Week_5

Thursday, January 30, 2020 1:33 PM

STA255 Week-5

Shahriar Shams

03/02/2019

Review of Week-4

- Expectation
- Variance and standard deviation
- Moment generating function and it's use

Learning goals

- Joint distribution for two discrete variables
  - Joint probability mass function(pmf)
  - Marginal probability mass functions
- Joint distribution for two continuous variables
  - Joint probability density function (pdf)
  - Marginal probability mass functions
- Independence of two/or more variables (discrete or continuous)
- Expectation, covariance and correlation
- Conditional distributions

Joint distribution

- Up until last week, everything we have learned relates to a single variable.
- For example, we now know how to calculate these followings:
  - Tossing three fair coins, X = number of heads. $P(X = 1) =?$
  - Rolling two fair dice, S = sum of the two faces. $P(S = 4) =?$ or $E(S) =?$
  - $X \sim Exp(\lambda = 2)$. P(X<1) = ? or $E(X) =?$
- This chapter introduces the idea of distributions with more than one variable
- Though Joint distributions deal with any number of variables, we will keep it simple and learn distribution for two variables which is often called *bi-variate* distributions.

Joint distribution for two discrete variables

Joint probability mass function(pmf)

X and Y are two discrete random variables defined on the sample space S. The joint probability mass function $p(x, y)$ is defined for each pair of numbers (x, y) by

$$p(x, y) = P[X = x \text{ and } Y = y]$$

- $p(x, y) \geq 0$
- $\sum_x \sum_y p(x, y) = 1$

Example:

- Suppose we are rolling two fair dice.
- There are 36 outcomes in the sample space.

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- Let's define two discrete random variables:
  - Let S = sum of the two faces
  - Let M = maximum of the two faces
- S will take value 2,3,4,...,12
- M will take value 1,2,3,...,6
- A joint pmf of S and M will give us the probability of S equals certain value AND M equals certain value.
- $P(S = 2, M = 1) =?$
  - Out of 36 outcomes, only one (1,1) gives us $S = 2$ and $M = 1$.
  - Therefore, the probability is $1/36$
- We can calculate this for all the different values of S and M which gives us the table of the joint pmf.

a	b = 1	2	3	4	5	6
2	1/36	0	0	0	0	0
3	0	2/36	0	0	0	0
4	0	1/36	2/36	0	0	0
5	0	0	2/36	2/36	0	0
6	0	0	1/36	2/36	2/36	0
7	0	0	0	2/36	2/36	2/36
8	0	0	0	1/36	2/36	2/36
9	0	0	0	0	2/36	2/36
10	0	0	0	0	1/36	2/36
11	0	0	0	0	0	2/36
12	0	0	0	0	0	1/36

- Each cell in this table gives one specific probability, $P(S = a, M = b)$

Marginal probability mass functions

- Marginal probability mass functions take us back to single variable distributions.
- For example,
  - based on the join pmf, P(S=4) = ?

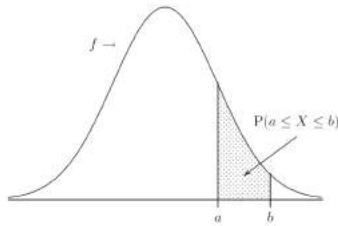
$$\begin{aligned} P(S=4) &= P(S=4, M=1 \text{ OR } S=4, M=2 \text{ OR } ... \text{ OR } S=4, M=6) \\ &= P(S=4, M=1) + P(S=4, M=2) + ... + P(S=4, M=6) \\ &= 3/36 \end{aligned}$$

  - This is the sum of all probabilities corresponding to the row of S=4
  - Similarly we can calculate probabilities of S taking other values.
  - This gives the marginal pmf of S (this is the same pmf we have learned in Week-2,page-2)
- Similarly, by taking the column totals we will get the marginal pmf of M.
- The last column and the last row of this following table are the marginal probabilities of S and M respectively.

a	b = 1	2	3	4	5	6	$p_S(a)$
2	1/36	0	0	0	0	0	1/36
3	0	2/36	0	0	0	0	2/36
4	0	1/36	2/36	0	0	0	3/36
5	0	0	2/36	2/36	0	0	4/36
6	0	0	1/36	2/36	2/36	0	5/36
7	0	0	0	2/36	2/36	2/36	6/36
8	0	0	0	1/36	2/36	2/36	5/36
9	0	0	0	0	2/36	2/36	4/36
10	0	0	0	0	1/36	2/36	3/36
11	0	0	0	0	0	2/36	2/36
12	0	0	0	0	0	1/36	1/36
$p_M(b)$	1/36	3/36	5/36	7/36	9/36	11/36	1

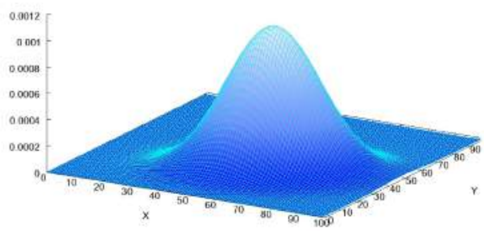
Joint distribution for two continuous variables

- **Recall:**
  - A *pdf* is a non negative function which gives us a line.
  - The area underneath the curve gives us probability.
  - Total area under the curve equals 1.



- In two variables case the function doesn't give us a line rather it gives us a surface.

- Using the similar idea of single variable pdf
  - The joint pdf is a non negative function.
  - The volume underneath the surface gives us probability.
  - Total volume under the surface equals 1.



- Here is the **formal definition** of joint *pdf*

Let X and Y be two continuous random variables. $f(x, y)$ is called the joint probability density function for X and Y if

- $f(x, y) \geq 0$
- $\int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} f(x, y) dx dy = 1$

And it defines probability as

- $P[a \leq X \leq b, c \leq Y \leq d] = \int_{y=c}^{d} \int_{x=a}^{b} f(x, y) dx dy$

Example: Suppose a function is given to us

$$f(x, y) = \frac{6}{5}(x + y^2) \ \ for\, 0 \leq x \leq 1;\ 0 \leq y \leq 1$$

Is this a valid joint *pdf*?

- $f(x,y) \geq 0$ for the ranges of x and y
- $\int_{y=0}^{1}\int_{x=0}^{1} \frac{6}{5}(x+y^2)\,dx\,dy$

$$= \int_{y=0}^{1} \frac{6}{5}\left(\frac{x^2}{2}+y^2x\right)\Big|_0^1 \,dy$$

$$= \int_0^1 \frac{6}{5}\left(\frac{1}{2}+y^2\right) dy$$

$$= \frac{6}{5}\left(\frac{1}{2}y + \frac{y^3}{3}\right)\Big|_0^1$$

$$= \frac{6}{5}\left(\frac{1}{2}+\frac{1}{3}\right)$$

$$= 1$$

Calculate $P[0 \leq X \leq 1/4, 0 \leq Y \leq 1/4]$

$$= \int_{y=0}^{1/4}\int_{x=0}^{1/4} \frac{6}{5}(x+y^2)\,dx\,dy$$

$$= \int_{y=0}^{1/4} \frac{6}{5}\left(\frac{x^2}{2}+y^2x\right)\Big|_0^{1/4} dy$$

$$= \int_{y=0}^{1/4} \frac{6}{5}\left(\frac{1}{32}+\frac{y^2}{4}\right) dy$$

$$= \frac{6}{5}\left(\frac{1}{32}y + \frac{y^3}{12}\right)\Big|_0^{1/4}$$

$$= \frac{6}{5}\left(\frac{1}{32}\cdot\frac{1}{4} + \frac{1}{12}\cdot\frac{1}{64}\right)$$

$=$

Marginal probability density functions

- Similar to the discrete cases, Marginal distributions takes us back to single variable distributions.
- Marginal pdf of X,
$$f_X(x) = \int_{y=-\infty}^{\infty} f(x,y)dy$$
- Marginal pdf of Y,
$$f_Y(y) = \int_{x=-\infty}^{\infty} f(x,y)dx$$
- continuing with the same example the marginal distributions are

$$f_X(x) = \int_{y=0}^{1} \frac{6}{5}(x+y^2)dy = \frac{6}{5}x + \frac{2}{5} \;\; ; \;\; 0 \le x \le 1$$

$$f_Y(y) = \int_{x=0}^{1} \frac{6}{5}(x+y^2)dx = \frac{6}{5}y^2 + \frac{3}{5} \;\; ; \;\; 0 \le y \leqslant 1$$

Independence of two variables (discrete or continuous)

- Recall:
  - Let A and B are two events
  - A is independent of B if $P(A \cap B) = P(A) * P(B)$
  - Independence suggests probability of A and B both happening is equal to the multiplication of two individual probabilities.
- We will use the same intuition.
- If the joint can be written as a multiplication of the marginals we call the two variables independent.

Discrete case

Two discrete random variables X and Y are called independent if

$$P(X = x, Y = y) = P(X = x) * P(Y = y)$$

for every pair of x and y.

Example: For the table at the bottom of page 3 of this pdf,

- $P(S = 2, M = 1) = 1/36$
- $P(S = 2) = 1/36$
- $P(M = 1) = 1/36$
- $P(S = 2, M = 1) \neq P(S = 2) * P(M = 1)$

- Therefore S and M are not independent.
- To prove independence, we have to check this for all the cells of the joint pmf.

Continuous case

Two continuous random variables X and Y are called independent if

$$f(x,y) = f_X(x) * f_Y(y)$$

- For more than two random variables the idea remains same.
- The joint density can be written as the product of all the marginal densities.

$$f(x_1, x_2, ..., x_n) = f(x_1) * f(x_2) * ... * f(x_n)$$

Expectation using joint distribution

- Last week, we learned about joint pmf and joint pdf of two variables.
- We are now interested in calculating expectation of any real valued function of these two variables.
- Suppose $g(X,Y)$ is that real valued function. We want to calculate $E[h(X,Y)]$
- If X and Y both are discrete,

$$E[h(X,Y)] = \sum_x \sum_y h(x,y) * P(X = x, Y = y)$$

- As an example, for this following joint pmf,

	a			
b	0	1	2	P$(Y = b)$
-1	1/6	1/6	1/6	1/2
1	0	1/2	0	1/2
P$(X = a)$	1/6	2/3	1/6	1

- calculate $E[XY]$
- calculate $E[X+Y]$

$E[XY] = (0*-1)*\frac{1}{6} + (1*-1)*\frac{1}{6} + (2*-1)*\frac{1}{6} +$
$(0*1)*0 + (1*1)*\frac{1}{2} + (2*1)*0$
$= 0$

$E[X+Y] = (0-1)*\frac{1}{6} + (1-1)*\frac{1}{6} + (2-1)*\frac{1}{6} +$
$(0+1)*0 + (1+1)*\frac{1}{2} + (2+1)*0$
$= -\frac{1}{6} + 0 + \frac{1}{6} + 0 + 1 + 0 = 1$

- If X and Y both are continuous random variables with joint *pdf* $f(x, y)$

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

- For two continuous random variables with joint *pdf* $f(x, y)$, show that $E[X + Y] = E[X] + E[Y]$

For two cont. random variable
X and Y, show
$E[X+Y] = E[X] + E[Y]$

$$E[X+Y] = \int_y \int_x (x+y) f(x,y)\, dx\, dy$$
$$= \int_y \int_x x f(x,y)\, dx\, dy + \int_y \int_x y f(x,y)\, dx\, dy$$
$$= \int_x \int_y x f(x,y)\, dy\, dx + \int_y \int_x y f(x,y)\, dx\, dy$$
$$= \int_x x \int_y f(x,y)\, dy\, dx + \int_y y \int_x f(x,y)\, dx\, dy$$
$$= \int_x x\, f_X(x)\, dx + \int_y y\, f_Y(y)\, dy$$
$$= E[X] + E[Y]$$

- Homework:
  - X and Y are two **independent** continuous random variables. Show $E[X + Y] = E[X] + E[Y]$
  - X and Y are two continuous random variables. Show $E[rX + sY + t] = rE[X] + sE[Y] + t$, where r, s and t are constants.

Covariance

- Irrespective of whether X and Y are independent or not,

$$E[X + Y] = E[X] + E[Y]$$

- For variance this relationship only holds if X and Y are independent.
- What happens if X and Y are not independent?
- Show

$$V[X + Y] = V[X] + V[Y] + 2E[(X - E[X])(Y - E[Y])]$$

$$\begin{aligned}
V[X+Y] &= E\left[\left(X+Y - E[X+Y]\right)^2\right] \\
&= E\left[\left\{X+Y - \left(E[X]+E[Y]\right)\right\}^2\right] \\
&= E\left[\left(\underline{X - E[X]} + \underline{Y - E[Y]}\right)^2\right] \\
&= E\left[\left(X-E[X]\right)^2 + \left(Y - E[Y]\right)^2 + 2\left(X-E[X]\right)\left(Y-E[Y]\right)\right] \\
&= E\left[\left(X-E[X]\right)^2\right] + E\left[\left(Y-E[Y]\right)^2\right] + 2E\left[\left(X-E[X]\right)\left(Y-E[Y]\right)\right] \\
&= V[X] + V[Y] + 2\,cov(X, Y)
\end{aligned}$$

Definition of covariance

Covariance between X and Y are defined as

$$cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

- An alternative formula

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

- So we can write

$$V[X + Y] = V[X] + V[Y] + 2cov(X, Y)$$

and

$$V[X - Y] = V[X] + V[Y] - 2cov(X, Y)$$

- Covariance measures the dependency/association between two variables.
- Unlike variance, covariance can be negative.
- A negative covariance indicates a negative association between the two variables.
- Just like variance, covariance is affected by multiplication and division but not addition/subtraction

$$cov(rX + s, tY + u) = rt\, cov(X, Y)$$

- If X and Y are independent, $cov(X, Y) = 0$

If X and Y are indep.

$$E[XY] = E[X]\,E[Y]$$

X is indep of $Y \Rightarrow f(x,y) = f_X(x) * f_Y(y)$

$$\begin{aligned}
\therefore E[XY] &= \int_y \int_x x y\, f(x,y)\, dx\, dy \\
&= \int_y \int_x x y\, f_X(x)\, f_Y(y)\, dx\, dy \\
&= \int_y y\, f_Y(y) \underbrace{\int_x x\, f_X(x)\, dx}_{E[X]}\, dy \\
&= \int_y y\, f_Y(y)\, E[X]\, dy \\
&= E[X] \underbrace{\int_y y\, f_Y(y)\, dy}_{E[Y]} \\
&= E[X]\, E[Y]
\end{aligned}$$

$$\begin{aligned}
\therefore cov(X,Y) &= E[XY] - E[X]\,E[Y] \\
&= 0
\end{aligned}$$

$\Rightarrow$ if X is indep of Y, $cov(X,Y) = 0$

Correlation

- Correlation is a unit free measure of dependence between two random variables.
- correlation coefficient denoted by $\rho(X, Y)$ is defined as

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{V(X)V(Y)}}$$

- ρ is a number between -1 and 1.
- ρ measures linear relationship.
- A negative value suggests negative relationship $\implies$ when one variable is increased the other other is found to be decreased.
- A positive value suggests positive relationship $\implies$ both variables increase or decrease at the same time.
- $\rho = 0$ means the variables are uncorrelated.

Scatter Plot

- If we plot two variables on a regular 2D graph it might look like the followings

- Scatter plot suggests the linear association between two variables.
  - If the points suggest a downward trend, it depicts a negative correlation (3rd graph)
  - If the points suggests an upward trend, it depicts a positive correlation (2nd graph)
  - If the points look completely random, it depicts zero correlation.
  - If we think of an imaginary line through the points, the closeness of the points to the line represents the strength of the relationship.
    - 2nd plot suggests a moderate association.
    - 3rd plot suggests a strong association.

Correlation and independence

- zero covariance or zero correlation $\implies$ no *linear* relationship. But there could be other types of non-linear relationships.
- If X and Y are independent then $cov(X,Y) = 0 \implies \rho(X,Y) = 0$
- But zero covariance or zero correlation does not imply independence.

Conditional Distributions

- **Recall:**
  - For two events A and B, $P[A|B] = \frac{P[A\cap B]}{P[B]}$
- Using the same idea,
  - conditional probability mass function of Y given $X = x$,

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)} = \frac{\text{join pmf of x and y}}{\text{marginal pmf of x}}$$

  - conditional probability density function of Y given $X = x$,

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\text{join pdf of x and y}}{\text{marginal pdf of x}}$$

- In a conditional distribution, the variable that we put the condition on is treated as constant (something known)
- Example of conditional distribution using discrete case

	b						
a	1	2	3	4	5	6	$p_S(a)$
2	1/36	0	0	0	0	0	1/36
3	0	2/36	0	0	0	0	2/36
4	0	1/36	2/36	0	0	0	3/36
5	0	0	2/36	2/36	0	0	4/36
6	0	0	1/36	2/36	2/36	0	5/36
7	0	0	0	2/36	2/36	2/36	6/36
8	0	0	0	1/36	2/36	2/36	5/36
9	0	0	0	0	2/36	2/36	4/36
10	0	0	0	0	1/36	2/36	3/36
11	0	0	0	0	0	2/36	2/36
12	0	0	0	0	0	1/36	1/36
$p_M(b)$	1/36	3/36	5/36	7/36	9/36	11/36	1

Calculate the pmf of M given $S = 6$.

M:	1	2	3	4	5	6
$p_{M\|S=6}(b)$:	0	0	$\frac{1/36}{5/36}$	$\frac{2/36}{5/36}$	$\frac{2/36}{5/36}$	0
	0	0	1/5	2/5	2/5	0

$$E[M \mid S=6] = (1*0)+(2*0)+(3*\tfrac{1}{5})+(4*\tfrac{2}{5})+(5*\tfrac{2}{5})+(6*0)$$
$$= 0 + 0 + \tfrac{3}{5} + \tfrac{8}{5} + 2 + 0$$
$$= \tfrac{21}{5}$$

Conditional mean and variance

- The conditional mean of Y given $X = x$ is defined as
  - for discrete case
$$E[Y|X = x] = \sum_y y p_{Y|X}(y|x)$$
  - for continuous case
$$E[Y|X = x] = \int_{y=-\infty}^{\infty} y f_{Y|X}(y|x)$$
- Basically it's a regular mean calculated based on the conditional distribution.
- Similarly, the variance of the conditional distribution is called conditional variance.

Chapter 5.4 and 5.5 are not needed.

Homework

Chapter 5.1

1, 3, 6, 9, 12, 13, 14

Chapter 5.2

18, 22, 25, 26, 29, 31

Chapter 5.3

39, 44, 45-48