SPRING 2020

TEXAS
The University of Texas at Austin

CE 311S : PROBABILITY AND STATISTICS

Week 8 – Class 2

03/11/2020

PRIYADARSHAN PATIL
Teaching Assistant, The University of Texas at Austin

Administrative stuff

- Online assignment 4 is due tomorrow
- Spring break

Agenda

- Jointly distributed random variables
    - Multiple discrete random variables
    - Multiple continuous random variables
    - Covariance and correlation

Learning goals

- By the end of this class, you should be able to:
    - Understand a joint PMF (PDF) and CDF
    - Calculate marginal PMF (PDF)
    - Calculate expected values for the RV and functions of the RV
    - Compute covariance and correlation coefficient

Introduction to joint random variables

- Random variables are often linked with each other
- Examples: Years in college and Credits completed, Years of work experience and salary, Auto and Renters insurance
- We are interested in understanding how random variables behave when studied together

Example: Insurance

- Some customers purchase both auto and homeowner's 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:

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

Joint random variables

- In general, PMF $P_{XY}(x, y)$ is the probability of $X = x$ **and** $Y = y$
- For a valid PMF, $P_{XY}(x, y) \geq 0 \; \forall \; (x, y) \; and \; \sum_X \sum_Y P_{XY}(x, y) = 1$
- The **marginal PMF of X** provides us the distribution of X when we aren’t concerned with Y

$$\boldsymbol{P_X(x) = \sum_{y \in R_Y} P_{XY}\,(x, y)}$$

Things to note:

- Sum of all entries equals 1
- Each value is non-negative
- Sum of all values in the first row is P(X=0) when not considering Y
- Applies to all rows and columns
- Joint CDF is written as:

$$F_{XY}(x,y) = P(X \le x \cap Y \le y)$$

Y

		0	**50**	**150**	**Sum**
	0	0.25	0.06	0.15	**0.46**
X	100	0.07	0.15	0.04	**0.26**
	200	0.14	0.05	0.09	**0.28**
	Sum	**0.46**	**0.26**	**0.28**	**1**

Example: Insurance

- Calculate the marginal PMFs of X and Y
  - $P_X(0) = 0.46$
  - $P_X(100) = 0.26$
  - $P_X(200) = 0.28$
  - $P_Y(0) = 0.46$
  - $P_Y(50) = 0.26$
  - $P_Y(150) = 0.28$

Y

X

	0	**50**	**150**	**Sum**
0	0.25	0.06	0.15	**0.46**
100	0.07	0.15	0.04	**0.26**
200	0.14	0.05	0.09	**0.28**
Sum	**0.46**	**0.26**	**0.28**	**1**

Independence

- Two RVs are independent if $P_{XY}(x,y) = P_X(x)P_Y(y)\ for\ all\ (x,y)$
- Are X and Y independent?

Y

X

	0	**50**	**150**	**Sum**
0	0.25	0.06	0.15	**0.46**
100	0.07	0.15	0.04	**0.26**
200	0.14	0.05	0.09	**0.28**
Sum	**0.46**	**0.26**	**0.28**	**1**

Expected value

- The expected value of any function h_{XY} is $\sum_X \sum_Y h(x,y) P_{XY}(x,y)$
- What is the expected value of the total deductible (X+Y)?

Y

		0	**50**	**150**	**Sum**
X	0	0.25	0.06	0.15	**0.46**
	100	0.07	0.15	0.04	**0.26**
	200	0.14	0.05	0.09	**0.28**
	Sum	**0.46**	**0.26**	**0.28**	**1**

Expected value

- Create two tables, one with $P_{XY}(x, y)$ and one with $h(x, y)$ values
- Take the product of corresponding values and add
- $E[h(x, y)]$ = 0∗0.25 + 50∗0.06+…
- $E[h(x, y)]$ = 137

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

Y

X

	0	50	150
0	0	50	150
100	100	150	250
200	200	250	350

Expected value

- Two ways to calculate $E[X]\ and\ E[Y]$:
- Take the product of corresponding values and add.
- Solve it using the marginal PMF
- $E[X]\ =\ 82\ and\ E[Y]\ =\ 55$

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

Y

X

	0	50	150
0	0	0	0
100	100	100	100
200	200	200	200

Joint continuous random variables

- All the concepts we studied apply to continuous distributions
- Similar changes as applied to single random variables
- Mass changes to density, summation to integration, etc.

Joint continuous random variables

- The joint density function $f_{XY}(x, y)$ is valid if $f_{XY}(x, y) \geq 0 \; \forall x, y$ 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 $\mathrm{f}_Y(\mathrm{y}) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$

Joint continuous random variables

- X and Y are independent if $f_{XY}(x, y) = f_X(x) f_Y(y) \forall x, y$
- $E\left(h(X, Y)\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{XY}(x, y) f_{XY}(x, y)\, dy\, dx$

Example: Column lifetime

- A test column you built for your materials class can either fail via the rebars rusting, or by the concrete flaking off.
- Let X be the years before the rebars rust to failure, and Y be the years before the concrete flakes off.
- The joint pdf is: $f_{XY}(x, y) = ce^{-x}e^{-2y}$ for $x \geq 0, y \geq 0$

Example: Column lifetime

- $f_{XY}(x, y) = 2e^{-x}e^{-2y}$ for $x \geq 0, y \geq 0$
- What is the marginal distribution of X? This is the pdf for years till rebar rusting
- What is the marginal distribution of Y? This is the pdf for years till concrete flaking

Example: Column lifetime

- $f_{XY}(x,y) = 2e^{-x}e^{-2y}$ for $x \geq 0, y \geq 0$
- X and Y are independent if $f_{XY}(x,y) = f_X(x)f_Y(y)\forall x, y$
- Are X and Y independent?

Example: Column lifetime

- $f_{XY}(x, y) = 2e^{-x}e^{-2y}$ for $x \geq 0, y \geq 0$
- What is the expected time till the rebars rust to failure?

Covariance and correlation

- When two RVs are not independent, we require a measure of how dependent they are.
- **Covariance** of RVs X and Y is defined as
  $$Cov(X, Y) = E[X - E[X]]E[Y - E[Y]]$$
- Equivalently,

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

Covariance

- Recall, $E[X] = 82$ and $E[Y] = 55$
- $E[XY] = 4550$
- $Cov(X,Y) = \ 4550 - 82 * 55 = 40$

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

Y

X

	0	50	150
0	0	0	0
100	0	5000	15000
200	0	10000	30000

Covariance

- Interpretation
  - If covariance is positive, when X is above average, Y usually is too; and when X is below average, Y usually is too.
  - If covariance is negative, when X is above average, Y is usually below average, and vice versa.
  - If X and Y are independent, their covariance is zero. (The converse is **not** true).
  - The magnitude does not mean much (depends on units of X and Y)

Correlation

- To gain more insight from the magnitude, we define the **correlation coefficient** as follows:

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

- The correlation coefficient is always between $[-1, +1]$
- It quantifies the strength of the linear relationship between X and Y

Correlation

- If $\rho_{XY} = 1$, then $Y = aX + b$ for some $a > 0$
- If $\rho_{XY} = -1$, then $Y = aX + b$ for some $a < 0$
- If $\rho_{XY} = 0$, there is no linear relationship between X and Y
- If ρ_{XY}=0, it does not imply that X and Y are independent

Covariance - properties

- $Cov(X,X) = Var[X]$
- If X and Y are independent, $Cov(X,Y) = 0$
- $Cov(X,Y) = Cov(Y,X)$
- $Cov(aX,Y) = aCov(X,Y)$
- $Cov(X+c,Y) = Cov(X,Y)$
- $Cov(X+Y,Z) = Cov(X,Z) + Cov(Y,Z)$

Covariance – special formulae

- $Cov\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j Cov(X_i, Y_j)$
- $Var(aX + bY) = a^2 Var(x) + b^2 Var(Y) + 2ab Cov(X, Y)$

Covariance - examples

- $Cov(X_1 + 2X_2, 3Y_1 + 4Y_2) = 3Cov(X_1, Y_1) + 6Cov(X_2, Y_1) + 4Cov(X_1, Y_2) + 8Cov(X_2, Y_2)$
- Let X and Y be independent standard normal random variables. What is $Cov(1 + X + XY^2, 1 + X)$

Summary

- Joint discrete (continuous) random variables have a joint PMF (PDF) and CDF
- Marginal distributions for each of the RVs can be calculated by summing (integrating) across the other random variable
- Expected values for functions of joint random variables are like expected values for single random variables
- Covariance and correlation coefficient are measures for determining the linear relation between two RVs

Any Questions?

- Thank you for attending
- Have a fun (and safe) spring break