Multivariate Distributions

In many applications we measure several r.v.'s per individual.

Examples:

- For a set of households, we can measure both income and the number of children.
- For a set of giraffes, we can measure their height and weight.
- We can measure the IQ and birth weight of children.
- The frequency of exercise and the rate of heart disease in adults.
- The level of air pollution and rate of respiratory illness in cities.
- The sex and survival indicator for Titanic's passengers.
- The age of a Facebook member and the number of Facebook friends.

The set of r.v.'s measured on each individual is called a random vector, usually denoted as $(Y_1, \ldots, Y_n)$.

Bivariate Probability Distributions

Consider a bivariate random vector $Y = (Y_1, Y_2)$.

Discrete r.v.'s: The joint pmf is $p(y_1, y_2) = \mathbb{P}(Y_1 = y_1, Y_2 = y_2)$.

All r.v.'s (discrete, continuous, and mixed): The joint cdf is

$$F(y_1, y_2) = \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2).$$

Continuous r.v.'s: The joint pdf is $f(y_1, y_2) = \frac{d}{dy_1}\frac{d}{dy_2}F(y_1, y_2)$.

Example

Roll two dice: $X =$ # on the first die, $Y =$ # on the second die.

The joint pmf of X and Y can be represented with a two-way table

$X \backslash Y$	1	2	3	4	5	6
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
2	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
3	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
4	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
5	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
6	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$

$p(i,j) = \frac{1}{36}$ for all i and j.

Example

Roll two dice: $X = \#$ on the first die, $T =$ total on both dice.

The joint pmf of X and T:

$X \backslash T$	2	3	4	5	6	7	8	9	10	11	12
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36

Example

Researches at three universities (in New York, California, and Florida) are applying for two separate grants. Suppose all proposals are equally good, and so which university get the contracts can be seen as a random selection.

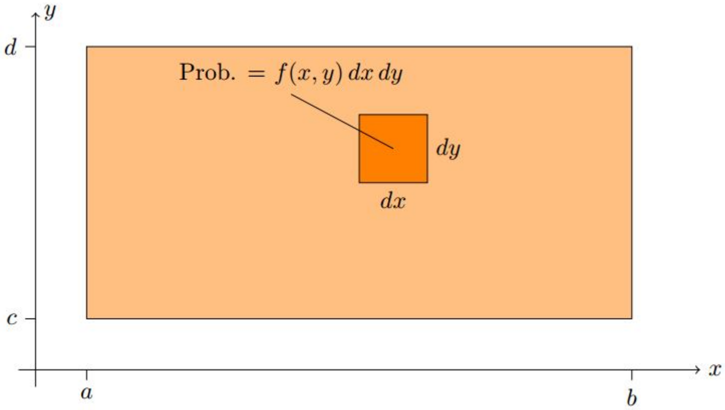
Let Y_1 = number of grants awarded to New York and Y_2= number of grants awarded to California.

Sample space ?

The joint pmf of Y_1 and Y_2?

Example

- X takes values in $[a, b]$, and Y takes values in $[c, d]$.
- (X, Y) take values in $[a, b] \times [c, d]$.
- The joint probability density function (pdf) of X and Y: $f(x, y)$.



$f(x, y)dxdy$ is the probability of being in the small rectangle.

Properties of joint pmf and pdf

Discrete case: probability mass function (pmf)

1. $0 \le p(x_i, y_j) \le 1$.
2. Total probability is 1:

$$\sum_{x_i, y_j} p(x_i, y_j) = 1.$$

Continuous case: probability density function (pdf)

1. $0 \le f(x_i, y_j)$.
2. Total probability is 1:

$$\int f(x, y) dx dy = 1.$$

Note that the density can be greater than 1.

Cumulative distribution function

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \int_c^y \int_a^x f(x, y) dxdy$$

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

Properties of cdf:

1. $F(x, y)$ is non-decreasing. That is, as x or y increases, $F(x, y)$ increases or remains constant.
2. $F(x, y) = 0$ at the lower left of its range. If the lower left is $(-\infty, -\infty)$, then this means $\lim F(x, y) = 0$ as $(x, y) \to (-\infty, -\infty)$.
3. $F(x, y) = 1$ at the upper right of its range.

Calculating Probabilities

Researches at three universities (in New York, California, and Florida) are applying for two separate grants. Suppose all proposals are equally good, and so which university get the contracts can be seen as a random selection.

Let Y_1 = number of grants awarded to New York and Y_2= number of grants awarded to California.

Find the probability that New York obtains at least as many grants as California.

Quiz

Roll two dice: X = # on the first die, Y = # on the second die.

What is the probability of the event that $Y - X \geq 2$?

$X \backslash Y$	1	2	3	4	5	6
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
2	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
3	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
4	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
5	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
6	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$

(A) 1/4
(B) 5/18
(C) 11/36
(D) 1/3
(E) Other

Quiz Explanation

Roll two dice: $X = \#$ on the first die, $Y = \#$ on the second die.

What is the probability of the event that $Y - X \geq 2$?

We can visualize the event by shading the cells.

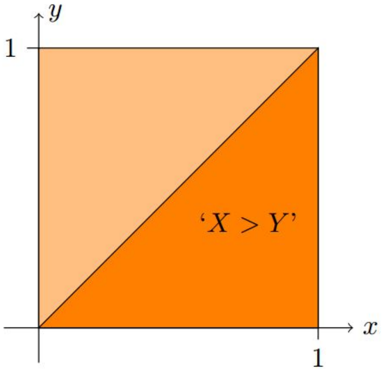
$X \backslash Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

$\mathbb{P}(Y - X \geq 2)$ = sum of the numbers in the shaded cells = $\frac{10}{36}$

Example

Suppose (X, Y) take values in the square $[0, 1] \times [0, 1]$.
Assume the uniform density: $f(x, y) = 1$.

What is the probability that $X > Y$?

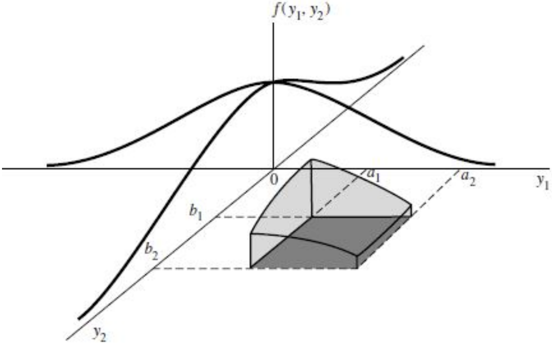


The event is the half of the square. Since the density is uniform, its probability is half of the total probability.
$P(X > Y) = 1/2$.

Probabilities as volumes

A (bivariate) joint pdf may be plotted as a 3-d function on the (y_1, y_2) plane.

Volumes under this surface correspond to probabilities (like areas under the pdf curve in the univariate case).



Calculating Probabilities

Finding a probability amounts to integrating the joint pdf over a particular region of that support.

Example: Suppose the random vector (Y_1, Y_2) has joint pdf

$$f(y_1, y_2) = \begin{cases} y_1 + y_2, & \text{for } 0 < y_1 < 1, 0 < y_2 < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $\mathbb{P}(Y_1 + Y_2 < 1)$.

Calculating Probabilities

Example: Suppose the time (in hours) to complete task 1 and task 2 for a random employee has the joint pdf:

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)}, & \text{for } 0 < y_1, 0 < y_2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that a random employee takes less than 2 hours on task 1 and between 1 and 3 hours on task 2.

What is the probability the employee takes longer on task 2 than on task 1?

Marginal Distribution

- For a bivariate r.v. (X, Y) the marginal distribution of X is the distribution of X seen as a single r.v.

Example: Roll two dice: X = # on first die, T = total on both dice. What are the marginal distributions of X and T?

Marginal Distribution

- For a bivariate r.v. (X, Y) the marginal distribution of X is the distribution of X seen as a single r.v.

Example: Roll two dice: X = # on first die, T = total on both dice. What are the marginal distributions of X and T?

Marginal Distribution

- For a bivariate r.v. (X, Y) the marginal distribution of X is the distribution of X seen as a single r.v.

Example: Roll two dice: X = # on first die, T = total on both dice. What are the marginal distributions of X and T?

The marginal pmf of X is found by summing the rows. The marginal pmf of T is found by summing the columns.

$X \backslash T$	2	3	4	5	6	7	8	9	10	11	12	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0	1/6
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	1/6
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	1/6
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	1/6
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	1/6
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(t_j)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1

Example

Researches at three universities (in New York, California, and Florida) are applying for two separate grants. Suppose all proposals are equally good, and so which university get the contracts can be seen as a random selection.

Let Y_1 = number of grants awarded to New York and Y_2= number of grants awarded to California.

Find the marginal distribution of Y_1.

Marginal Distribution: Continuous r.v.'s

For a jointly continuous r.v. (X,Y), the marginal pdf $f_X(x)$ is found by integrating out the y. Likewise for $f_Y(y)$.

$$f_X(x)) = \int f(x,y)dy,$$

and

$$f_Y(y)) = \int f(x,y)dx.$$

Example

Suppose we model two proportions Y_1 and Y_2 with the joint pdf

$$f(y_1, y_2) = \begin{cases} 0.4(y_1 + 4y_2), & \text{for } 0 < y_1 < 1, 0 < y_2 < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the marginal pdf of Y_1 and find $\mathbb{P}(Y_1 > 0.3)$.

Example

Suppose for two proportions Y_1 and Y_2, we know that $Y_1 < Y_2$. We assume that the joint pdf is

$$f(y_1, y_2) = \begin{cases} 6y_1, & \text{for } 0 < y_1 < y_2 < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the marginal pdf of Y_1 and Y_2.

Quiz

Suppose X and Y are random variables and (X, Y) takes values in $[0, 1] \times [0, 1]$.
Suppose that the pdf is

$$k(x^2 + y^2).$$

Find k.

(A) 2/3
(B) 3/2
(C) 2
(D) 3
(E) Other

Quiz

Suppose X and Y are random variables and (X, Y) takes values in $[0, 1] \times [0, 1]$.
Suppose that the pdf is
$$k(x^2 + y^2).$$

Find the marginal pdf $f_X(x)$. Use this to find $\mathbb{P}(X < .5)$.

(A) 5/16
(B) 7/32
(C) 1/3
(D) 1/2
(E) Other

Quiz

$X \backslash Y$	1	2	3	4	5	6
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
2	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
3	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
4	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
5	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
6	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$

Compute $F(3.5, 4)$.

(A) 1/4
(B) 1/3
(C) 4/9
(D) 1/2

Conditional Distributions

The conditional probability distribution of Y given X is the probability distribution we should use to describe Y after we have seen X.

In the discrete case,

$$p(y|x) = \mathbb{P}(Y = y|X = x) = \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)} = \frac{p(x, y)}{p_X(x)} = \frac{\text{joint pmf}}{\text{marginal pmf}}.$$

Similarly,

$$p(x|y) = \frac{p(x, y)}{p_Y(y).}$$

Example

Researches at three universities (in New York, California, and Florida) are applying for two separate grants. Suppose all proposals are equally good, and so which university get the contracts can be seen as a random selection.

Let Y_1 = number of grants awarded to New York and Y_2= number of grants awarded to California.

Find the conditional distribution of Y_1 given $Y_2 = 0$ and find $\mathbb{P}(Y_1 \geq 1 | Y_2 = 0)$.

Conditional Distribution: Continuous r.v.'s

For continuous r.v.'s, the conditional pdf of Y given X is defined as

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\text{joint pdf}}{\text{marginal pdf}}.$$

Similarly, the conditional pdf of X given Y is

$$f(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

Note that the conditional pdf $f(x|y)$ is a proper pdf in x, and y is a parameter. For example,

$$\int f(x|y)dx = 1$$

for every y in the range of r.v. Y.

Conditional Expectations

We can also define conditional expectations:

$$\mathbb{E}[g(X)|Y=y] = \int g(x)f(x|y)dx.$$

For example,

$$\mathbb{E}[X|Y=y] = \int xf(x|y)dx.$$

A conditional expectation is a function of parameter y.

Example

Suppose for two proportions X and Y, we know that $X < Y$. We assume that the joint pdf is

$$f(x, y) = \begin{cases} 6x, & \text{for } 0 \leq x < y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $f(y|x)$, find $\mathbb{P}(Y < 0.8|X = 0.4)$, and find $\mathbb{E}(Y|X = x)$

Quiz

Suppose we model two proportions X and Y with the joint pdf

$$f(x,y) = \begin{cases} \frac{2}{5}(x+4y), & \text{for } 0 \le x \le 1, 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $f(y|x)$ and compute $\mathbb{P}(Y < 0.5 | X = 0.5)$.

(A) 1/5
(B) 3/10
(C) 7/20
(D) 1/9
(E) Other

Example

Suppose that X and Y are independent Poisson distributed random variables with means λ_1 and λ_2, respectively. Let $W = X + Y$.

Find the conditional distribution of X, given that $W = w$.

Independent r.v.

- Random variables X and Y are independent if
$$F(x,y) = F_X(x)F_Y(y).$$
Otherwise, they are called dependent.

- Discrete random variables are independent if
$$p(x_i, y_j) = p_X(x_i)p_Y(y_j).$$

- Continuous random variables are independent if
$$f(x,y) = f_X(x)f_Y(y).$$

Independence and conditional probabilities

If X and Y are independent, then

$$f_Y(y|x) = f_Y(y),$$

and

$$f_X(x|y) = f_X(x).$$

Quiz

Roll two dice: X = # on the first die, Y = # on the second die.

$X\backslash Y$	1	2	3	4	5	6	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(y_j)$	1/6	1/6	1/6	1/6	1/6	1/6	1

Are these random variables independent?

(A) Yes
(B) No

Quiz

Roll two dice: X = # on the first die, T = total on both dice.

$X\backslash T$	2	3	4	5	6	7	8	9	10	11	12	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0	1/6
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	1/6
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	1/6
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	1/6
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	1/6
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(y_j)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1

Are these random variables independent?

(A) Yes
(B) No

Quiz

Which of the following pdf's corresponds to independent random variables X and Y?

(i) $f(x, y) = 4x^2y^3$.
(ii) $f(x, y) = \frac{1}{2}(x^3y + xy^3)$.
(iii) $f(x, y) = 6e^{-3x-2y}$.

(Assume that the range of X and Y is a rectangle chosen in such a way that these are valid densities.)

Put a 1 for independent and a 0 for non-independent.

(A) 111
(B) 110
(C) 100
(D) 010
(E) 101

Examples

If the joint pdf for X and Y can be factored into two non-negative functions,

$$f(x, y) = g(x)h(y),$$

and the support of the joint pdf is a rectangle, then X and Y are independent.
The functions $g(x)$ and $h(y)$ need not be valid pdf's.

If the support of the joint pdf of X and Y is not a rectangle, then these variables are not independent.

Quiz

Let the joint density be

$$f(x, y) = \begin{cases} 2y, & \text{for } 0 < x < 1, 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Are these r.v.'s independent?

(A) Yes
(B) No

Quiz

Let the joint density of X and Y be

$$f(x, y) = \begin{cases} 6x, & \text{for } 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Are these variable independent?

(A) Yes
(B) No

Independence of n random variables

The continuous r.v.'s Y_1, Y_2, ..., Y_n are independent if and only if

$$f(y_1, y_2, \dots, y_n) = f_1(y_1) \dots f_n(y_n),$$

where $f_i(y_i)$ is the pdf of the i-th random variable.

If, in addition, they have the same distribution then they are said to be independent and identically distibuted (i.i.d.).

Example: If we take a take a random sample of n measurements from a large population, these measurements are i.i.d random variables.

Example

Lifelengths (in hours) of a population of batteries follow an exponential distribution with parameter $\lambda = 30$ distribution. We take a random sample of 5 batteries and observe their lifelengths.

Find the joint pdf of these 5 measurements.

Expected Value of a Function of a Random Vector

Let (X, Y) be jointly dicrete r.v.'s having joint pmf $p(x, y)$. Then for any function $g(X, Y)$:

$$\mathbb{E}g(X, Y) = \sum_{x_i, y_j} g(x_i, y_j)p(x_i, y_j).$$

Similarly, for jointly continuous r.v.'s X and Y with joint pdf $f(x, y)$:

$$\mathbb{E}g(X, Y) = \int\int g(x, y)f(x, y)dxdy.$$

Example

Consider a product that contains impurities. Some impurities are toxic and some nontoxic. Let X be the proportion of a sample of the product that is impure, and let Y represent the proportion of the impurities that are toxic. The joint pdf of X and Y is:

$$f(x,y) = \begin{cases} 2(1-x), & \text{for } 0 < x < 1, 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected proportion of the sampled product containing toxic

impurities.

We want to compute $\mathbb{E}(XY)$.

Some simple facts

Given the joint pdf, we can find the expected values of functions of X alone or Y alone as well. For example,

$$\mathbb{E}Y = \int\int yf(x,y)dxdy.$$

$$\operatorname{Var}Y = \int\int (y - \mathbb{E}Y)^2 f(x,y)dxdy.$$

If a and b are constants, then

$$\mathbb{E}ag(X,Y) + b = a\mathbb{E}g(X,Y) + b.$$

$$\mathbb{E}[g_1(X,Y) + \ldots + g_k(X,Y)] = \mathbb{E}g_1(X,Y) + \ldots + \mathbb{E}g_k(X,Y).$$

Example

Let the joint density be

$$f(x,y) = \begin{cases} 6x, & \text{for } 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $\mathbb{E}(Y - X)$.

One approach is to integrate using the joint pdf.

Alternatively, note that $\mathbb{E}(Y - X) = \mathbb{E}Y - \mathbb{E}X$.

Covariance

Measures the degree to which two random variables vary together, e.g. height and weight of people.

If X and Y are random variables with means μ_X and μ_Y, then their covariance is

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

If $\mathrm{Cov}(X, Y) > 0$, then X and Y are positively related (as X increases, Y tends to increase).

If $\mathrm{Cov}(X, Y) < 0$, they are negatively related.

Covariance is similar to dot product.

Properties of Covariance

1. $\mathrm{Cov}(aX + b, cY + d) = ac\mathrm{Cov}(X, Y)$ for constant a, b, c, d.
2. $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$.
3. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.
4. $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mu_X \mu_Y$.
5. If X and Y are independent then $\mathrm{Cov}(X, Y) = 0$.

Quiz

Suppose we have the following probability mass table:

$Y \backslash X$	-1	0	1	$p(y_j)$
0	0	$1/2$	0	$1/2$
1	$1/4$	0	$1/4$	$1/2$
$p(x_i)$	$1/4$	$1/2$	$1/4$	1

What is the covariance of X and Y?

Quiz

Suppose we have the following probability mass table:

$Y \backslash X$	-1	0	1	$p(y_j)$
0	0	$1/2$	0	$1/2$
1	$1/4$	0	$1/4$	$1/2$
$p(x_i)$	$1/4$	$1/2$	$1/4$	1

Are X and Y independent?
(A) True
(B) False

Quiz

Flip a fair coin 3 times.
Let X = number of heads in the first 2 flips

Let Y = number of heads in the last 2 flips.

Compute $\mathrm{Cov}(X, Y)$.

(A) 0
(B) 1/4
(C) 1/3
(D) 1/2
(E) Other

Correlation coefficient

Like covariance, but removes scale.

$$\mathrm{Cor}(X,Y) = \rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

Properties:

1. ρ is the covariance of the standardized versions of X and Y.
2. $-1 \le \rho \le 1$.
   $\rho = 1$ if and only if $Y = aX + b$ with $a > 0$.
   $\rho = -1$ if and only if $Y = aX + b$ with $a < 0$.
3. $\mathrm{Cor}(aX + b, cY + d)$ equals $\mathrm{Cor}(X,Y)$ if $ac > 0$,
   and it equals $-\mathrm{Cor}(X,Y)$ if $ac < 0$.

Example

Let

$$f(x,y) = \begin{cases} 6x, & \text{for } 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $\mathrm{Cov}(X, Y)$ and find ρ.

We have seen that marginally, $X \sim \mathrm{beta}(2,2)$ and $Y \sim \mathrm{beta}(3,1)$.

A MATLAB solution

```
f = @(x,y) 6*x; %density 6x on the region 0 < x < y < 1.

format rat %represent results as ratios
E_XY = integral2(@(x,y) x.*y.*f(x,y), 0, 1, @(x) x, 1)
E_X = integral2(@(x,y) x.*f(x,y), 0, 1, @(x) x, 1)
E_Y = integral2(@(x,y) y.*f(x,y), 0, 1, @(x) x, 1)
Cov_XY = E_XY - E_X * E_Y

E_XY = 2/5    E_X = 1/2    E_Y = 3/4    Cov_XY = 1/40

V_X = integral2(@(x,y) x.^2.*f(x,y), 0, 1, @(x) x, 1) ..
                                                    - (E_X)^2
V_Y = integral2(@(x,y) y.^2.*f(x,y), 0, 1, @(x) x, 1) ..
                                                    - (E_Y)^2
format shortG
Cor_XY = Cov_XY / sqrt(V_X * V_Y)

V_X = 1/20    V_Y = 3/80    Cor_XY = 0.57735
```

Empirical cdf, expectation, variance and covariance

Sample S with n observations: $(x_1, y_1), \ldots, (x_n, y_n)$.
Empirical cdf:

$$F_{X,n}(x) = \frac{\# \text{ of } x_i < x}{n}.$$

Empirical expectation (mean of the sample) and variance:

$$\widehat{\mu}_X = \frac{1}{n}(x_1 + \ldots + x_n).$$

$$\widehat{\sigma}_X^2 = \frac{1}{n}(x_1^2 + \ldots + x_n^2) - \widehat{\mu}_X^2.$$

Empirical covariance and correlation:

$$\widehat{\mathrm{Cov}}(X, Y) = \frac{1}{n}(x_1 y_1 + \ldots + x_n y_n) - \widehat{\mu}_X \widehat{\mu}_Y.$$

$$\widehat{\mathrm{Cor}}(X, Y) = \frac{\widehat{\mathrm{Cov}}(X, Y)}{\widehat{\sigma}_X \widehat{\sigma}_X}.$$

Examples of real-world correlations

- Over time, amount of ice cream consumption is correlated with number of pool drownings.
- Hormone replacement therapy (HRT) is correlated with a lower rate of coronary heart disease (CHD).
- Correlation is not causation.

Linear functions of r.v.'s

Let $U = aX + bY$ and $W = cX + dY$.

Then, for expectations:

$$\mathbb{E}U = a\mathbb{E}X + b\mathbb{E}Y \text{ and } \mathbb{E}W = c\mathbb{E}X + d\mathbb{E}Y.$$

For covariance:

$$\mathrm{Cov}(U, W) = \mathrm{Cov}(aX + bY, cX + dY)$$

$$= acVar(X) + (ad + bc)Cov(X, Y) + bdVar(X).$$

And for variances:

$$\mathrm{Var}(U) = \mathrm{Var}(aX + bY) = a^2\mathrm{Var}X + 2ab\mathrm{Cov}(X, Y) + b^2\mathrm{Var}Y.$$

Special Cases

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X),$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + 2\mathrm{Cov}(X, Y) + \mathrm{Var}(Y),$$

If X and Y are independent, then

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$$

Example

Let

$$f(x, y) = \begin{cases} 6x, & \text{for } 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $\mathrm{Var}(Y - X)$, $\mathrm{Cov}(X, Y - X)$, and $\mathrm{Cov}(X + Y, Y - X)$.

Quiz

Let X_1, X_2 and X_3 be independent random variables with mean $\mu = -1$ and variance $\sigma^2 = 4$.

If $Y = 3X_1 + X_2 + 1$ and $Z = X_1 - 3X_2 + X_3 + 2$,

then $\mathrm{Cov}(Y, Z) = ?$

Conditional Expectation

The conditional expectation of $g(X)$ given that $Y = y$, is

$$\mathbb{E}(g(x)|Y = y) = \int_{-\infty}^{\infty} g(X)f(x|y)dx,$$

if X and Y are jointly continuous and

$$= \sum_{x_i} g(x_i)p(x_i|y),$$

if X and Y are jointly discrete.

Example

Let X and Y have the joint density

$$f(x,y) = \begin{cases} \frac{1}{y} e^{-x/y} e^{-y}, & \text{for } x > 0, y > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $\mathbb{E}(X|Y = y)$.

Quiz

Let X and Y have the joint probability density function given by

$$f(x,y) = \begin{cases} 3x, & 0 \le y \le x \le 1; \\ 0, & \textit{elsewhere}; \end{cases}$$

Find $\mathbb{E}(X|Y = 1/2)$.

(A) 1/2
(B) 7/12
(C) 2/3
(D) 7/9
(E) Other

A different view on conditional expectation

$\mathbb{E}(Y|X=x)$ is a function of x. So we can think about $\mathbb{E}(Y|X)$ as a random variable which is a function of the random variable X.

Example: Roll a fair die. Y = # on the die. $X = 1$ if the # is odd, and $X = 0$ if it is even.

$X = 1$ if $\# \in \{1, 3, 5\}$ and $X = 0$ if $\# \in \{2, 4, 6\}$

Then $\mathbb{E}(Y|X) = 3$ with probability $1/2$ (when $X = 1$),
and $\mathbb{E}(Y|X) = 4$ with probability 1/2 (when $X = 0$).

What is the expectation of this new random variable?

What is the expectation of Y ?

Law of Iterated Expectations

$$\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}(Y).$$

Proof in the discrete case:

$$\mathbb{E}[\mathbb{E}(Y|X)] = \sum_{x_i} \left\{ \sum_{y_j} y_j p(y_j|x_i) \right\} p_X(x_i)$$

$$= \sum_{y_j} y_j \left\{ \sum_{x_i} p(y_j|x_i) p_X(x_i) \right\}$$

$$= \sum_{y_j} y_j p_Y(y_j) = \mathbb{E}(Y)$$

The continuous case is similar.

Example

A miner is trapped in a mine containing 3 doors.
The first door leads to a tunnel that will take him to safety after 3 hours of travel.
The second door leads to a tunnel that will return him to the mine after 5 hours of travel.
The third door leads to a tunnel that will return him to the mine after 7 hours of travel.
If we assume that the miner is at all times equally likely to choose any one of the doors, find the expected time for him to get to safety.

Properties of conditional expectation

1. $$\mathbb{E}(aX + bY|Z) = a\mathbb{E}(X|Z) + b\mathbb{E}(Y|Z).$$

2. In general, $\mathbb{E}(f(X)g(Y)|Z) \neq \mathbb{E}(f(X)|Z)\mathbb{E}(g(Y)|Z)$.
   If it does hold true for all functions f and g, then X and Y are called conditionally independent.
   Independent variables are always conditionally independent, however the converse is not true.

3. $$\mathbb{E}(f(X)Y|X) = f(X)\mathbb{E}(Y|X).$$

Independence and Conditional Independence

Three random variables: X, Y, Z with joint pmf in the first row, and the marginal distribution of X and Y in the second raw.

$Z = 0$

X\Y	0	1	10
0	1/8	1/8	0
1	1/8	1/8	0
10	0	0	0

$Z = 1$

X\Y	0	1	10
0	1/8	0	1/8
1	0	0	0
10	1/8	0	1/8

X\Y	0	1	10
0	1/4	1/8	1/8
1	1/8	1/8	0
10	1/8	0	1/8

Example

Let X and Y denote the values of two stocks at the end of a five-year period. X is uniformly distributed on the interval $(0, 12)$. Given $X = x$, Y is uniformly distributed on the interval $(0, x)$.
Calculate $\mathrm{Cov}(X, Y)$ according to this model.

Example

Let Y be the number of defective parts from an assembly line out of 20 produced produced per day.
Suppose the probability p of a part being defective is random (It changes from day to day.) The distribution of p is $\mathrm{Beta}(1, 9)$.

Find $\mathbb{E}(Y)$.

Quiz

There is an urn which contains 3 marbles (2 white and 1 black). The person who gets to pick the black marble gets to win \$111 whereas the individual who picks the white marble walks away with nothing. The marbles that are taken out are not replaced. Assume that there are only 2 individuals.

What is the expectation of the second guy's earnings?

Law of Iterated Variance

Theorem If X and Y are any r.v.'s, then

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}[\mathbb{E}(Y|X)].$$

Proof:

$$\begin{aligned}
\mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}[\mathbb{E}(Y|X)] &= \mathbb{E}\{\mathbb{E}(Y^2|X) - (\mathbb{E}(Y|X))^2\} \\
&+ \mathbb{E}[(\mathbb{E}(Y|X))^2] - [\mathbb{E}(\mathbb{E}(Y|X))]^2 \\
&= \mathbb{E}\{\mathbb{E}(Y^2|X)\} - \left\{\mathbb{E}(\mathbb{E}(Y|X))\right\}^2 \\
&= \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 \\
&= \mathrm{Var}(Y).
\end{aligned}$$

Example

Let Y be the number of defective parts from an assembly line out of 20 produced produced per day.
Suppose the probability p of a part being defective is random (It changes from day to day.) The distribution of p is $\mathrm{Beta}(1, 9)$.

Find $\mathrm{Var}(Y)$.

Gambler's ruin example

A gambler has no money. He goes to a casino and plays a game. A fair coin is thrown. If it shows head, the gambler gets a dollar, if tail, he pays a dollar. If he cannot pay, he is ruined.

What is the probability that he will eventually be ruined?
Let $p_n = \mathbb{P}($ "eventually ruined if starts with n dollars"$)$. So, p_0 = ?

What is the expected time to the ruin?
Let $t_n = \mathbb{E}($ "time to the ruin if starts with n dollars"$)$. So, t_0 = ?

Quiz

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} = ?$$

(A) $\begin{bmatrix} 2 & 2 \\ -2 & 2 \end{bmatrix}$

(B) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(C) $\begin{bmatrix} -3 & 2 \\ -2 & 1 \end{bmatrix}$

(D) $\begin{bmatrix} 3 & 2 \\ -2 & -1 \end{bmatrix}$

(E) $\begin{bmatrix} 1 & 2 \\ -2 & -3 \end{bmatrix}$

Quiz

$$\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = ?$$

(A) 2
(B) -2
(C) 5
(D) 24
(E) 10

The multivariate normal distribution

Let $\mu = (\mu_X, \mu_Y)^t$ and

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix},$$

where $\sigma_X^2 > 0$, $\sigma_Y^2 > 0$, and $|\sigma_{XY}| < \sigma_X \sigma_Y$.

Random vector $Z = (X, Y)$ has the bi-variate normal (or Gaussian) distribution with parameters μ and Σ, if its density function is

$$f(x, y) = \frac{1}{(2\pi)\sqrt{\det(\Sigma)}} \exp - \left(-\frac{1}{2}(z - \mu)^t \Sigma^{-1} (z - \mu) \right),$$

where $z = (x, y)^t$ and Σ^{-1} is the inverse of the matrix Σ.

Notation: $Z \sim N(\mu, \Sigma)$.

Basic properties of multivariate Gaussian

$f(x, y)$ is a valid density function.

If $Z \sim N(\mu, \Sigma)$, then

- $\mathbb{E}X = \mu_X$.
- $\mathbb{E}Y = \mu_Y$.
- $\mathrm{Var}(X) = \sigma_X^2$.
- $\mathrm{Var}(Y) = \sigma_Y^2$.
- $\mathrm{Cov}(X, Y) = \sigma_{XY}$.

The bi-variate normal distribution

The density of the bi-variate normal distribution with parameters σ_X, σ_Y, $\rho = \sigma_{XY}/(\sigma_X\sigma_Y)$ is

$$f(x,y) = \frac{1}{(2\pi)\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp - \left(\frac{1}{2}Q(x,y)\right),$$

where

$$Q = \frac{1}{1-\rho^2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right].$$

Facts about the bi-variate normal:

- Marginal distributions: $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$.
- The conditional distribution of $Y|X = x$ is

$$N\left(\mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(x - \mu_X), \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}\right).$$

Example

Suppose (X, Y) is a bi-variate normal vector with $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, $\rho = 1/2$. Find $\mathbb{P}(Y > 0 | X = 1)$.

The multinomial distribution

This is a generalization of the binomial distribution.

Rather than 2 possible outcomes per trial, we assume that there are k possible outcomes per trial.

Multinomial experiment

- There are n identical trials.
- The outcome of each trial falls into one of k categories.
- The probability that the outcome falls into category i is p_i (for $i = 1, 2, \ldots, k$), where $p_1 + p_2 + \ldots + p_k = 1$, and this set of probabilities is the same across trials.
- The trials are independent.

The multinomial random vector is $X = (X_1, X_2, \ldots, X_k)$, where X_i is the number of trials resulting in category i.

Note that $X_1 + X_2 + \ldots + X_k = n$.

The multinomial distribution: definition

The random vector $(X_1, X_2, \ldots, X_k)$ is distributed as a multinomial r.v. with parameters n and $p = (p_1, p_2, \ldots, p_k)$ if its joint probability function is

$$p(x_1, x_2, \ldots, x_k) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \ldots p_n^{x_n},$$

where each $x_i \in \{0, 1, \ldots, n\}$, $\sum_{i=1}^{k} x_i = n$, and $\sum_{i=1}^{k} p_i = 1$.

Example

Three card players play a series of matches. The probability that player A will win any game is 20%, the probability that player B will win is 30%, and the probability player C will win is 50%.

If they play 6 games, what is the probability that player A will win 1 game, player B will win 2 games, and player C will win 3?

Means, variances and covariances

If $(X_1, \ldots, X_n) \sim \mathrm{Multinom}(n, p_1, \ldots, p_k)$, then

1. $\mathbb{E}(X_i) = np_i$.
2. $\mathrm{Var}(X_i) = np_i(1 - p_i)$.
3. $\mathrm{Cov}(X_i, X_j) = -np_ip_j$, if $i \neq j$.

Results (1) and (2) follow because the marginal distribution of X_i is binomial with parameters n, p_i.

(3) Let $U_s = 1$ if the trial s results in outcome i, and $U_s = 0$ otherwise.

Similarly, let $W_t = 1$ if the trial t results in outcome j, and $W_t = 0$ otherwise.

$$X_i = U_1 + U_2 + \ldots + U_n.$$

$$X_j = W_1 + W_2 + \ldots + W_n.$$

...

Conditional Distribution: Example

Let $(X_1, X_2, X_3) \sim \mathrm{Multinom}(n, p_1, p_2, p_3)$.
Then

$$p(x_1, x_2 | x_3) \sim \mathrm{Multinom}(n - x_3, \frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}).$$

Dirichlet's distribution

A generalization of Beta distribution:

$$f(x_1, \ldots, x_n) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_n)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_n)} x_1^{\alpha_1 - 1} \ldots x_n^{\alpha_n - 1} \delta(x_1 + \ldots + x_n - 1),$$

where $x_i > 0$.