ENM 207

Lecture 14

Joint Probability Distribution

Joint Probability Distribution

- Generally we must deal with two or more random variable simulatneously.
- For example, we might select fabricated vidget and measure
  - height
  - weight of it.
  - Thus, both height and weight are the random variables of interest.
- The objective of this subject is to
  - formulate joint probability distribution for two or more random variables
  - present methods for obtaining both marginal and conditional distributions.
  - present a definition of independence for random variables
  - define covariance and correlation

Joint Probability Distribution for Two R.Vs

- If the possible values of $[x_1,x_2]$ are either finite or countably infinite in number , then $[x_1,x_2]$ will be a two dimensional discrete random vector.

- If possible values of $[x_1,x_2]$ are some uncountable set in the Euclidean plane, then $[x_1,x_2]$ will be a two dimensional continuous random vector.

Joint Probability Distribution for Two R.Vs

Discrete Case

$$p(x_1, x_2) = p(X_1 = x_1, X_2 = x_2)$$

$$p(x_1, x_2) \geq 0 \quad and$$

$$\sum_{all\ j} \sum_{all\ i} p(x_1, x_2) = 1$$

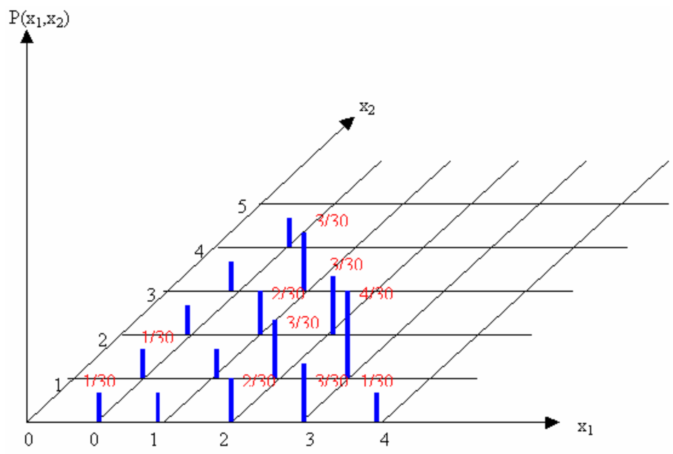
Joint Probability Distribution for Two R.Vs

Example:

x1		0	1	2	3	4	P(x2)
x2							
0		1/30	1/30	1/15	1/10	1/30	4/15
1		1/30	1/30	1/10	2/15		3/10
2		1/30	1/15	1/10			1/5
3		1/30	1/10				2/15
4		1/10					1/10
P(x1)		7/30	7/30	4/15	7/30	1/30	top p(x)=1

this is the tabular representation.
This means that $x_1=0$ and $x_2=0$ with probability 1/30.

Joint Probability Distribution for Two R.Vs



Graphical representation of a bivariate probability distribution

Joint Probability Distribution for Two R.Vs

- If [X_1,X_2] is a continuous random vector with range space R in the Euclidean plane , then the joint density function has the following properties.

$$f(x_1, x_2) \geq 0 \quad for\ all\ (x_1, x_2) \in R$$

$$\iint_R f(x_1, x_2)\, dx_1 dx_2 = 1 \qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1 dx_2 = 1$$

Joint Probability Distribution for Two R.Vs

2. Continuous Case

$$P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2) = \int_{a_2}^{b_2}\int_{a_1}^{b_1} f(x_1, x_2)\, dx_1 dx_2$$

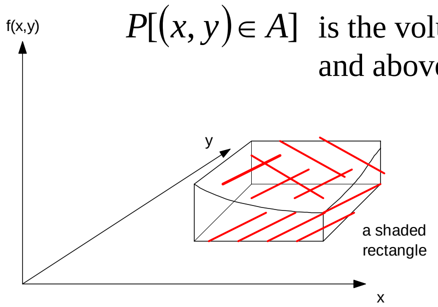
- It should again be noted that $f(x_1, x_2)$ doesn't represent the probability of anything , and the convention that

$$f(x_1, x_2) = 0 \quad for\ (x_1, x_2) \notin R$$

Joint Probability Distribution for Two R.Vs

We can think of f(x,y) as specifying a surface at height f(x,y) above the point (x,y) in a three-dimensional coordinate system.

$P[(x,y)\in A]$ is the volume under this surface and above the region A.



Joint Probability Distribution for Two R.Vs

Example A bank operates both a drive-up facility and walk up window. On a randomly selected day , let

x=the proportion of time that the drive up facility is in use (at least one customer is being served or waiting to be served)

y=the proportion of time that the walk up window is in use.

Then the set of possible values for (x,y) is the rectangle

$$D=\{(x,y): 0\le x\le 1, 0\le y\le 1\}$$

the joint p.d.f of (x,y) is given by

$$f(x,y)=\begin{cases}\frac{6}{5}\left(x+y^2\right) & 0\le x\le 1, 0\le y\le 1\\ 0 & o.w.\end{cases}$$

Joint Probability Distribution for Two R.Vs

a) Is this function a p.d.f ?

b) Find probability that neither facility is busy more than one quarter of time.

Joint Probability Distribution for Two R.Vs

Example

A nut company markets cans of deluxe mixed nuts containing almonds, and peanuts. Suppose the net weight of each can is exactly 1 pound, but the weight contribution of each type of nut is random. Let ;

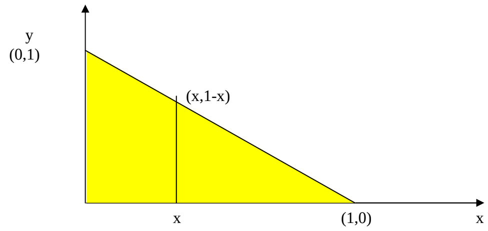
X=the weight of almonds in a selected can
Y=the weight of cashews

Then the region of positive density is $R = \{(x, y): 0 \le x \le 1, 0 \le y \le 1, x + y \le 1\}$

$$f(x, y) = \begin{cases} 24xy & 0 \le x \le 1, 0 \le y \le 1,\ x + y \le 1 \\ 0 & o.w \end{cases}$$

Joint Probability Distribution for Two R.Vs

The shaded region pictured in figure is as follows.



Show that this function is p.d.f.

Joint Probability Distribution for Two R.Vs

Example

$$f(x_1, x_2) = \begin{cases} \dfrac{1}{500} & 0 \le x_1 \le 0.25 \quad , \quad 0 \le x_2 \le 200 \\ 0 & o.w. \end{cases}$$

Compute $P(0.1 \le x_1 \le 0.2 \quad , \quad 100 \le x_2 \le 200)$

Marginal Distributions

- Having defined the bivariate probability distribution, called the the joint probability distribution (or in the continuous case the joint density) , a natural question arises as to the distribution of x_1 or x_2 alone. These distributions are called "marginal disributions".

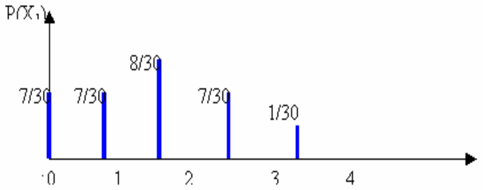
- In discrete case , the marginal distribution of X_1 is;

$$P(x_1) = \sum_{all\ j} p(x_1, x_2) \quad i = 1,2,....$$

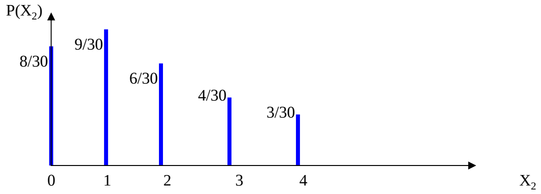
- the marginal distribution of X_2 is

$$P(x_2) = \sum_{all\ i} p(x_1, x_2) \quad j = 1,2,3...$$

Marginal Distributions



Graphical representation of marginal distribution for X_1



Graphical representation of marginal distribution for X_2

Marginal Distributions

In continuous case, the marginal distribution of X_1 is

$$f(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$$

the marginal distribution of X_2 is

$$f(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$$

Marginal Distributions

Example

$$f(x,y) = \begin{cases} \frac{6}{5}\left(x + y^2\right) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & o.w. \end{cases}$$

a) Find marginal p.d.f. of x, which gives the probability distribution of busy time for the drive-up facility without reference to the walk-up window

Marginal Distributions

c) Find the marginal p.d.f of Y

d)find probability that $P\left(\frac{1}{4} \le y \le \frac{3}{4}\right)$

Expected Value and Variance

1. Discrete Case

$$E(x_1) = \mu_1 = \sum_{all\ i} x_1 p(x_1) = \sum_{all\ i}\sum_{all\ j} x_1 p(x_1, x_2)$$
$$= \sum_{all\ i} x_1 \sum_{all\ j} p(x_1, x_2)$$
$$= \sum_{all\ i} x_1 p(x_1)$$

$$v(x_1) = \sigma_1^{\ 2} = \sum_{all\ i} (x_1 - \mu_1)^2 p(x_1) = \sum_{all\, i}\sum_{all\ j} (x_1 - \mu_1)^2 p(x_1, x_2)$$
$$= \sum x_1^2 p(x_1) - \mu_1^{\ 2} = \sum_i \sum_j x_1^{\ 2} p(x_1, x_2) - \mu_1^{\ 2}$$

Expected Value and Variance

$$E(x_2) = \mu_2 = \sum_j x_2 p(x_2) = \sum_j \sum_i x_2 p(x_1, x_2)$$

$$V(x_2) = \sigma_2^2 = \sum_j (x_2 - \mu_2)^2 p(x_2) = \sum_j \sum_i (x_2 - \mu_2)^2 p(x_1, x_2)$$

$$= \sum x_2^2 p(x_2) - \mu_2^2 = \sum_j \sum_i x_2^2 p(x_1, x_2) - \mu_2^2$$

Expected Value and Variance

Example.

The mean and variance of x_2 could be determined using the marginal distribution of x_2.

$$E(x_1) = \mu_1 = 0.\frac{7}{30} + 1.\frac{7}{30} + 2.\frac{8}{30} + 3.\frac{7}{30} + 4.\frac{1}{30} = \frac{8}{5}$$

$$V(x_1) = \sigma_1^2 = [0^2.\frac{7}{30} + 1^2.\frac{7}{30} + 2^2.\frac{8}{30} + 3^2.\frac{7}{30} + 4^2.\frac{1}{30}]^2 - [\frac{8}{5}]^2 = \frac{103}{75}$$

Expected Value and Variance

2. Continuous Case

$$E(x_1) = \mu_1 = \int_{-\infty}^{\infty} x_1 f(x_1) dx_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1, x_2) dx_2 dx_1$$

$$V(x_1) = \sigma_1^2 = \int_{-\infty}^{\infty} (x_1 - \mu_1)^2 f(x_1) dx_1$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1 - \mu_1)^2 f(x_1, x_2) dx_2 dx_1 = \int_{-\infty}^{\infty} x_1^2 f(x_1) dx_1 - \mu_1^2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1^2 f(x_1, x_2) dx_2 dx_1 - \mu_1^2$$

Expected Value and Variance

$$E(x_2) = \mu_2 = \int_{-\infty}^{\infty} x_2 f(x_2) dx_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_1, x_2) dx_1 dx_2$$

$$V(x_2) = \sigma_2^2 = \int_{-\infty}^{\infty} (x_2 - \mu_2)^2 f(x_2) dx_2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_2 - \mu_2)^2 f(x_1, x_2) dx_1 dx_2$$

$$= \int_{-\infty}^{\infty} x_2^2 f(x_2) dx_2 - \mu_2^2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2^2 f(x_1, x_2) dx_1 dx_2 - \mu_2^2$$

Expected Value and Variance

Example:

$$f(x_1, x_2) = \begin{cases} \dfrac{1}{500} & 0 \le x_1 \le 0{,}25 \ , \ 0 \le x_2 \le 2000 \\ 0 & o.w. \end{cases}$$

a) Find marginal distribution for X_1
b) Compute $E(X_1)$ and $Var(X_1)$

Conditional Distribution

Let *X* and *Y* be two random variables, discrete or continuous. The conditional distribution of the random variable Y, given that *X* = *x* is

$$f(y \mid x) = \frac{f(x, y)}{f(x)}, \qquad f(x) > 0.$$

Similarly, the conditional distribution of the random variable *X*, given that *Y*=*y*, is

$$f(x \mid y) = \frac{f(x, y)}{f(y)}, \qquad f(y) > 0.$$

Conditional Distribution

- Example

Referring to example for the discrete case, find conditional distribution of X_1, given that $X_2 = 1$, and use it to determine $P(X_1 = 0 \mid X_2 = 1)$.

Conditional Distribution

Example

The joint density for the random variables (X,Y), where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic article produces is

$$f(x,y) = \begin{cases} 10xy^2 & 0 < x < y < 1 \\ 0, & \text{o.w.} \end{cases}$$

a) Find the marginal densities f(x), f(y) and the conditional density f(y|x).
b) Find the probability that spectrum shifts more than half of the total observations, given the temperature is increased to 0.25 unit.

Independent Random Variabes

Two random variables *X* and *Y* are said to be independent if for every pair of *x* and *y* values ,

$p(x,y) = p(x)p(y)$ when *X* and *Y* are discrete.

$f(x,y) = f(x)f(y)$ when x and y are continuous.

If this equation is not satisfied for all (*x*,*y*) then *x* and *y* are said to be dependent.

Independent Random Variabe

Example

Referring to example for the discrete case, show that random variables are not statistically independent.

$$p(0,1) = \frac{1}{30}$$

Marginal distribution for X_1

$$p(0) = \sum_{x_2=0}^{4} p(0,x_2) = \frac{1}{30} + \frac{1}{30} + \frac{1}{30} + \frac{1}{30} + \frac{1}{10} = \frac{7}{30}$$

Marginal distribution for X_2

$$p(1) = \sum_{x_1=0}^{4} p(x_1,1) = \frac{1}{30} + \frac{1}{30} + \frac{1}{10} + \frac{2}{15} = \frac{3}{10}$$

$$p(0,1) \neq p(0)\,p(1)$$

X and Y are not statistically independent.

Independent Random Variabes

Example

$$f(x,y)=\begin{cases}10xy^2 & 0<x<y<1\\ 0, & \text{o.w.}\end{cases}$$

Show that X and Y are not statistically independent.

Covariance

- When two random variables *X* and *Y* are not independent, it is frequently of interest to measure how strongly they are related to one another.
- The covariance between two r.v. *X* and *Y* is;

$$\operatorname{cov}(x,y)=E(x-\mu_x)(y-\mu_y)=$$

$$\begin{Bmatrix}\sum\sum(x-\mu_x)(y-\mu_y)\,p(x,y) & x,y \quad \textit{discrete}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)\,f(x,y)dxdy & x,y \quad \textit{continuous}\end{Bmatrix}$$

$$\operatorname{cov}(x,y)=E(xy)-\mu_x\ \mu_y=E(xy)-E(x)E(y)$$

Independent Random Variabes

Correlation:

The correlation coefficent of X and Y denoted by Corr(X,Y) , ρ_{xy} or just only ρ is denoted by ;

$$\rho_{xy} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

Independent Random Variabes

Example

The fraction X of male runners and the fraction Y of female runners who compete in marathon races is described by the joint density function

$$f(x,y) = \begin{cases} 8xy & 0 \le x \le 1,\ 0 \le y \le x \\ 0, & \text{o.w.} \end{cases}$$

Find covariance of X and Y

References

- Walpole, Myers, Myers, Ye, (2002),
  - *Probability & Statistics for Engineers & Scientists*
- Dengiz, B., (2004),
  - *Lecture Notes on Probability,* http://w3.gazi.edu.tr/web/bdengiz
- Hines, Montgomery, (1990),
  - *Probability & Statistics in Engineering* & Management *Science*