Course No: MATH F113

Probability and Statistics

BITS Pilani
Pilani Campus

Joint Distributions

Sumanta Pasari

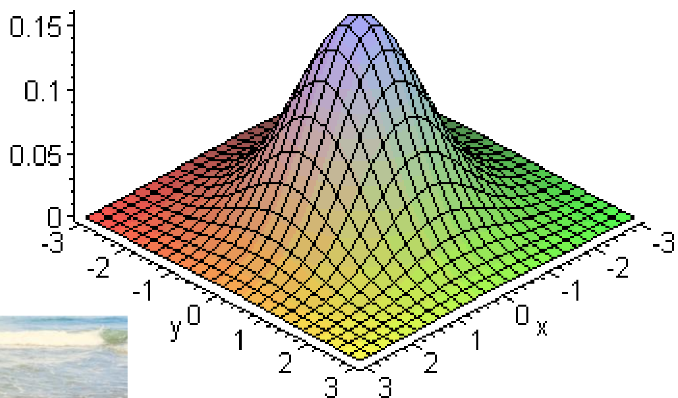
Office: 6067 C, NAB

sumanta.pasari@pilani.bits-pilani.ac.in

Overall: Step-by-Step

1. **Data Analysis**
   - Collection, classification, interpretation and analysis of data
   - Visualization using charts and graphs
   - Measurement of central tendency and dispersion

2. **Probability**
   - Probability axioms and rules
   - Important probability distributions
   - Joint distributions; moments, correlation



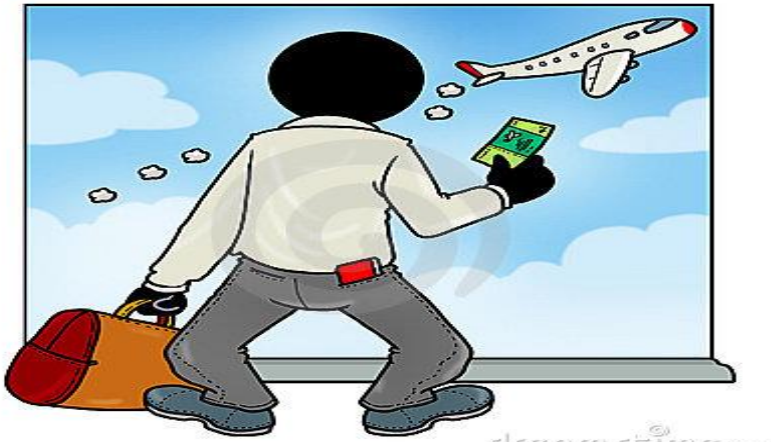
3. **Statistics**
   - Sample statistics
   - Parameter estimation
   - Statistical Inference
   - Simple Linear Regression

What is Probability?

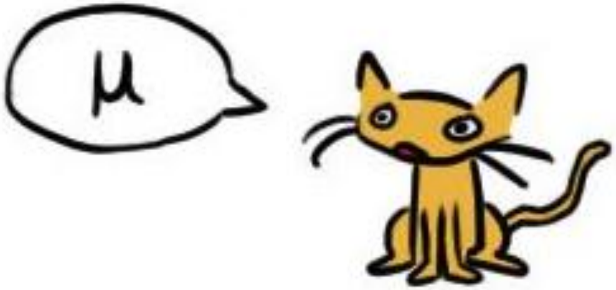
Discrete Distributions

Gambler's fallacy…

Poop by a bird…



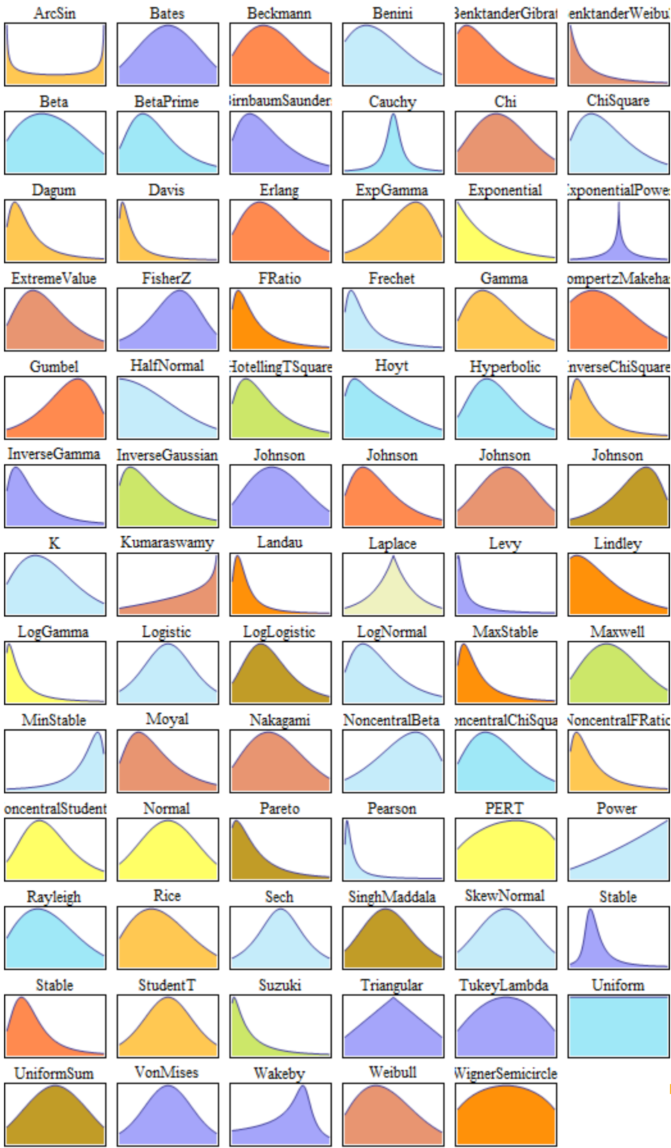
Missed flight…

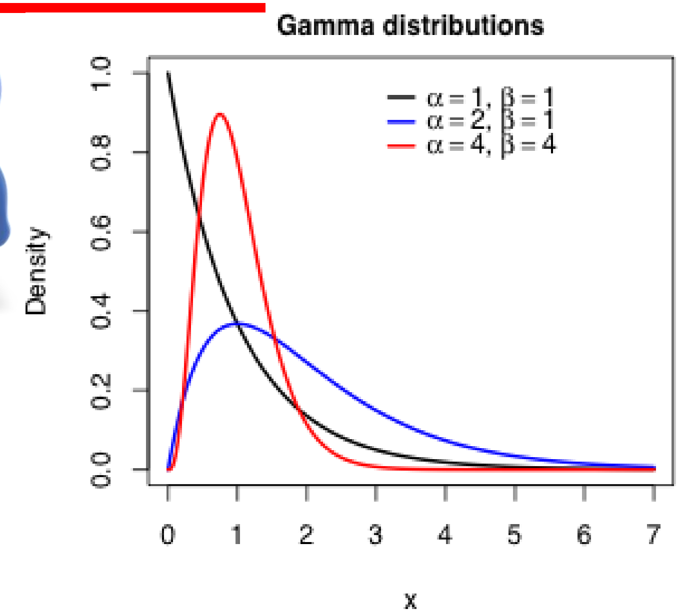


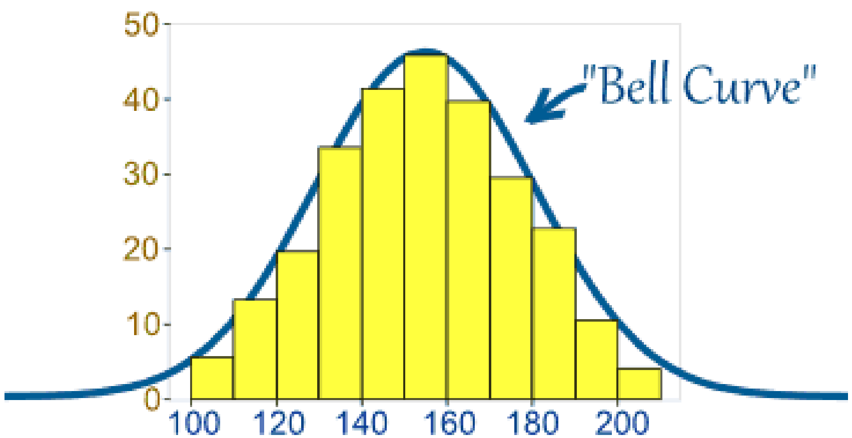
Expectation and Moments

(http://blog.minitab.com/blog/fun-with-statistics/poisson-processes-and-probability-of-poop)

Continuous Distribution







Let's Watch…

https://www.youtube.com/watch?v=Qfs_mvpA-vQ&t=112s

Joint Distributions

Single Random Variable (Univariate)

Discrete **Continuous**

Two Dimensional Random variables (Bivariate)

Discrete **Continuous**

Joint Distributions

- When two or more random variables are involved in a random experiment
- Toss a coin twice; ***S = {HH, HT, TH, TT}***;
  - Let ***X*** and ***Y*** be random variables associated with the first and second tossing. If 'head' appears, then associated RV is 1, otherwise 0. Events are {(X=0, Y=0), (X=0, Y=1), (X=1, Y=0), (X=1, Y=1)} . Can you find out what is ***P(X=i, Y=j)****, i=0,1, j= 0,1*?
  - Let ***X*** denote the number of 'heads' and ***Y*** denote the number of 'tails'. Notice that ***X+Y=2***. Then possible events are {(X=0, Y=2), (X=1, Y=1), (X=2, Y=0)} . Can you find out what is ***P(X=1, Y=1)?*** Note that in this example, events like (X=0, Y=0), (X=0, Y=1), (X=1, Y=0), (X=1, Y=2), (X=2, Y=1), (X=2, Y=2) are impossible events, as ***X +Y =2***.
- Two balls are selected at random from a box that contains 3 blue balls, 2 red balls and 3 green balls. If ***X*** is the number of blue balls and ***Y*** is the number of red balls selected, find ***P(X=1, Y=1), P(X+Y≤1).***
- Similarly, let ***X*** and ***Y*** be random variables associated with the height and weight of a group of students. Find ***P(150≤X≤180, 50≤Y≤80).***

Joint Distributions

Bivariate distribution occurs when we observe two nondeterministic quantities, one followed by another.

For example: Record the atmospheric temperature *T* in Celsius followed by atmospheric pressure *P* in pounds per square foot at a random place and time. This gives two dimensional r.v. *(T, P)*.

n-dimensional random variable $\leftrightarrow$ we observe *n* nondeterministic quantities in sequence.

Discrete Joint Density

- Let ***X*** and ***Y*** be discrete RV associated with a random experiment. Then the ordered pair ***(X,Y)*** is called two-dimensional discrete RV.
- The joint pdf for ***(X,Y)*** is $f_{XY}(x,y) = P(X = x \text{ and } Y = y)$

Necessary and sufficient condition for a function to be a discrete joint density:

$$\text{(i)} \quad f_{XY}(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$$

$$\text{(ii)} \sum_{\text{all } x} \sum_{\text{all y}} f_{XY}(x,y) = 1$$

- For any region ***A*** in the *xy* plane, the probability of ***A*** is defined as

$$P(A) = \sum \sum f_{XY}(x,y) \ \ \forall \ (x,y) \in A$$

Marginal Density of Discrete RVs

- From a known joint density ***f(x,y)*** for discrete RVs ***(X,Y)***, it is easy to find out their respective density functions.

(i) The marginal density for X, denoted by $f_X(x)$, is given by

$$f_X(x) = P(X = x) = \sum_{\text{all y}} f_{XY}(x,y)$$

(ii) The marginal density for Y, denoted by $f_Y(y)$, is given by

$$f_Y(y) = P(Y = y) = \sum_{\text{all x}} f_{XY}(x,y)$$

Ex.5.1 For a discrete random variable (X,Y), $P(X=i, Y=j)=0,\ i, j=-1,0,1$ except $P(X=0,Y=1)=\frac{1}{3}, P(X=1,Y=-1)=\frac{1}{3}, P(X=1,Y=1)=\frac{1}{3}$.

(i) Find the joint density for (X,Y) and (ii) marginal densities for X and Y.

Examples: Discrete Joint Distribution

Sol (Ex.5.1). We represent pdf and marginal densities in a tabular form.

Y=y \ X=x	-1	0	1	marginal Y f(y)
-1	0	0	1/3	**1/3**
0	0	0	0	**0**
1	0	1/3	1/3	**2/3**
marginal X f(x)	**0**	**1/3**	**2/3**	**Total Sum 1.0**

$$f_X(x) = \begin{cases} 1/3; & x = 0 \\ 2/3; & x = 1 \\ 0; & e.w \end{cases}$$

$$f_Y(y) = \begin{cases} 1/3; & y = -1 \\ 2/3; & y = 1 \\ 0; & e.w \end{cases}$$

HW.5.1 For (X,Y), $f_{XY}(x,y) = P(X = x, Y = y) = \dfrac{x^2 + y}{32}$; $x = 0,1,2,3$; $y = 0,1$.

Find the marginal densities $f_X(x)$ and $f_Y(y)$.

HW.5.2 Verify whether (a) $f_{XY}(x,y) = \dfrac{x^2 - y}{16}$; $x = 1,2$; $y = 0,1$ (b) $f_{XY}(x,y) = \dfrac{xy}{12}$; $x, y = 0,1$ can define discrte pdfs.

Examples: Discrete Joint Distribution

Ex.5.2. Two balls are selected at random from a box that contains 3 blue balls, 2 red balls and 3 green balls. If ***X*** is the number of blue balls and ***Y*** is the number of red balls selected, find pdf for (X,Y) and marginal pdfs. What is P(X+Y≤1)?

Y=y \ X=x	0	1	2	marginal Y f(y)
0	3/28	9/28	3/28	15/28
1	6/28	6/28	0	12/28
2	1/28	0	0	1/28
marginal X f(x)	10/28	15/28	3/28	Total Sum 1.0

Formula used here:

$$f(x,y)=\frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$$

$x=0,1,2$

$y=0,1,2$

First we find out the events $\{(x,y)$ such that $x+y\le 1\}$

$$P(X+Y\le 1)=f(0,0)+f(0,1)+f(1,0)=3/28+6/28+9/28=9/14$$

Examples: Discrete Joint Distribution

Ex.5.3. *Number of automobiles sold at two dealerships over 300 days*

	Loharu Dealership						
Pilani Dealership	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
0	21	30	24	9	2	0	**86**
1	21	36	33	18	2	1	**111**
2	9	42	9	12	3	2	**77**
3	3	9	6	3	5	0	**26**
Total	**54**	**117**	**72**	**42**	**12**	**3**	**300**

(a) Develop a joint probability distribution table

(b) Find E(X), V(X), E(X+Y), V(X+Y), P(X+Y>4)

Examples: Discrete Joint Distribution

Sol.5.3.

	Y: number of sales at Loharu						
X: number of sales at Pilani	0	1	2	3	4	5	f(x)
0	0.0700	0.1000	0.0800	0.0300	0.0067	0.0000	0.2867
1	0.0700	0.1200	0.1100	0.0600	0.0067	0.0033	0.3700
2	0.0300	0.1400	0.0300	0.0400	0.0100	0.0067	0.2567
3	0.0100	0.0300	0.0200	0.0100	0.0167	0.0000	0.0867
f(y)	0.1800	0.3900	0.2400	0.1400	0.0400	0.0100	1.0000

Joint Density for Continuous RVs

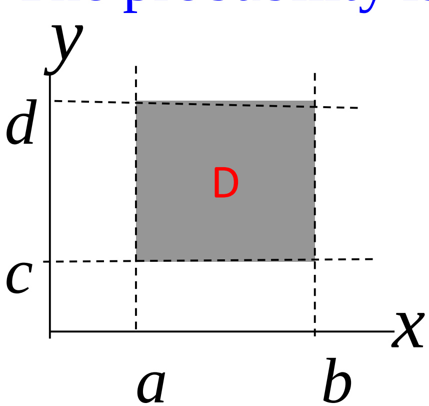
- Let ***X*** and ***Y*** be continuous RV associated with a random experiment. Then the ordered pair ***(X,Y)*** is called two-dimensional continuous RV.

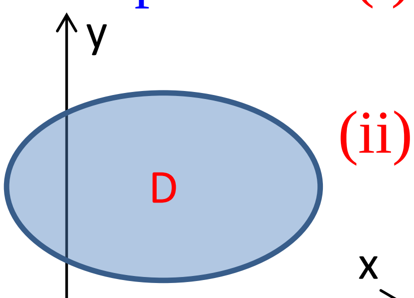
Necessary and sufficient condition for ***f(x,y)*** to be a continuous joint density

(i) $f_{XY}(x,y) \geq 0, \quad \forall (x,y) \in \mathbb{R}^2$

(ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy\,dx = 1$

The probability is computed as (i) $P(a < X \leq b,\ c < Y \leq d) = \int_a^b \int_c^d f_{XY}(x,y)\,dy\,dx$

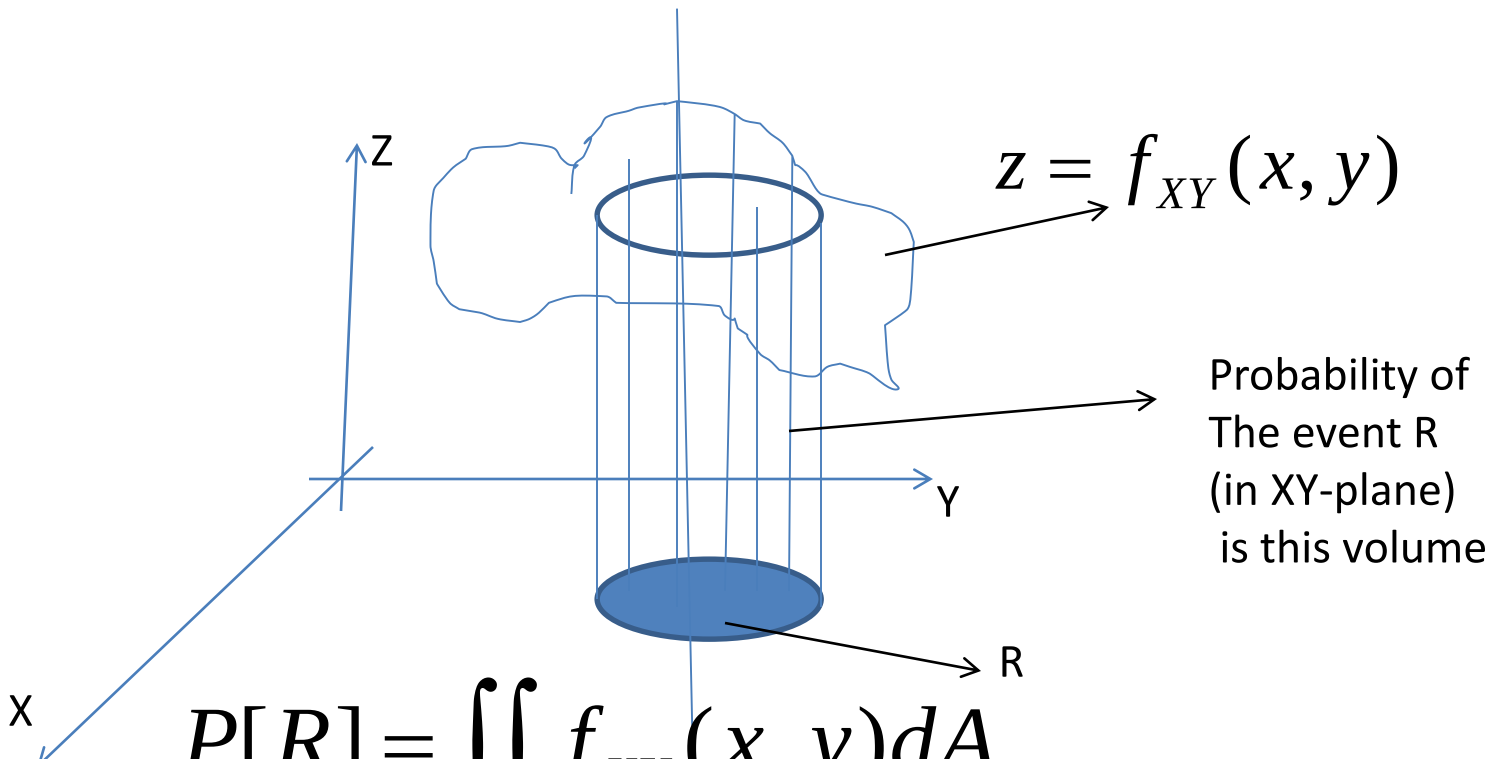




(ii) $D \subset \mathbb{R}^2, P(A) = \iint_D f(x,y)\,dA$

Joint Density for Continuous RVs

<u>Note</u>: In one dimensional continuous case, the probabilities correspond to areas under density curve while in the case of 2-D, they corresponds to volumes under density surfaces.



$$P[R] = \iint_R f_{XY}(x, y)\, dA$$

Marginal Density of Continuous RVs

- From a known joint density ***f(x,y)*** for continuous RVs ***(X,Y)***, it is easy to find out their respective density functions.

(i) The marginal density for X, denoted by $f_X\left(x\right)$, is given by

$$f_X\left(x\right)=\int_{-\infty}^{\infty} f_{XY}(x,y)dy$$

(ii) The marginal density for Y, denoted by $f_Y\left(y\right)$, is given by

$$f_Y\left(y\right)=\int_{-\infty}^{\infty} f_{XY}(x,y)dx$$

Remarks:

(i) The marginal densities are also density functions (Verify!)

(ii) Thus, $f_X\left(x\right)$ and $f_Y\left(y\right)$ will follow the properties of density functions

For n-dimensional RV

- The concept can be extended in an analogous manner to n dimensions.
- The joint density $f_{X_1 \ldots X_n}(x_1,\ldots,x_n)$ is a function of n random variables.
- Replicate the definition.

Independent Random Variables

Let (X, Y) be a two-dimensional RV with joint density f_{XY} and marginal densities f_X, f_Y, respectively.

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

Independent Random Variables

In general for n RVs,

$X_1, \ldots, X_n$ are independent iff

$$f_{X_1 \ldots X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

Here $f_{X_1 \ldots X_n}(x_1, \ldots, x_n)$ is the joint density and $f_{X_i}(x_i)$ are the marginal densities of X_i.

Exercise 5.1 (From Book)

3. The joint density for (X,Y) is given by:

$$f_{XY}(x,y) = \frac{1}{n^2}, x = 1,2,3,..n$$

$$y = 1,2,3,..n$$

and $f_{XY}(x,y) = 0$ elsewhere.

(a) Verify that $f_{XY}(x,y)$ satisfies the conditions necessary to be a density.

Exercise 5.1 (Q. 3)

(b) Find the marginal densities for X and Y

(c) Are X and Y independent?

Sol:

$$for\ \ n>0,\quad f_{XY}(x,y)=\frac{1}{n^2}\geq 0,\ \forall\ \ x=1,\ldots,n$$

$$\text{and}\ \ y=1,\ldots,n;\ \text{and equals 0 elsewhere.}$$

$$also,$$

$$\sum_{x=1}^{n}\sum_{y=1}^{n}\frac{1}{n^2}=\sum_{x=1}^{n}\frac{1}{n^2}n=1$$

Exercise 5.1 (Q. 3)

(b) $$f_X(x) = \sum_{y=1}^{n} \frac{1}{n^2} = 1/n, \; x = 1, 2, \ldots n$$

$$= 0 \qquad \text{e.w.}$$

$$f_Y(y) = \sum_{x=1}^{n} \frac{1}{n^2} = 1/n, \; y = 1, 2, \ldots n$$

$$= 0 \qquad \text{e.w.}$$

(c) The random variables X, Y are independent:

Exercise 5.1 (Q. 3)

$$f_X(x)\,f_Y(y) = \frac{1}{n}\frac{1}{n}$$

$$= \frac{1}{n^2} = f_{XY}(x, y)$$

$$\text{for } 1 \le x \le n, 1 \le y \le n \text{ while}$$

$$f_{XY}(x, y) = 0 = f_X(x)\,f_Y(y)$$

$$\text{elsewhere.}$$

innovate achieve lead

BITS Pilani
Pilani Campus

Course No: MATH F113

Probability and Statistics

Joint Distributions

Exercise 5.1 (Q. 4)

4. The joint density for (X,Y) is given by

$$f_{XY}(x,y) = \frac{2}{n(n+1)}, 1 \le y \le x \le n$$

x, y **integers** n is a positive integer

(a) Verify that $f_{XY}(x,y)$ satisfies the conditions necessary to be a density.

(b) Find the marginal densities for X and Y

Exercise 5.1 (Q. 4)

(c) Are X and Y independent?

(d) Assume that n=5, use the density to find

$$P[X \leq 3 \ \text{ and } \ Y \leq 2]$$

Find also

$$P[X \leq 3] \ \text{ and } \ P[Y \leq 2]$$

Exercise 5.1 (Q. 4)

Sol:

$$\text{for } 1 \le y \le x \le n$$

$$\text{Since } f(x,y) = \frac{2}{n(n+1)} > 0,$$

as n is a +ve integer

and $f(x,y) = 0$ elsewhere.

$$\text{also, } \sum_{\text{all } (x,y)} f(x,y) = \sum_{y=1}^{n} \sum_{x=y}^{n} \frac{2}{n(n+1)}$$

$$= \frac{2}{n(n+1)} \sum_{y=1}^{n} (n - y + 1)$$

Exercise 5.1 (Q. 4)

$$= \frac{2}{n(n+1)}\{n+(n-1)+(n-2)+\ldots+1\} = 1$$

$$b) \quad f_X(x) = \sum_{y=1}^{x} \frac{2}{n(n+1)}$$

$$= \frac{2x}{n(n+1)}, x = 1, 2, \ldots n$$

$$f_Y(y) = \sum_{x=y}^{n} \frac{2}{n(n+1)} = \frac{2(n-y+1)}{n(n+1)}, y = 1, 2, ..n$$

Exercise 5.1 (Q. 4)

$$c)\ X \text{ and } Y \text{ are not independent}$$

$$f_X(1)\ f_Y(2) = \frac{2\times 1}{n(n+1)}\frac{2(n-2+1)}{n(n+1)}$$

$$\neq 0 = f_{XY}(1,2)$$

Exercise 5.1 (Q. 4)

d) when $n = 5$,

$$P[X \le 3 \text{ and } Y \le 2] = \sum_{y=1}^{2} \sum_{x=y}^{3} \frac{2}{5(6)}$$

$$= \frac{2}{30} \sum_{y=1}^{2} (3 - y + 1) = 1/3$$

$$P[X \le 3] = \sum_{x=1}^{3} \frac{2x}{5(6)} = \frac{2}{30} \frac{(2+4+6)}{2} = 2/5$$

$$= \frac{2}{30} \sum_{y=1}^{2} (3 - y + 1) = 1/3$$

Exercise 5.1 (Q. 4)

$$P[Y \leq 2]$$

$$= \sum_{y=1}^{2} \frac{2(5-y+1)}{5(6)} = \frac{2}{30}(5+4) = 3/5$$

Exercise 5.1 (Q. 9)

(9)An engineer is studying early morning traffic pattern at a particular intersection. The observation period begins at 5.30 a.m. Let X denote the time of arrival of the first vehicle from the north – south direction. Let Y denote the first arrival time from the

Exercise 5.1 (Q. 9)

east – west direction. Time is measured in fractions of an hour after 5.30 a.m. Assume that density for (X,Y) is given by

$$f_{XY}(x,y) = \frac{1}{x}, 0 < y < x < 1$$

(a) Verify that this is a valid density for a two dimensional random variable.

Exercise 5.1 (Q. 9)

$$(a)\int_0^1\int_0^x \frac{1}{x}\,dy\,dx = \int_0^1 (1/x)[y]_{y=0}^{x} = 1$$

(b) $Find\ \ P[X \le 0.5 \text{ and } Y \le 0.25]$

$$Sol:\ P[X \le 0.5\ and\ Y \le 0.25]$$

$$= \int_0^{0.25}\int_y^{0.5} \frac{1}{x}\,dx\,dy = 0.4233$$

Exercise 5.1 (Q. 9)

$$(d)\ Find\ \ P[X > 0.5 \text{ and } Y \geq 0.5]$$

$$= \int_{0.5}^{1} \int_{y}^{1} \frac{1}{x}\, dx\, dy = 0.15345$$

Examples: Continuous Joint Densities

Ex.5.4. Let $f(x,y)=\begin{cases} x^2+\dfrac{xy}{3}; & 0\le x\le 1,\ 0\le y\le 2 \\ 0 & ;\text{ e.w.} \end{cases}$

(i) Find $P\left(X>\dfrac{1}{2}\right)$ (ii) $P(Y<X)$

Examples: Continuous Joint Densities

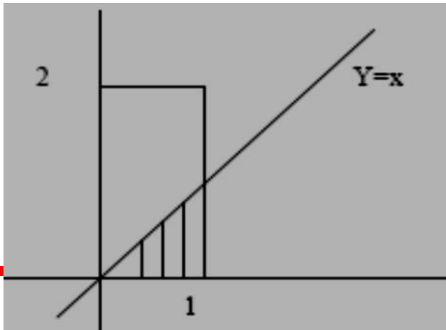
innovate achieve lead

Sol. (i) $P\left(X > \frac{1}{2}\right) = \int_{1/2}^{1}\int_{0}^{2} f(x,y)\,dydx = \int_{1/2}^{1}\int_{0}^{2}\left(x^2 + \frac{xy}{3}\right)dydx$

$$= \int_{1/2}^{1}\left[x^2y + \frac{xy^2}{6}\right]_{y=0}^{y=2} dx = \int_{1/2}^{1}\left(2x^2 + \frac{2x}{3}\right)dx = \left[\frac{2}{3}x^3 + \frac{x^2}{3}\right]_{x=1/2}^{x=1} = \frac{5}{6}$$

(ii) $P(Y < X) = \int_{0}^{1}\int_{0}^{x} f(x,y)\,dydx = \int_{0}^{1}\int_{0}^{x}\left(x^2 + \frac{xy}{3}\right)dydx$

$$= \int_{0}^{1}\left[x^2y + \frac{xy^2}{6}\right]_{y=0}^{y=x} dx = \int_{0}^{1}\left(x^3 + \frac{x^3}{6}\right)dx = \left[\frac{7x^4}{24}\right]_{0}^{1} = \frac{7}{24}$$



Examples: Continuous Joint Densities

HW5.3. Let $f(x,y)=\begin{cases} k(6-x-y);\ 0\le x\le 2,\ 2\le y\le 4 \\ 0 \qquad ;\ \text{e.w.} \end{cases}$

(i) Find k

(ii) Find $f_X(x),\ f_Y(y)$

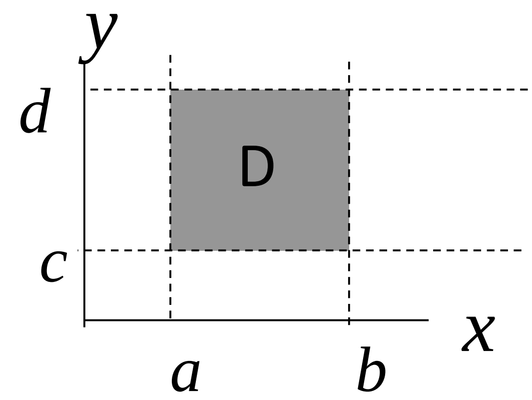
(iii) $P(Y<3)$

(iv) $P(X<1\cap Y<3)$

(v) $P(X<1 \mid Y<3)$

Examples: Continuous Joint Densities

Hint. (i) $\int_0^2 \int_2^4 f(x,y)\,dy\,dx = 1 \Rightarrow k = 1/8$



(ii) $f_X(x) = \int_2^4 f(x,y)\,dy$ where $0 \le x \le 2$;

$f_Y(y) = \int_0^2 f(x,y)\,dx$ where $2 \le y \le 4$

(iii) $P(Y < 3) = \int_2^3 f_Y(y)\,dy = \frac{5}{8}$,

(iv) $P(X < 1 \cap Y < 3) = \int_0^1 \int_2^3 f(x,y)\,dy\,dx = \frac{3}{8}$

(v) $P(X < 1 \mid Y < 3) = \frac{P(X < 1 \cap Y < 3)}{P(Y < 3)} = \frac{3}{5}$

Examples: Continuous Joint Densities

HW5.4. If $f(x,y)=\begin{cases} kx^2y; \ 0<x<1, 0<y<1 \\ 0 \qquad\qquad ; \text{e.w.} \end{cases}$ is a pdf,

(i) Find k $\left(\text{Sol. } k=6\right)$

(ii) Find $f_X(x)$, $f_Y(y)$

(iii) $P(X>Y)$ (iv) $P\left(X<\frac{1}{2}\,\middle|\,Y<\frac{1}{2}\right)$

Examples: Continuous Joint Densities

HW5.5. If $f(x,y)=\begin{cases} k(3x+y); 1<x<3, 0<y<2 \\ 0 \qquad\qquad ;\text{ e.w.} \end{cases}$ is a pdf,

(i) Find k (Sol. $k=1/28$)

(ii) Find $f_X(x),\ f_Y(y)$

(iii) Are X and Y independent?

Examples: Continuous Joint Densities

HW5.6. If $f(x, y) = \begin{cases} 2k; & 0 < x < 1, 0 < y < x \\ 0 & ; \text{e.w.} \end{cases}$ is a pdf,

(i) Find k (Sol. $k = 1$)

(ii) Find $f_X(x)$, $f_Y(y)$.

Problem Solving

HW.5.7. For the following pdfs, check whether X and Y are independent

(i) $f(x,y)=\begin{cases}\dfrac{x\left(1+3y^2\right)}{4}; & 0\le x\le 2,\ 0\le y\le 1\\ 0 & ;\ \text{e.w.}\end{cases}$

(ii) $f(x,y)=\begin{cases}2; & 0<x<1, 0<y<x\\ 0 & ;\ \text{e.w.}\end{cases}$

(iii) $f(x,y)=\begin{cases}6x^2y; & 0<x<1, 0<y<1\\ 0 & ;\ \text{e.w.}\end{cases}$

(iv) $f(x,y)=\begin{cases}\dfrac{1}{8}(6-x-y); & 0\le x\le 2,\ 2\le y\le 4\\ 0 & ;\ \text{e.w.}\end{cases}$

Problem Solving (HW 5.8)

innovate achieve lead

Food Safety and Standards Authority of India (FSSAI) has conducted an evaluation of 300 restaurants in the National Capital Region (NCR) Delhi. Each restaurant received a rating on a 3-point scale on typical meal price (1 least expensive to 3 most expensive) and quality (1 lowest quality to 3 greatest quality). A cross-tabulation of the rating data is provided below.

	Meal Price (y)			
Quality (x)	**1**	**2**	**3**	**Total**
1	42	39	3	**84**
2	33	63	54	**150**
3	3	15	48	**66**
Total	**78**	**117**	**105**	**300**

(a) Develop a bivariate probability distribution for quality and meal price of a randomly selected restaurant in the NCR Delhi. Assume that X and Y are the respective random variables corresponding to quality rating and meal price.

(b) Compute the expected value and variance for quality rating, X.

(c) Compute the expected value and variance for meal price, Y.

(d) Compute the mean and variance of $(X+Y)$. Hence, compute the covariance and the correlation coefficient between quality and meal price.

(e) Using above results, do you suppose it is very likely to find a low cost restaurant in the NCR Delhi that is also high quality? [12]

Problem Solving

innovate achieve lead

Ex. 5.5: Suppose that the two dimensional continuous random variable (X,Y) has joint pdf given by:

$$f(x,y)=\begin{cases} x^2+\dfrac{xy}{3}, 0\le x\le 1, 0\le y\le 2 \\ 0, \ \text{elsewhere} \end{cases}$$

Sol.

Check that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)$$

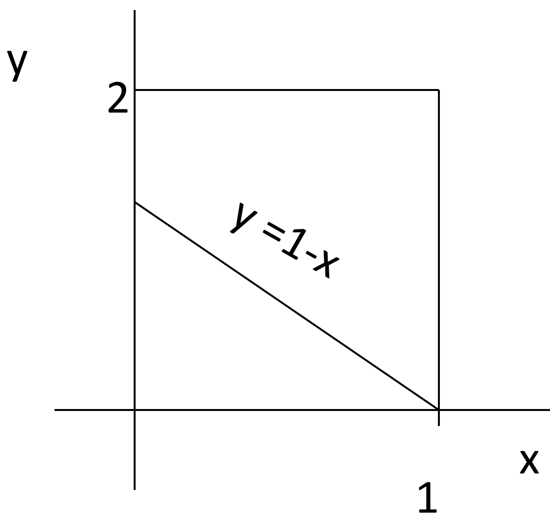
$$=\int_{0}^{2}\int_{0}^{1}(x^2+\frac{xy}{3})dxdy$$

$$= 1$$

Sol.

Let B={X+Y≥1}, compute P(B).

$$P(B) = 1 - \int_0^1 \int_0^{1-x} \left(x^2 + \frac{xy}{3}\right) dy\, dx = \frac{65}{72}$$



P[B]= 1-P(B′), where
B′ = (X+Y<1}, Hence

innovate achieve lead

BITS Pilani
Pilani Campus

Course No: MATH F113

Probability and Statistics

Cumulative distribution function

Def: Let (X,Y) be a two dimensional random variable. The cdf F of the two dimensional random variable (X,Y) is defined by:

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

Remark: Function of two variables and has a number of properties analogous to one dimensional random variable.

CDF and PDF

If $F(x,y)$ is the cdf of a two dimensional random variable with joint pdf, $f(x,y)$,

then $\frac{\partial^2}{\partial x\,\partial y}F(x,y) = f(x,y)$,

wherever F is differentiable.

(a) $F_X(x) = F(x,\infty)$ and $F_Y(y) = F(\infty,y)$

(b) X and Y are independent

$$\Leftrightarrow F(x,y) = F_X(x)F_Y(y)$$

Multinomial Law

Ex 5.6: A die is rolled 10 times in succession. Find the probability of the occurrence of six 4 times, five twice, and all other faces once each. Ans: $\dfrac{10!}{4!\ 2!\ 1!1!1!\ 1!}\left(\dfrac{1}{6}\right)^{10}$

Ex 5.7:

An urn contains $N = N_1 + N_2 + \cdots + N_m$ balls; $N_i : i^{\text{th}}$ color

With replacement, $n = i_1 + i_2 + \cdots + i_m$ balls are picked.

Find $\mathrm{P}\left(X_1 = i_1, X_2 = i_2, \cdots, X_m = i_m\right)$

$$= \frac{n!}{i_1!\ i_2! \cdots i_m!}\left(\frac{N_1}{N}\right)^{i_1}\left(\frac{N_2}{N}\right)^{i_2}\cdots\left(\frac{N_m}{N}\right)^{i_m}$$

Book Exercise 5.1 (Q. 12)

Ex. 12: Items coming off an assembly line are classified as being either non-defective, defective but salvageable, or defective but non-salvageable. The probabilities of observing items in each of these categories are 0.90, 0.08 and 0.02 respectively. The probabilities do not change from trial to trial. Twenty items are randomly selected and classified. Let X_1 denote the number of non-defective items, X_2 the number of non-defective but salvageable and X_3 the number of non-defective but non-salvageable items obtained.

Exercise 5.1 (Q. 12)

(a)Find P[X_1=15,X_2=3,X_3=2].

(b)Find the general formula for joint density of (X_1, X_2, X_3).

$$\textbf{\textit{Soln}}:(a)\frac{20!}{(15!)(3!)(2!)}(0.90)^{15}(0.08)^{3}(0.02)^{2}$$

$$(b)f(x_1,x_2,x_3)=\frac{n!}{n_1!\ n_2!\ n_3!}p_1^{n_1}p_2^{n_2}p_3^{n_3}$$

(This is called multinomial distribution).

Expectation

Let (X, Y) be a two-dimensional RV with joint density f_{XY} and marginal densities $f_X,\ f_Y,$ respectively.

Expectation of $H(X,Y)$, denoted by $E\left[H(X,Y)\right]$, is given as

$$E\left[H(X,Y)\right]=\sum_{\text{all } x}\sum_{\text{all y}} H(x,y)\, f_{XY}(x,y), \text{ if } (X,Y) \text{ is discrete RV.}$$

provided $\sum_{all\ x}\sum_{all\ y} |H(x,y)|\ f_{XY}(x,y)$ exists.

$$E\left[H(X,Y)\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,y)\, f_{XY}(x,y)dydx, \text{ if } (X,Y) \text{ is continuous RV}$$

provided $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |H(x,y)|\ f_{XY}(x,y)\ dy\,dx$ exists

Expectation: Properties

(a) $E(c) = c;\ c$ constant

(b) $E[aX + bY + c] = aE(X) + bE(Y) + c$

(c) $E[a_0 + a_1 X_1 + \ldots + a_n X_n]$

$$= a_0 + a_1 E[X_1] + \ldots + a_n E[X_n]$$

If two random variables X and Y are independent,

$E(XY) = E(X)E(Y)$ (Verify!)

Proof (for continuous case)

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \; f_{XY}(x,y)\,dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \; f_X(x)\; f_Y(y)\; dxdy$$

$$= \int_{-\infty}^{\infty} x \; f_X(x) \; \int_{-\infty}^{\infty} y \, f_Y(y)\; dy \; dx = \int_{-\infty}^{\infty} x \; f_X(x) \; E[Y] \; dx$$

$$= E[Y]\int_{-\infty}^{\infty} x \; f_X(x)\; dx = E[X]\,E[Y]$$

Mean and Variance

Let (X,Y) be a two-dimensional RV with joint density f_{XY} and marginal densities f_X, f_Y, respectively.

$$\mu_X = E(X) = \sum_{\text{all } x}\sum_{\text{all y}} x f_{XY}(x,y) = \sum_x x f_X(x), \text{ if } (X,Y) \text{ is discrete RV.}$$

$$\mu_Y = E(Y) = \sum_{\text{all } x}\sum_{\text{all y}} y f_{XY}(x,y) = \sum_y y f_Y(y), \text{ if } (X,Y) \text{ is discrete RV.}$$

$$\mu_X = E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)\,dy\,dx = \int_{-\infty}^{\infty} x f_X(x)\,dx, \text{ if } (X,Y) \text{ is continuous RV}$$

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{XY}(x,y)\,dy\,dx = \int_{-\infty}^{\infty} y f_Y(y)\,dy, \text{ if } (X,Y) \text{ is continuous RV}$$

Similarly,

$$\sigma_X^2 = E(X-\mu_x)^2 = \sum_{\text{all } x}\sum_{\text{all y}} (x-\mu_X)^2 f_{XY}(x,y) = \sum_{\text{x}} (x-\mu_X)^2 f_X(x);\ (X,Y) \text{ discrete}$$

$$\sigma_X^2 = E(X-\mu_x)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)^2 f_{XY}(x,y)\,dy\,dx = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\,dx;\ (X,Y) \text{ continuous}$$

Examples

HW5.9. For a discrete random variable (X,Y), $P(X=i,Y=j)=0,\ i,j=-1,0,1$ except $P(X=0,Y=1)=\frac{1}{3}, P(X=1,Y=-1)=\frac{1}{3}, P(X=1,Y=1)=\frac{1}{3}$.

Find $E(X), E(Y), E(X+3Y), \mathrm{Var}(X), \mathrm{Var}(Y), \mathrm{Var}(7X+9)$

HW5.10. Two balls are selected at random from a box that contains 3 blue balls, 2 red balls and 3 green balls. If X is the number of blue balls and Y is the number of red balls selected, then

Find (i) $f(x,y)$ (ii) $f_X(x),\ f_Y(y)$, (iii) $E(X), E(X^2), E(Y), E(Y^2)$, $E(XY)$ (iv) $\mathrm{Var}(-2X)$, $\mathrm{Var}(-2Y+5)$

innovate achieve lead

Examples

HW5.11 $f(x,y)=\begin{cases} k;\ 0<y<x<1 \\ 0 \qquad\qquad ;\ \text{e.w.} \end{cases}$

Find (i) k (ii) $f_X(x),\ f_Y(y)$,

(iii) $E(X), E(X^2), E(Y), E(Y^2), E(XY)$

(iv) $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$, $\mathrm{Var}(XY+5)$

Variance Properties

Ex.5.8. Show that $\text{Var}(X+Y) \neq \text{Var}(X) + \text{Var}(Y)$, in general

Sol. $\text{Var}(X+Y) = E\left[(X+Y-\mu_X-\mu_Y)^2\right]$

$$= E\left[\left((X-\mu_X)+(Y-\mu_Y)\right)^2\right]$$

$$= E\left[(X-\mu_X)^2+(Y-\mu_Y)^2+2(X-\mu_X)(Y-\mu_Y)\right]$$

$$= E(X-\mu_X)^2+E(Y-\mu_Y)^2+2E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

$$\neq \text{Var}(X)+\text{Var}(Y).$$

Here $E\left[(X-\mu_X)(Y-\mu_Y)\right]$ is called the covariance.

Covariance

- It is a measure (absolute) of how much two variables change together; e.g., height and weight, income-expenditure, age-blood pressure, advertisement-sales, demand-supply, fertilizer-yield.
- If two variables tend to show similar behaviour, then the covariance is positive, otherwise negative.
- The sign of the covariance shows the tendency in the **linear relationship** between the variables.
- The magnitude of covariance does not really produce a fruitful meaning.

Covariance

Covariance provides a measure of joint variability.

If X, Y tend to show similar behaviour, then the covariance is +ve, else it is -ve.

Let X, Y be RVs with means μ_X, μ_Y, respectively.

$$\begin{aligned}\operatorname{Cov}(X,Y) &= E\left[\left(X-\mu_Y\right)\left(Y-\mu_Y\right)\right] \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

Covariance Properties

(i) If two random variables X and Y are independent, then $\text{Cov}\left(X,Y\right)=0$

(ii) $\text{Cov}\left(X,Y\right)=\text{Cov}\left(Y,X\right)$

(iii) $\text{Cov}\left(X,X\right)=\text{Var}\left(X\right)$

(iv) $\text{Cov}\left(cX,Y\right)=c\,\text{Cov}\left(X,Y\right)$, c is a constant

(v) $\text{Cov}\left(a+bX,Y\right)=b\,\text{Cov}\left(X,Y\right)$

(vi) $\text{Cov}\left(a+bX,c+dY\right)=bd\,\text{Cov}\left(X,Y\right)$

(vii) $\text{Var}\left[a_0+\sum_{i=1}^{n}a_iX_i\right]=\sum_{i=1}^{n}a_i^2\text{Var}[X_i]+2\sum_{i<j}a_ia_j\,\text{Cov}\left(X_i,X_j\right)$

(viii) $\text{Var}\left[a_0+\sum_{i=1}^{n}a_iX_i\right]=\sum_{i=1}^{n}a_i^2\text{Var}[X_i]$, if $X_1,\ldots,X_n$ are independent

Remark on Covariance

If X and Y are independent, then $\mathrm{Cov}(X,Y)=0$, but the converse is NOT true.

Let X be uniformly distributed in $[-1,1]$ and let $Y = \mathrm{X}^2$.

Now $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$

$= E(X^3) - E(X)E(\mathrm{X}^2)$

$= 0$

Problem Solving

Ex.5.9. Let X, Y be independent RVs with means $1,\ -3,$ and variances $4,\ 5,$ respectively. Then, find $E(5X+3Y-9), \mathrm{Var}(2X+4Y+5), \mathrm{Var}(X-2)$

Hint. $E(5X+3Y-9)=5E(X)+3E(Y)-9=??$

$\mathrm{Var}(2X+4Y+5)=2^2\mathrm{Var}(X)+4^2\mathrm{Var}(Y)$, as X, Y are independent

$\mathrm{Var}(X-2)=\mathrm{Var}(X)=4$

HW5.12. Let $f(x,y)=\begin{cases} kxy;\ 0\le x\le 1,\ 0\le y\le x \\ 0 \qquad ;\ \text{e.w.} \end{cases}$

(i) Find $k, E(X^2), E(Y^2), E(XY)$ (ii) $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$ (iii) $\mathrm{Cov}(X,Y)$

Problem Solving

HW5.13. $f(x,y)=\begin{cases} 1; \ |x|<1 \text{ and } |y|<\dfrac{x}{2} \\ 0 \qquad\qquad ; \text{ e.w.} \end{cases}$

Show that $\text{Cov}(X,Y)=0$, but X,Y are not independent.

HW5.14. Let $\text{Cov}(X,Y)=K$. Then, show that

$$\text{Cov}(aX+bY,cX+dY)=0, \text{ if } ac\sigma_x^2+(ad+bc)K+bd\sigma_y^2=0.$$

innovate achieve lead

BITS Pilani
Pilani Campus

Course No: MATH F113

Probability and Statistics

Note on Correlation

Sec 5.3 to be covered later. In Sec. 5.4, only conditional densities will be discussed.

Conditional Density Functions

Let (X, Y) be a RV with joint density f_{XY} and marginal densities f_X, f_Y

(i) The conditional density for X given $Y = y$, denoted by $f_{X|y}$, is given by

$$f_{X|y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}; \quad f_Y(y) > 0$$

(ii) The conditional density for Y given $X = x$, denoted by $f_{Y|x}$, is given by

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)}; \quad f_X(x) > 0$$

Remarks:

(i) The conditional densities $f_{X|y}$ and $f_{Y|x}$ are also density functions

Conditional Density Functions

Let (X,Y): RV, joint pdf f_{XY}, and marginal densities f_X, f_Y, respectively.

Notice that whenever X and Y are independent,

$$f_{X|y}(x) = f_X(x) \text{ and } f_{Y|x}(y) = f_Y(y)$$

Examples: Conditional Density

Ex.5.10. $f(x,y)=\begin{cases}\dfrac{x\left(1+3y^2\right)}{4} & ;\ 0\le x\le 2,\ 0\le y\le 1\\ 0 & ;\ \text{e.w.}\end{cases}$

(i) Find $f_X(x),\ f_Y(y),\ f_{X|y},\ f_{Y|x}$

(ii) Evaluate $P\left(1/4 < X < 1/2 \middle| Y = 1/3\right)$.

Examples: Conditional Density

innovate achieve lead

Sol. (i) $f_X(x)=\int_0^1 \frac{x(1+3y^2)}{4}dy=\frac{x}{2}, 0<x<2;$

$$f_Y(y)=\int_0^2 \frac{x(1+3y^2)}{4}dx=\frac{1+3y^2}{2}, 0<y<1$$

$$f_{X|y_0}(x)=\frac{f_{XY}(x,y_0)}{f_Y(y_0)}=\frac{x}{2}, 0<x<2$$

$$f_{Y|x_0}(y)=\frac{f_{XY}(x_0,y)}{f_X(x_0)}=\frac{1+3y^2}{2}, 0<y<1$$

(If $y_0 \notin [0,1]$, then $f_Y(y_0)=0$ hence the conditional density $f_{X|y_0}(x)$ is undefined for any x.)

(ii) $P(1/4<X<1/2 | Y=1/3)=\int_{1/4}^{1/2} f_{X|Y=1/3}(x)dx=\int_{1/4}^{1/2}\frac{x}{2}dx=\frac{3}{64}$

Examples: Conditional Density

HW 5.15. Let $f(x,y)=\begin{cases} 2; \ 0<x<1, 0<y<x \\ \quad 0 \qquad\qquad ; \text{ e.w.} \end{cases}$ is a pdf,

(i) Find $f_X(x), f_Y(y), f_{X|y}, f_{Y|x}$ and

(ii) evaluate $P\left(1/4<X<3/4 \middle| Y=1/2\right)$.

Examples: Conditional Density

Hint. (i) $f_X(x) = 2x, 0 < x < 1;\ f_Y(y) = 2(1-y), 0 < y < 1$

$$f_{X|y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{1-y}, y < x < 1;$$

$$f_{Y|x}(y) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1}{x}, 0 < y < x$$

Note that $f_{X|y}$ is a function of x, as y is a constant. Similar is the case for $f_{Y|x}$.

(ii) When $y = 1/2$, $f_{X|y}(x) = \dfrac{1}{1-(1/2)}, \dfrac{1}{2} < x < 1$, that is, $f_{X|Y=1/2}(x) = 2; \dfrac{1}{2} < x < 1$

$$P(1/4 < X < 3/4 | Y = 1/2) = \int_{1/4}^{3/4} f_{X|Y=1/2}(x)\,dx = \int_{1/2}^{3/4} 2\,dx = \frac{1}{2}$$

Conditional Density Functions

innovate achieve lead

HW5.16. For the following pdfs, use conditional density functions to check whether X and Y are independent

(i) $f(x,y)=\begin{cases} x^2+\dfrac{xy}{3}, 0\le x\le 1, 0\le y\le 2 \\ 0 \qquad ;\text{ e.w.} \end{cases}$

(ii) $f(x,y)=\begin{cases} \sin x \sin y;\ 0<x<\dfrac{\pi}{2}, 0<y<\dfrac{\pi}{2} \\ 0 \qquad ;\text{ e.w.} \end{cases}$

(iii) $f(x,y)=\begin{cases} 6x^2y;\ 0<x<1, 0<y<1 \\ 0 \qquad ;\text{ e.w.} \end{cases}$

Exercise 5.4 (Q. 54, Page: 188)

An electronic device is designed to switch house lights on and off at random times after it has been activated. Assume that the device is designed in such a way that it will be switched on and off exactly once in a 1-hour period.

Exercise 5.4 (Q. 54, Page: 188)

Let Y denote the time at which the lights are turned on and X the time at which they are turned off. Assume that the joint density for (X,Y) is given by

$$f_{XY}(x, y) = 8xy, \quad 0 < y < x < 1$$

Exercise 5.4 (Q. 54, Page: 188)

(g) Find the conditional distribution of X given Y

(h) Find the probability that the lights will be switched off within 45 minutes of the system being activated given that they were switched on 10 minutes after the system was activated.

Exercise 5.4 (Q. 56, Page: 189)

Let X denote the number of "do loops" in a Fortran program and Y the number of runs needed to debug the program. Assume that the joint density for (X,Y) is given as:

x/y	1	2	3	4
0	0.059	0.100	0.050	0.001
1	0.093	0.120	0.082	0.003
2	0.065	0.102	0.100	0.010
3	0.050	0.075	0.070	0.020

Exercise 5.4 (Q. 56, Page: 189)

(d) Find the probability that a randomly selected program requires at least two runs to debug given that it contains exactly one "do loop"

Sol: $P[Y \geq 2 \mid X = 1] = \dfrac{P[Y \geq 2 \text{ and } X = 1]}{P[X = 1]} = \dfrac{0.205}{0.298} = 0.688$

(This can also be evaluated using $f_{Y|x}$)

$$P[Y \geq 2 \mid X = 1] = \sum_{y \geq 2} f_{Y|X=1}(y).$$

Problem Solving

HW 5.17: The joint density of *X* and *Y* is given:

$$f(x,y)=\begin{cases}\frac{1}{2}ye^{-xy}, & 0<x<\infty, 0<y<2\\ 0 & , \text{ elsewhere}\end{cases}$$

(1) Find *E[X|Y=1]*. (also denoted by $\mu_{X|y=1}$)
(this is the conditional mean of $f_{X\,|\,1}$)
(2) Find *P[X>2|Y=1]*.

Solution

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^{\infty} \frac{1}{2} ye^{-xy} dx = \frac{1}{2}; 0 < y < 2$$

$$\text{Hence } f_{X|Y=1}(x) = \begin{cases} \dfrac{f(x,1)}{f_Y(1)} = e^{-x}; & 0 < x < \infty \\ 0; & \text{elsewhere} \end{cases}$$

$$E[X \mid Y = 1] = \int_{-\infty}^{\infty} x f_{X|Y=1}(x) dx \text{ (Evaluate)}$$

Solution

$$P[X > 2 \mid Y = 1] = \int_{2}^{\infty} f_{X|Y=1}(x)\,dx$$

$$= \int_{2}^{\infty} e^{-x}\,dx \ \ (\text{evaluate})$$