Lecture 8: Joint Probability Distributions

MSU-STT-351-Sum-19B

Joint Probabilty Distributions

Earlier, we discussed how to display and summarize the data $x_1, \ldots, x_n$ on a variable X. Also, we discussed how to describe the population distribution of a random variable X through *pmf* or *pdf*. We now extend these ideas to the case where $X = (X_1, X_2, \ldots, X_p)$ is a random vector and we will focus mainly for the case $p = 2$.

First, we introduce the joint distribution for two random variables or characteristics X and Y.

1. Discrete Case:

Let X and Y be two discrete random variables. For example,

X=number of courses taken by a student.

Y=number of hours spent (in a day) for these courses.

Our aim is to describe the joint distribution of X and Y.

Definition:

(a) The joint distribution of X and Y (both discrete) is defined by

$$p(x, y) = P(X = x; Y = y),$$

satisfying (i) $p(x, y) \geq 0$; (ii) $\sum_{x,y} p(x, y) = 1$.

(b) Also,

$$p(x) = P(X = x) = \sum_y p(x, y); \quad p(y) = P(Y = y) = \sum_x p(x, y)$$

are respectively called the **marginal** distributions of X and Y.

(c) The mean (or the expected value) of a function $h(X, Y)$ is

$$\mu_{h(x,y)} = E(h(X, Y)) = \sum_x \sum_y h(x, y)p(x, y).$$

Also, $P((X, Y) \in A) = \sum_{(x,y)\in A} p(x, y)$, where $A \subseteq R^2$.

Example 1: The joint distribution of $p(x, y)$ of X (number of cars) and Y (the number of buses) per signal cycle at a traffic signal is given by

		y		
p(x,y)		0	1	2
x	0	0.025	0.015	0.010
	1	0.050	0.030	0.020
	2	0.125	0.075	0.050
	3	0.150	0.090	0.060
	4	0.100	0.060	0.040
	5	0.050	0.030	0.020

(a) Find $P(X = Y)$.
(b) Find the marginal distribution of X and Y.
(c) Suppose a bus occupies **three** vehicle spaces and a car occupies just one. What is the mean number of vehicle spaces occupied during a signal cycles? That is, find the mean or expected value of $h(X, Y) = X + 3Y$.

Solution:
(a) The number of cars equals the number of buses if $X = Y$. Hence,
$P(X = Y) = p(0,0) + p(1,1) + p(2,2) = .025 + .030 + .050 = .105.$
That is, about 10.5% of the time.

(b) Adding the row values yields the marginal distribution of the x values:

x	0	1	2	3	4	5
p(x):	0.05	0.1	0.25	0.3	0.2	0.1

So, the mean number of cars is $\mu_x = E(X)$ is

$$\mu_x = 0(0.05) + 1(.10) + 2(.25) + 3(.30) + 4(.20) + 5(.10) = 2.8.$$

Similarly, adding the column values yields the marginal distribution of the y as:

y:	0	1	2
p(y):	0.50	0.30	0.20

(c) Let X and Y denote the number of cars and buses at a signal cycle. Then the number of vehicle spaces occupied is $h(X, Y) = X + 3Y$. Hence, the mean number of spaces occupied is then $E(h(X, Y)) = E(X + 3Y)$ is

$$\begin{aligned}
\mu_{h(x,y)} &= \sum\sum(x + 3y)p(x, y) \\
&= [0 + 3(0)](.025) + [0 + 3(1)](.015) + ... + [5 + 3(2)](.020) \\
&= 4.90.
\end{aligned}$$

Note that $E(h(X, Y)) = E(X + 3Y)$ can also be computed from the result $E(X + 3Y) = E(X) + 3E(Y)$, using the marginal distributions of X and Y.

2. Continuous Case

Bivariate Continuous Distributions

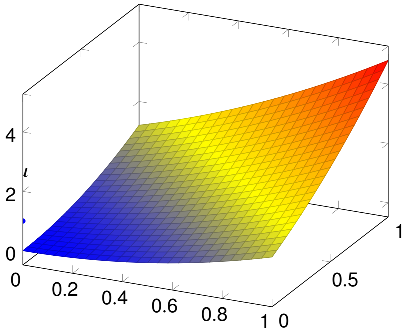
Definition: Let X and Y be continuous variables. The joint probability density of X and Y, denoted by $f(x, y)$, satisfies

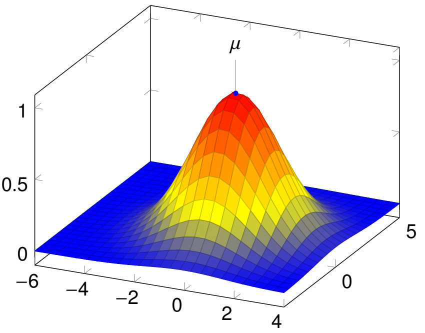
(i) $f(x, y) \geq 0$

(ii) $\int\int f(x, y)dxdy = 1.$

The graph $(x, y, f(x, y))$ is a surface in 3-dimensional space. The second condition shows the volume of this density surface is 1.

Joint Probability Distributions





Joint Probability Distributions

Properties (i) If X and Y are two continuous rvs with density $f(x, y)$, then

$$P[(X, Y) \in A] = \int \int_A f(x, y) dxdy,$$

which is the volume under density surface above A.
(ii) The marginal probability density functions of X and Y are respectively

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy; \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

(iii) The mean (expected value) of $h(x, y)$ is

$$\mu_{h(x,y)} = \int \int h(x, y) f(x, y) dxdy.$$

(iv) The mean functions μ_x and μ_y are defined as
$$\mu_x = \int x f_X(x) dx; \quad \mu_y = \int y f_Y(y) dy.$$

Example 1 (Eg. 5.3) Let X and Y have joint density

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

Then,

$$\begin{aligned} P(0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}) &= \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x+y^2)dxdy \\ &= \frac{6}{5}\int_0^{1/4}\left\{[\frac{x^2}{2} + y^2 x]_0^{\frac{1}{4}}\right\}dy \\ &= \frac{6}{5}\int_0^{1/4}\left(\frac{1}{32} + \frac{1}{4}y^2\right)dy \\ &= \frac{7}{640} = 0.0109. \end{aligned}$$

Example 2 (Ex.16)
Let X_1, X_2 and X_3 have density

$$f(x_1, x_2, x_3) = \begin{cases} k(x_1 x_2(1-x_3)), & 0 \le x_i \le 1; x_1 + x_2 + x_3 \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Compute the joint marginal density function of X_1 and X_3 alone.
(b) What is $P(X_1 + X_3 \le .5)$?
(c) Compute the marginal pdf of X_1 alone.
Solution: It can be seen that the value of $k = 144$.

$$\begin{aligned} (a): \; f(x_1, x_3) &= \int_{-\infty}^{\infty} f(x_1, x_2, x_3) dx_2 \\ &= \int_0^{1-x_1-x_3} k x_1 x_2 (1-x_3) dx_2 \\ &= 72 x_1 (1-x_3)(1-x_1-x_3)^2, \; 0 \le x_1, \; 0 \le x_3, \; x_1 + x_3 \le 1. \end{aligned}$$

Solution:
(b):

$$
\begin{aligned}
P(X_1 + X_3 \leq 0.5) &= \int_0^{0.5} \int_0^{0.5 - x_1} 72x_1(1 - x_3)(1 - x_1 - x_3)^2 dx_3 dx_1 \\
&= 0.53125
\end{aligned}
$$

(c):

$$
\begin{aligned}
f_{x_1}(x_1) &= \int_{-\infty}^{\infty} f(x_1, x_3) dx_3 \\
&= \int 72x_1(1 - x_3)(1 - x_1 - x_3)^2 dx_3 \\
&= 18x_1 - 48x_1^2 + 36x_1^3 - 6x_1^5, \ 0 \leq x_1 \leq 1.
\end{aligned}
$$

Conditional Distributions

(i) Let X and Y be **continuous** rvs with with the joint *pdf* $f(x, y)$ and the marginal *pdfs* $f_X(x)$ and $f_Y(y)$. Then the conditional probability density of Y, given $X = x$, is

$$f_{Y|X}(y|x) = f(y|x) = \frac{f(x, y)}{f_X(x)}, \ -\infty \leq y \leq \infty,$$

provided that $f_X(x) > 0$. Similarly, $f_{X|Y}(x|y)$ is defined.

(ii) Let X and Y be **discrete** rvs with joint *pmf* $p(x, y)$ and marginal *pmfs* $p_X(x)$ and $p_Y(y)$. Then the conditional pmf of Y given $X = x$ is

$$p_{Y|X}(y|x) = p(y|x) = \frac{p(x, y)}{p_X(x)},$$

provided that $p_X(x) > 0$. Similarly, $p(x|y)$ is defined.

Example 3 (Ex. 18) Let X denote the number of hoses being used on the self-service island at particular time, and let Y denote the number of hoses on the full-service island in use at that time, in a service station. The joint *pmf* of X and Y is given in the following table:

		y		
p(x,y)		0	1	2
	0	0.10	0.04	0.02
x	1	0.08	0.20	0.06
	2	0.06	0.14	0.30

(a) Given that $X = 1$, determine the conditional pmf of Y, that is, $p_{y|x}(0|1), p_{y|x}(1|1)$ and $p_{y|x}(2|1)$.

(b) Given that two hoses are in use at the self-service island. What is the conditional pmf of the number of hoses in use on the full-service island?

(c) Use the result of part (b) to calculate the conditional probability $P(Y \leq 1|X = 2)$.

(d) Given that two hoses are in use at the full-service island, what is the conditional *pmf* of the number in use at the self-service island?

Solution

(a) $p_{Y|X}(y|1)$ results from dividing each entry in $X = 1$ row of the joint probability table by $p_x(1) = 0.34$, we obtain

$$p_{y|x}(0|1) = \frac{0.08}{0.34} = 0.2353$$

$$p_{y|x}(1|1) = \frac{0.20}{0.34} = 0.5882$$

$$p_{y|x}(2|1) = \frac{0.06}{0.34} = 0.1765$$

(b) To obtain $p_{Y|X}(y|2)$, divide each entry in the $y = 2$ row by $p_X(2) = 0.5$:

y:	0	1	2
$p_{Y|X}(y|2)$:	0.12	0.28	0.6

(c) $P(Y \leq 1|x = 2) = p_{Y|X}(0|2) + p_{Y|X}(1|2) = 0.12 + 0.28 = 0.40$

(d) $p_{X|Y}(x|2)$ results from dividing each entry in the $y = 2$ column by $p_Y(2) = 0.38$:

x:	0	1	2
$p_{X\|Y}(x\|2)$:	0.0526	0.1579	0.7895

Example 4 (Ex. 19): Let X and Y (pressures for right and left tires) have density

$$f(x,y) = \begin{cases} K(x^2+y^2), & 20 \leq x, y \leq 30 \\ 0, & \text{otherwise;} \end{cases}$$

(a) Determine the conditional pdf of Y given that $X = x$ and the conditional pdf of X given that $Y = y$.
(b) If the pressure in the right tire is found to be 22 psi, what is the probability that the left tire has a pressure of at least 25 psi? Compare this with $P(Y \geq 25)$
(c) If the pressure in the right tire is found to be 22 psi, what is the expected pressure in the left tire, and what is the standard deviation of pressure in this tire?

Solution:

(a) $$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{k(x^2+y^2)}{10kx^2+0.05}, \quad 20 \le y \le 30.$$

$$f_{X|Y}(x|y) = \frac{k(x^2+y^2)}{10ky^2+0.05}, \quad 20 \le x \le 30. \ \left(k = \frac{3}{380000}\right)$$

$$\begin{aligned}(b):\ P(Y \ge 25|X=22) &= \int_{25}^{30} f_{Y|X}(y|22)dy \\ &= \int_{25}^{30} \frac{k((22)^2+y^2)}{10k(22)^2+0.05}dy \\ &= 0.783;\end{aligned}$$

Also,

$$P(Y \ge 25) = \int_{25}^{30} f_Y(y)dy = \int_{25}^{30}(10ky^2+0.05)dy = 0.75;$$

$$\begin{aligned}(c):\ E(Y|X=22) &= \int_{-\infty}^{\infty} y.f_{Y|X}(y|22)dy \\ &= \int_{20}^{30} y\frac{k((22)^2+y^2)}{10k(22)^2+0.05}dy \\ &= 25.372912.\end{aligned}$$

Note also that

$$E(Y^2|X=22) = \int_{20}^{30} y^2\frac{k((22)^2+y^2)}{10k(22)^2+0.05}dy = 652.028640.$$

$$Var(Y|X=22) = E(Y^2|X=22) - [E(Y|X=22)]^2 = 8.243976.$$

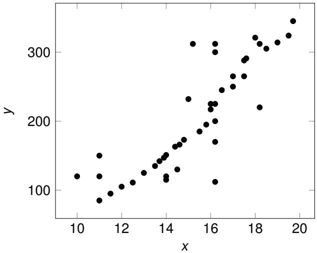
Covariance and Correlation:

Definition: The covariance between two random variables X and Y is

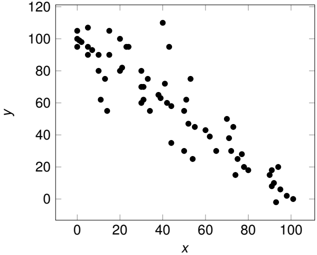
$$
\begin{aligned}
Cov(X,Y) &= E[(X-\mu_x)(Y-\mu_y)] \\
&= E(XY) - \mu_x\mu_y \\
&= \begin{cases} \sum_x \sum_y (x-\mu_x)(y-\mu_y)p(x,y), & \text{if } X \text{ \& } Y \text{ are discrete} \\ \\ \int\int (x-\mu_x)(y-\mu_y)f(x,y)\,dxdy, & \text{if } X \text{ \& } Y \text{ are continuous} \end{cases}
\end{aligned}
$$

Look at the following graphs:

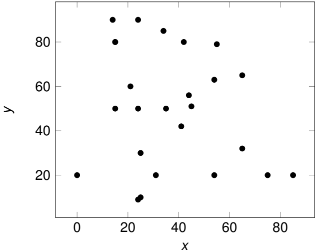
Joint Probability Distributions



(a) Positive correlation



(b) Negative correlation



(c) No correlation

Definition: The correlation coefficient between X and Y, denoted Corr(X,Y) or $\rho_{X,Y}$ or simply ρ is defined as (here σ_x denotes the SD of X)

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

Properties of $\rho_{x,y} = \rho$**:**
(i) For two rvs X and Y, $-1 \leq \rho \leq 1$
(ii) If a and c are either both positive or negative then

$$Corr(aX + b, cY + d) = Corr(X,Y)$$

(iii) If X and Y are independent, then $\rho = 0$. However, $\rho = 0$ does not imply that X and Y are independent.
(iv) $\rho = -1$ or $\rho = 1$ if and only if $Y = aX + b$ for some reals a and b.

Example 5 (Ex. 31): Let X and Y have joint density (with $K = 3/380000$)

$$f(x, y) = \begin{cases} K(x^2 + y^2), & 20 \leq x, y \leq 30 \\ 0, & \text{otherwise,} \end{cases}$$

(a) Compute the covariance and the ρ between X and Y.

Solution (a): By symmetry,

$E(X) = \int_{20}^{30} x f_X(x) dx = \int_{20}^{30} x[10Kx^2 + 0.05] dx = 25.329 = E(Y)$;

$E(XY) = \int_{20}^{30} \int_{20}^{30} xy.K(x^2 + y^2) dxdy = 641.447.$

$Cov(X, Y) = 641.447 - (25.329)^2 = -.111$

(b) $E(X^2) = \int_{20}^{30} x^2 [10Kx^2 + 0.05] dx = 649.8246 = E(Y^2)$;

$Var(X) = Vay(Y) = 649.8246 - (25.329)^2 = 8.2664.$

Hence,

$$\rho = \frac{-0.111}{\sqrt{(8.2664)(8.2664)}} = -0.0134.$$

Example 6 (Ex 35)
(a) Use the rules of expected value to show that

$$Cov(aX + b, cY + d) = acCov(X, Y).$$

(b) Use part (a) along with the rules of variance and standard deviation to show that $Corr(aX + b, cY + d) = Corr(X, Y)$ when a and c have the same sign.
Solution: (a):

$$\begin{aligned}
Cov(aX + b, cY + d) &= E[(aX + b)(cY + d)] - E(aX + b)E(cY + d) \\
&= E[acXY + adX + bcY + bd] \\
&\quad -(aE(X) + b)(cE(Y) + d) \\
&= acE(XY) - acE(X)E(Y) \\
&= acCov(X, Y)
\end{aligned}$$

(b):

$$
\begin{aligned}
Corr(aX+b, cY+d) &= \frac{Cov(aX+b, cY+d)}{\sqrt{Var(aX+b)}\sqrt{Var(cY+d)}} \\
&= \frac{acCov(X,Y)}{|a||c|\sqrt{Var(X)Var(Y)}} \\
&= Corr(X,Y),
\end{aligned}
$$

since a and c have the same signs.

(c) When a and c differ in sign,

$$Corr(aX+b, cY+d) = -Corr(X,Y).$$

Example 7 **Bivariate Normal Distribution.**

This is a very useful distribution in statistics and its joint density is

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\{-g(x,y)\},$$

where

$$g(x,y) = \frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}.$$

Here, $\rho = \rho_{x,y}$.

3. Independence

It is in general difficult to find the exact $p(x, y)$ or $f(x, y)$, from the marginal distributions. A simple approach is way is to assume **independence** between X and Y, which roughly means that knowing the values of X does not affect the values of Y (or vice-versa). This is an important concept in probability and statistics.

Definition: Two continuous random variables X and Y are **independent** if

$$f(x, y) = f_X(x)f_Y(y) \text{ for all } (x, y),$$

where for example $f_X(x)$ is the marginal density of X. Note under the independence, the joint distribution can be constructed from the marginal distributions. Similar definition applies to discrete random variables also. Equivalently, X and Y are **independent** if $f(x|y) = f(x)$ for all x and y.

If the above equation is not satisfied for some (x, y), then the variables are said to be **dependent**.

Remark: If X and Y are **independent**, then

$$P(a < X < b; c < Y < d) = P(a < X < b)P(c < Y < d).$$

Example 1 (revisited). Suppose the joint distribution of X (number of cars) and Y (the number of buses) per signal cycle at a particular left turn lane is given by

		y		
p(x,y)		0	1	2
x	0	0.025	0.015	0.010
	1	0.050	0.030	0.020
	2	0.125	0.075	0.050
	3	0.150	0.090	0.060
	4	0.100	0.060	0.040
	5	0.050	0.030	0.020

The marginal distributions are given by

x	0	1	2	3	4	5
p(x):	0.05	0.1	0.25	0.3	0.2	0.1

y:	0	1	2
p(y):	0.50	0.30	0.20

Note that
$p_{X,Y}(0,0) = p_X(0)p_Y(0) = 0.025;\;\; p_{X,Y}(0,1) = p_X(0)p_Y(1) = 0.015$ and etc. Indeed,

$$p_{X,Y}(x,y) = p_X(x)p_Y(y), \;\; \text{for all } (x,y).$$

Hence, X and Y are independent.

Example 8: Suppose the joint distribution of X and Y is

$$f(x,y) = \begin{cases} \lambda\mu e^{-\lambda x-\mu y}, & x>0, y>0 \\ 0, & \text{otherwise} \end{cases}$$

Note $f(x,y) = \lambda\mu e^{-\lambda x}\mu e^{-\mu y} = f_X(x)f_Y(y)$ for all (x,y). Hence, X and Y are independent.

Example 9: Let X and Y have density

$$f(x,y) = \begin{cases} xe^{-x(1+y)}, & x>0, y>0 \\ 0, & \text{otherwise} \end{cases}$$

(i) Find the marginal densities $f(x)$ and $f(y)$.
(ii) Check if the variables X and Y are independent.
(iii) Compute the $P(X \geq 2; Y \geq 4)$.

Solution:
(i) The marginal density of X is
$f(x) = e^{-x} \int_0^\infty xe^{-xy} dy = e^{-x}$, for $x > 0$ (take $\lambda = x$ in exponential density).
Also, the marginal density of Y is

$$\begin{aligned} f(y) &= \int_0^\infty xe^{-(y+1)x} dx \\ &= \frac{1}{(y+1)} \int_0^\infty x(y+1)e^{-(y+1)x} dx \text{ (take } \lambda = y+1) \\ &= \frac{1}{(y+1})\frac{1}{(y+1)} \\ &= \frac{1}{(y+1)^2} \text{ for } y > 0, \end{aligned}$$

using the mean of the exponential density.

(ii) Since $f(x, y) \neq f(x)f(y)$, it follows X and Y are not independent.
(iii)

$$\begin{aligned}
P(X > 2 : Y > 4) &= \int\int_{(x>2,y>4)} f(x, y)dxdy \\
&= \int_4^\infty \int_2^\infty xe^{-x(1+y)}dxdy \\
&= \int_2^\infty e^{-x}(\int_4^\infty xe^{-xy}dy)dx \quad \text{(interchanging the order)} \\
&= \int_2^\infty e^{-x}e^{-4x}dx \\
&= \int_2^\infty e^{-5x}dx = \frac{1}{5}e^{-10},
\end{aligned}$$

using the fact $P(X > t) = e^{-\lambda t}$ for the exponential density.

Home Work:

Sect 5.1: 1, 12, 14, 15

Sect 5.2: 22, 26, 29, 30