, saw this problem and made an unconventional suggestion that saved countless lives.

Don't arm the places that sustained the most damage on planes that came back. By virtue of the fact that these planes came back, these parts of the planes can sustain damage.

If an essential part of the plane comes back consistently undamaged, like the engines in the previous example, that's probably because all the planes with shot-up engines don't make it back.

Any questions so far?

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

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**

Red River Showdown

- Longhorns keep playing against OU every year.
- Starting from the current year, how many expected years will it take for them to **win** with a **score difference of more than 5**.
- Assume:
  - Independence of events every year
  - Every year difference of score is normal with mean=-1 and sd=8
- Y= no. of years before score difference is more than 5
- Negative binomial distribution
- $E(Y) = \frac{m}{p} = \frac{1}{0.31} = 3.23$ years

Joint RV review

In general, if X and Y are any two discrete variables, the **joint probability mass function** $P_{XY}(x, y)$ is the probability of seeing both $X = x$ and $Y = y$.

To be a valid joint PMF, $P_{XY}(x, y) \geq 0$ for all x and y, and $\sum_x \sum_y P_{XY}(x, y) = 1$.

The **marginal PMF of** X gives us the distribution of X when we aren't concerned with Y:

$$P_X(x) = \sum_y P_{XY}(x, y)$$

Two discrete random variables X and Y are **independent** if $P_{XY}(x, y) = P_X(x)P_Y(y)$ for *every* possible value of x and y. If this is not true (even for one value of x and y), they are **dependent**.

Joint RV example - discrete

- Some customers purchase both auto and homeowners insurance from the same company.
- Let X and Y represent the deductibles of the auto and homeowners policies for a randomly selected customer. X and Y follow the joint pmf shown in the table:
- Calculate the marginal pmfs of X and Y
  - $P_X(0) = 0.46;\ P_X(100) = 0.26; P_X(200) = 0.28$
  - $P_Y(0) = 0.46;\ P_Y(50) = 0.26; P_Y(150) = 0.28$
- Are X and Y independent?
  - No

Y

X	0	50	150
0	0.25	0.06	0.15
100	0.07	0.15	0.04
200	0.14	0.05	0.09

Joint RV example - discrete

- What is the probability that a randomly selected customer has a homeowners deductible of atleast $50?
  - Ans=0.54
- What is the probability that a randomly selected customer pays higher on auto insurance?
  - Ans=0.50
- Evaluate the probability: $P(X+Y\ >110)$
  - Ans=0.62

X \ Y	0	50	150
0	0.25	0.06	0.15
100	0.07	0.15	0.04
200	0.14	0.05	0.09

Joint RV review

The **joint density function** $f_{XY}(x,y)$ is valid if $f_{XY}(x,y) \geq 0$ for all x and y, and if $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\ dy\ dx = 1$.

The **marginal density functions** are $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\ dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\ dx$

X and Y are independent if $f_{XY}(x,y) = f_X(x)f_Y(y)$ for all x and y

$E[h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)f_{XY}(x,y)\ dy\ dx$

Commuting to campus

- In commuting to campus, Anna must first get on a bus near her house and then transfer to a second bus.
- **If the waiting time (in minutes) at stop 1 and 2 be X and Y, both of which have a uniform distribution with mean 2.5 minutes and variance 25/12 minutes**, then evaluate the following (assume the arrival of bus is **independent** on two stops):
- What is the joint PDF of X and Y? R_{XY}?
  - $f_{XY}(x,y) = \frac{1}{25}$; $R_{XY} = \{(x,y) \mid 0 \le x \le 5 \; and \; 0 \le y \le 5\}$
- What is the probability that wait time at bus stop 1 is more than the wait time at bus stop 2?
  - Ans=1/2
- What is the probability that Ann will wait more than 6 minutes in total?
  - Ans= 8/25

Airplanes

- Two airplanes are flying in the **same direction in adjacent parallel corridors**. At time $t = 0$, the first aircraft is $10\ km$ ahead of the second one.
- Suppose the speed X of the first plane(km/hr) is normally distributed with mean 520 and standard deviation 10 and second plane's speed Y is also normally distributed with mean and standard deviation 600 and 10 respectively. Assume speeds are **independent**.
- What is the joint PDF of X and Y? Is it a special distribution?
  - $f_{XY}(x, y) = f_X(x) f_Y(y)$ where $f_X(x)$ and $f_Y(y)$ are normal distribution PDFs
  - No
- What is the probability that after $2\ hr$ of flying, the second plane has not caught up to the first plane?
  - $P(2X + 10 > 2Y) = P(Y - X < 5) = P(Z < -7.5) \approx 0$
  - Use the property that difference of normal distributions is also normal
- Determine the probability that the planes are separated by at most $10\ km$ after 2 hour?
  - $P(|2X - 2Y| < 10)$

Review

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y].$$

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- Properties of covariance
  - $Cov(X, X) =$
  - X, Y independent implies $Cov(X, Y) =$
  - $Cov(X, Y) \quad Cov(Y, X)$
  - $Cov(aX, Y) =$
  - $Cov(X + c, Y) =$
  - $Cov(X + Y, Z) =$

For **any** random variables $X_1, \ldots, X_n$

$$E[a_1X_1 + a_2X_2 + \cdots a_nX_n] = a_1E[X_1] + a_2E[X_2] + \cdots + a_nE[X_n]$$

$$V[a_1X_1 + a_2X_2 + \cdots a_nX_n] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{Cov}(X_i, X_j)$$

Auto and homeowners insurance

- Let X and Y represent the deductibles of the auto and homeowners policies for a randomly selected customer. X and Y follow the joint pmf shown in the table:

- Find covariance of X and Y
- Find correlation of X and Y

Y

X	0	50	150
0	0.25	0.06	0.15
100	0.07	0.15	0.04
200	0.14	0.05	0.09

Auto and homeowners insurance

- E(XY)= 3270
- E(X)= 59.6
- E(Y)= 55
- Var(X)= 6632-59.6^2= 3079.84
- Var(Y)= 6950- 55^2 = 3925
- Cov(X,Y)= -8
- Correlation(X,Y)= -0.0023

- Find
  - $E(25X+30Y)$ = 3140
  - $E(X^2)$ = 6632
  - $Cov(25+X-10Y, X+Y)$ =
    - Ans= -36098.76

Y

		0	50	150
	0	0.25	0.06	0.15
X	100	0.07	0.15	0.04
	120	0.14	0.05	0.09

Continuous distribution covariance

- Suppose the air pressure in each of the front tires of a particular vehicle is a random variable X for the right tire and Y for the left tire (psi units), with joint pdf

- $f(x,y) = \begin{cases} K(x^2+y^2) & 20 \le X \le 30, 20 \le Y \le 30 \\ 0 & otherwise \end{cases}$

- (a) What is the value of K ? =(3/38)*10^{-4}
- Find Cov(X,Y) = -0.1082

Central limit theorem review

If $X_1, \ldots X_n$ are a random sample (with each X_i having mean μ and variance σ^2, then

$$E[\overline{X}] = \mu$$

and

$$V[\overline{X}] = \frac{\sigma^2}{n}$$

Let $X_1, \ldots, X_n$ be a random sample. Then if n is sufficiently large, $\overline{X}$ has approximately a normal distribution, with mean and standard deviation given

Tips and tricks

- If it is given that variables are independent, then multiply the marginal PDF to get the joint PDF
- Joint uniform distribution
  - Probability over a region is the ratio of the area of the region with the total area of the feasible region
- Sum of exponential is Gamma distribution
- Sum of normal distribution is also a normal distribution

Question

Problem 3. (25 points) In a vain attempt to halt the thefts, the administration decides to levy a fine on any student caught stealing bricks. The fine involves a fixed charge of \$100, and a surcharge of \$15 per pound of bricks stolen. Assume that the weight of the bricks students are caught with is given by a gamma distribution with a mean of 4 pounds and a variance of 4 pounds^2.

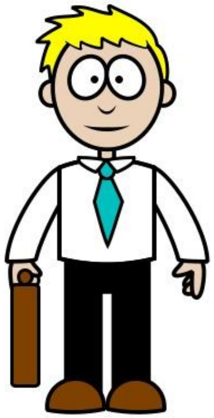
\$160 \$30

(a) (10) What is the expected fine levied by UT for a single student? What is the standard deviation?

(b) (10) This week, 100 students will be caught and fined. What is the mean and standard deviation of the average fine for these 100 students? Mean= \$160, Std dev = \$3

(c) (5) What is the probability that the average fine this week is greater than \$166? 0.0228

Lawyer question



Lawyer A

- You seek to hire one of two lawyers A and B for fighting your case. The cost of legal services at A and B have independent exponential distributions with mean \$60,000 and \$30,000 respectively.
  - $\lambda_1 = \left(\frac{1}{60000}\right)$ *and* $\lambda_2 = \left(\frac{1}{30000}\right)$
- You will pick the law firm with lower total cost. Let C be the actual cost you pay.
  - C=min(X,Y)
- What is the probability that you pay more than \$40,000 on legal services?
  - There is an easier way to do this problem
  - P(C> 40000)= P(X>40000 AND Y>40000)= P(X>40000)P(Y>40000)= $e^{-40000\lambda_1} e^{-40000\lambda_2} = 0.1353$

Lawyer B

Lawyer question

Lawyer A

- What is the CDF of C? =
- $P(C \leq c) = 1 - P(C \geq c)$
- $= 1 - P(X \geq c \text{ and } Y \geq c)$
- $= 1 - P(X \geq c)P(Y \geq c)$
- $= 1 - e^{-\lambda_1 c} e^{-\lambda_2 c}$
- $= 1 - e^{-(\lambda_1 + \lambda_2)c}$
- Find the mean and standard deviation of C?
  - Realize that C is exponential with lambda value as the sum of lambdas of X and Y
  - Mean= $1/(\lambda_1+\lambda_2)$, Std dev= $1/(\lambda_1+\lambda_2)$

Lawyer B

Thank you

Any Questions?