Joint probability distributions: Discrete Variables

Probability mass function (pmf) of a single discrete random variable X specifies how much probability mass is placed on each possible X value.

The joint pmf of two discrete random variables X and Y describes how much probability mass is placed on each possible pair of values (x, y):

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

Two Discrete Random Variables

Like single pmf, joint pmf has to be positive, and add up to 1:

$$p(x, y) \geq 0 \quad \text{and} \quad \sum_x \sum_y p(x, y) = 1$$

Events: sets consisting of elements (x, y). Examples:

$A = \{(x, y): x + y = 5\}$
$B = \{(x, y): \max(x, y) \leq 3\}$
$C = \{(x, y): x = 5\}$
$D = \{(x, y): x < 5 \text{ and } y < 5\}$

Probability $P[(X, Y) \in A]$ = sum the joint pmf over pairs (x,y) in A:

$$P[(X, Y) \in A] = \sum_{(x, y)} \sum_{\in A} p(x, y)$$

Example 1

An insurance agency has customers with both home and auto policy. For each type of policy, a deductible amount must be specified.

For auto policy, choices are \$100 and \$250,
for home policy, choices are \$0, \$100, and \$200.

Suppose a customer is selected at random. Let:

X = his deductible on the auto policy
Y = his deductible on the home policy.

Example 1

cont'd

All possible values for (X, Y) are then:
(100, 0), (100, 100), (100, 200),
(250, 0), (250, 100), (250, 200)

Suppose the joint pmf is given by the insurance company in the accompanying **joint probability table:**

$p(x, y)$		y = 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

So from the table, $P(100, 100) = P(X = 100 \text{ and } Y = 100) = 0.10$.
$P(Y \geq 100) = p(100,100) + p(250,100) + p(100,200) + p(250,200) = .75$

Example 1

What is the probability that $X = 100$?
$p(X=100) = p(100, 0) + p(100, 100) + p(100, 200) = .50$

What is the probability that $X = 250$?
$p(X=250) = p(250, 0) + p(250, 100) + p(250, 200) = .50$

The "**marginal**" pmf of X is

$$p_X(x) = \begin{cases} .5 & x = 100, 250 \\ 0 & \text{otherwise} \end{cases}$$

(focuses only on X, and doesn't care what Y is)

Example 2

cont'd

Similarly, the marginal pmf of Y is obtained from column totals as:

$p(x, y)$		y = 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

$$p_Y(y) = \begin{cases} .25 & y = 0, 100 \\ .50 & y = 200 \\ 0 & \text{otherwise} \end{cases}$$

so $P(Y \geq 100) = p_Y(100) + p_Y(200) = .75$ as before.

Two Continuous Random Variables

Let X and Y be continuous. **Joint probability density function** $f(x, y)$ is a function satisfying $f(x, y) \geq 0$ and

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

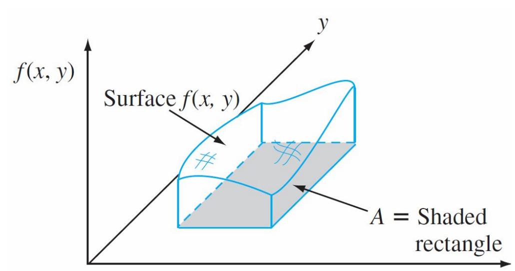
Then for any set A

$$P[(X, Y) \in A] = \iint_A f(x, y)\, dx\, dy$$

Two Continuous Random Variables

In particular, if A is the two-dimensional rectangle
$\{(x, y): a \leq x \leq b, c \leq y \leq d\}$, then

$$P[(X,Y) \in A] = P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\, dy\, dx$$



Example

A bank operates a drive-up and a walk-up window. Let
X = the proportion of time the drive-up facility is in use
Y = the proportion of time the walk-up window in use

Say the manager has given us the joint pdf based on his experience:

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

Example

cont'd

As a good probabilist, you first verify that this is a pdf:

1) $f(x, y) \ge 0$ (check)
2)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = \int_0^1\int_0^1 \frac{6}{5}(x + y^2)\,dx\,dy$$

$$= \int_0^1\int_0^1 \frac{6}{5}x\,dx\,dy + \int_0^1\int_0^1 \frac{6}{5}y^2\,dx\,dy$$

$$= \int_0^1 \frac{6}{5}x\,dx + \int_0^1 \frac{6}{5}y^2\,dy = \frac{6}{10} + \frac{6}{15} = 1$$

Example

cont'd

The probability that neither facility is busy more than one-quarter of the time is

$$P\left(0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}\right) = \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x + y^2)\,dx\,dy$$

$$= \frac{6}{5}\int_0^{1/4}\int_0^{1/4} x\,dx\,dy + \frac{6}{5}\int_0^{1/4}\int_0^{1/4} y^2\,dx\,dy$$

$$= \frac{6}{20}\cdot\frac{x^2}{2}\Big|_{x=0}^{x=1/4} + \frac{6}{20}\cdot\frac{y^3}{3}\Big|_{y=0}^{y=1/4}$$

$$= \frac{7}{640} = .0109$$

Marginal density

The **marginal probability density functions** of X and Y, denoted by $f_X(x)$ and $f_Y(y)$, respectively, are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy \qquad \text{for } -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx \qquad \text{for } -\infty < y < \infty$$

Independent Random Variables

In many situations, information about the observed value of one of the two variables *X* and *Y* gives information about the value of the other variable.

$p(x, y)$		y = 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

Here, $P(X = 250) = 0.5$ and $P(X = 100) = 0.5$

But, if we are told that the selected individual had $Y = 0$, then $X = 100$ is four times as likely as $X = 250$.

Independent Random Variables

Two random variables *X* and *Y* are said to be **independent** if for every pair *(x,y):*

$p(x, y) = p_X(x)\ p_Y(y)$ when *X* and *Y* are discrete

Or

$f(x, y) = f_X(x)\ f_Y(y)$ when *X* and *Y* are continuous

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

Example

In the insurance example,

$$p(100, 100) = .10$$

while

$$p_X(100)\ p_Y(100) = (.5)(.25) = .125$$

so *X* and *Y* are not independent.

Independence of *X* and *Y* requires that *every* entry in the joint probability table be the product of the corresponding row and column marginal probabilities.

Conditional Distributions

Suppose

X = number of major defects in a randomly selected car

Y = number of minor defects in that same car.

If we learn that the selected car has one major defect, what is the probability that the car will also have at least one minor defect?

That is, what is $P(Y >= 1 \mid X = 1)$?

Conditional Distributions

The conditional probability density function of Y given that $X = x$ is

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

If X and Y are discrete, replacing pdf's by pmf's in the above is the conditional probability mass function of Y when $X = x$.

The definition of $f_{Y|X}(y \mid x)$ parallels that of $P(B \mid A)$, the conditional probability that B will occur, given that A has occurred.

Example 1, cont

Reconsider the insurance example:

$p(x, y)$		y = 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

The conditional pmf of Y given that $X = 100$ is:

P(Y=0 | X=100) = P(Y=0 & X=100)/P(X=100) = 0.2/0.5 = 40%
P(Y=100 | X=100) = 0.1/0.5 = 20%
P(Y=200 | X=100) = 0.2/0.5 = 40%

Conditional expectation

cont'd

So what is the expected value of Y given that X=100?
("*conditional* expectation"):

0 * P(Y=0|X=100) +
100 * P(Y=100|X=100) +
200 * P(Y=200|X=100)

=

0*.4 + 100*.2 + 200*.4 = 20+80 = 100

More Than Two Random Variables

If $X_1, X_2, \ldots, X_n$ are all discrete random variables, the joint pmf of the variables is the function

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

If the variables are continuous, the joint pdf of $X_1, \ldots, X_n$ is the function $f(x_1, x_2, \ldots, x_n)$ such that for any n intervals $[a_1, b_1], \ldots, [a_n, b_n]$, we have:

$$P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n \ldots dx_1$$

Expected value of a function of two variables

Then the expected value of a function *h*(*X*, *Y*), denoted by *E*[*h*(*X*, *Y*)] is given by

$$E[h(X, Y)] = \begin{cases} \sum_x \sum_y h(x, y) \cdot p(x, y) & \text{when } X \text{ and } Y \text{ are discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y)\, dx\, dy & \text{when } X \text{ and } Y \text{ are continuous} \end{cases}$$

Example

You have purchased 5 tickets to a concert for you and 4 friends (one is Bob). Tickets are for seats 1–5 in one row. If the tickets are randomly distributed among you, what is the expected number of seats separating you from Bob?

Let *X* and *Y* denote the seat numbers for you and Bob, respectively. Possible (*X*, *Y*) pairs are

{(1, 2), (1, 3), . . . , (5, 4)}, and the joint pmf of (*X*, *Y*) is

$$p(x, y) = \begin{cases} \frac{1}{20} & x = 1, \ldots, 5;\ y = 1, \ldots, 5;\ x \neq y \\ 0 & \text{otherwise} \end{cases}$$

Example

cont'd

The number of seats between you and Bob is
h(*X*, *Y*) = |*X* – *Y*| – 1.

$h(x, y)$		x = 1	2	3	4	5
	1	—	0	1	2	3
	2	0	—	0	1	2
y	3	1	0	—	0	1
	4	2	1	0	—	0
	5	3	2	1	0	—

Example

cont'd

Thus

$$E[h(X, Y)] = \sum_{(x, y)}\sum h(x, y) \cdot p(x, y)$$

$$= \sum_{\substack{x=1 \\ x \neq y}}^{5} \sum_{y=1}^{5} (|x - y| - 1) \cdot \frac{1}{20}$$

Covariance

When two random variables X and Y are not independent, it is frequently of interest to assess how strongly they are related to one another.

The **covariance** between two rv's X and Y is

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

$$= \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y) & X, Y \text{ discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)\, dx\, dy & X, Y \text{ continuous} \end{cases}$$

Covariance

Since $X - \mu_X$ and $Y - \mu_Y$ are the deviations of the two variables from their respective mean values, the covariance is the expected product of deviations.

Note that $\text{Cov}(X, X) = E[(X - \mu_X)^2] = V(X)$.

If both variables tend to deviate in the same direction (both go above their means or below their means at the same time), then the covariance will be positive.

If they tend to deviate in the opposite direction from their means at the same time – ie, the signs of $(x - \mu_X)$ and $(y - \mu_Y)$ tend to be opposite -- the product (and covariance) will be negative

If X and Y are not strongly related, positive and negative products will tend to cancel one another, yielding a covariance near 0.

Covariance shortcut

The following shortcut formula for $\text{Cov}(X, Y)$ simplifies the computations.

Proposition

$\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y$

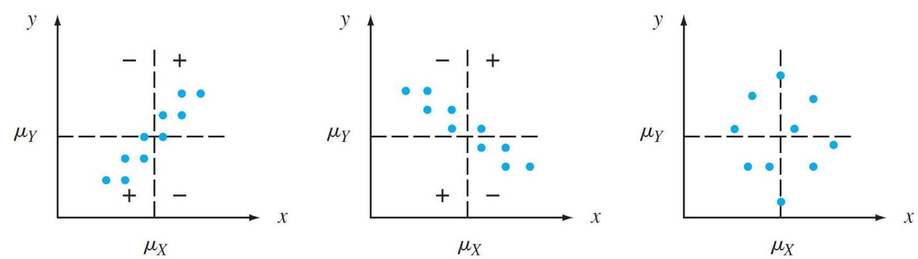
According to this formula, no intermediate subtractions are necessary; only at the end of the computation is $\mu_X \cdot \mu_Y$ subtracted from $E(XY)$.

This is analogous to the "shortcut" for the variance computation we saw earlier.

Covariance

The covariance depends on *both* the set of possible pairs and the probabilities of those pairs.

Below are examples of 3 types of "co-varying":



(a) positive covariance; (b) negative covariance; (c) covariance near zero

Example

The joint and marginal pmf's for
X = automobile policy deductible amount
Y = homeowner policy deductible amount

are:

$p(x, y)$		y: 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

x	100	250
$p_X(x)$	.5	.5

y	0	100	200
$p_Y(y)$	.25	.25	.5

from which $\mu_X = \Sigma x p_X(x) = 175$ and $\mu_Y = 125$.

Example, cont.

cont'd

Therefore,

$$\text{Cov}(X, Y) = \sum_{(x,y)}\sum (x - 175)(y - 125)p(x, y)$$

$$= (100 - 175)(0 - 125)(.20) + \ldots + (250 - 175)(200 - 125)(.30)$$

$$= 1875$$

This is a large positive covariance – but can we say anything about how large?

Correlation

Correlation

Definition

The **correlation coefficient** of X and Y, denoted by Cov(X, Y), $\rho_{X,Y}$, or just ρ, is defined by

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

It represents a "scaled" covariance – correlation ranges between -1 and 1.

Example

In the insurance example,

$E(X^2) = 36{,}250$, so

$\sigma_X^2 = 36{,}250 - (175)^2 = 5625$, so $\sigma_X = 75$

$E(Y^2) = 22{,}500$, so

$\sigma_Y{}^2 = 6875$, and $\sigma_Y = 82.92$.

This gives

$$\rho = \frac{1875}{(75)(82.92)} = .301$$

Correlation

Propositions

1. $\text{Cov}(aX + b, cY + d) = a\,c\,\text{Cov}(X, Y)$

2. $\text{Corr}(aX + b, cY + d) = sgn(ac)\,\text{Corr}(X, Y)$

3. For any two rv's X and Y, $-1 \le \text{Corr}(X, Y) \le 1$

4. $\rho = 1$ or -1 iff $Y = aX + b$ for some numbers a and b with $a \neq 0$.

Correlation

If X and Y are independent, then $\rho = 0$, but $\rho = 0$ does not imply independence.

The correlation coefficient ρ is actually not a completely general measure of the strength of a relationship.

This says that ρ is a measure of **linear** relationship between X and Y, and only when the two variables are perfectly related in a linear manner will ρ be as positive or negative as it can be.

A ρ less than 1 in absolute value indicates only that the relationship is not completely linear, but there may still be a very strong nonlinear relation.

Correlation

Also, $\rho = 0$ does not imply that X and Y are independent, but only that there is a complete absence of a linear relationship.

When $\rho = 0$, X and Y are said to be **uncorrelated.**

Two variables could be uncorrelated yet highly dependent because there is a strong nonlinear relationship, so be careful not to conclude too much from knowing that $\rho = 0$.

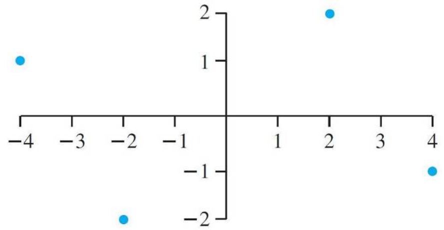
Let Y = X 50% of the time, and –X the other 50% of the time. Then Cor(X,Y) = 0, but there is a strong relationship.

Example

Let X and Y be discrete rv's with joint pmf

$$p(x, y) = \begin{cases} \frac{1}{4} & (x, y) = (-4, 1), (4, -1), (2, 2), (-2, -2) \\ 0 & \text{otherwise} \end{cases}$$

The points that receive positive probability mass are identified on the (x, y) coordinate system



Example

cont'd

The value of X is completely determined by the value of Y and vice versa, so the two variables are completely dependent. However, by symmetry $\mu_X = \mu_Y = 0$ and

$$E(XY) = (-4)\frac{1}{4} + (-4)\frac{1}{4} + (4)\frac{1}{4} + (4)\frac{1}{4} = 0$$

The covariance is then $\text{Cov}(X,Y) = E(XY) - \mu_X \mu_Y = 0$ and thus $\rho_{X,Y} = 0$. Although there is perfect dependence, there is also complete absence of any linear relationship!

Interpreting Correlation

A value of ρ near 1 does not necessarily imply that increasing the value of X *causes* Y to increase. It implies only that large X values are *associated* with large Y values.

For example, in the population of children, vocabulary size and number of cavities are quite positively correlated, but it is certainly not true that cavities cause vocabulary to grow.

Instead, the values of both these variables tend to increase as the value of age, a third variable, increases.

In summary, association (a high correlation) is not the same as causation.