

Chapter 7: Joint Distributions

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington
Department of Mathematics

Summer 2026

By the end of this lecture, students should be able to:

- ▶ define joint, marginal, and conditional distributions;
- ▶ understand independence of random variables;
- ▶ apply the two-dimensional LOTUS formula;
- ▶ define covariance and correlation;
- ▶ distinguish independence from zero correlation;
- ▶ define pairwise and mutual independence;
- ▶ define the Multinomial distribution;
- ▶ understand the Multivariate Normal distribution.

Joint Distributions

Why Joint Distributions?

Previously, we studied one random variable at a time.

But many probabilistic questions involve several random variables simultaneously.

Examples:

- ▶ height and weight;
- ▶ temperature and rainfall;
- ▶ several coin tosses;
- ▶ sums of random variables;
- ▶ sample means.

Main idea

Joint distributions describe how multiple random variables behave together.

Definition

For discrete random variables X, Y , the joint PMF is

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y).$$

Properties:

$$p_{X,Y}(x,y) \geq 0, \quad \text{and} \quad \sum_x \sum_y p_{X,Y}(x,y) = 1.$$

Similarly, for n discrete random variables $X_1, \dots, X_n$, the joint PMF is

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

Note $\{X \leq x, Y \leq y\} = \{X \leq x\} \cap \{Y \leq y\}$

Question How to find $p_X(x)$ (the PMF of X) and $p_Y(y)$ (the PMF of Y) using $p_{X,Y}(x,y)$ (the joint distribution of X and Y)?

Answer: The marginal PMFs are obtained by summing out the other variable.

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$

and

$$p_Y(y) = \sum_x p_{X,Y}(x,y).$$

Remark

Marginal's distribution can be obtained via (summing up) the joint distribution but not necessarily vice versa. We will see that independence enables us to also find the joint distribution through the marginal distributions.

Example: Two Fair Dice and Marginals

Let X be the outcome of the first die and Y the outcome of the second die.

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y) = \frac{1}{36}, \quad x, y \in \{1, 2, 3, 4, 5, 6\}.$$

$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$	$y = 6$	$p_X(x) = \mathbb{P}(X = x)$
$x = 1$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$x = 2$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$x = 3$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$x = 4$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$x = 5$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$x = 6$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$p_Y(y) = \mathbb{P}(Y = y)$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	

For example,

$$p_X(3) = \sum_{y=1}^6 p_{X,Y}(3,y) = \frac{1}{36} + \cdots + \frac{1}{36} = \frac{6}{36} = \frac{1}{6}, \quad \text{and} \quad p_Y(4) = \sum_{x=1}^6 p_{X,Y}(x,4) = \frac{6}{36} = \frac{1}{6}.$$

The marginal PMFs appear on the "margins" of the joint PMF table.

Proof of Marginal PMFs

We prove the formula for $p_X(x)$.

Fix a value x in the support of X . Then

$$p_X(x) = \mathbb{P}(X = x).$$

The event $\{X = x\}$ can be written as the union of all events

$$\{X = x, Y = y\}$$

over all possible values y of Y :

$$\{X = x\} = \bigcup_y \{X = x, Y = y\}.$$

Reason:

$$\{X = x\} = (\{X = x\} \cap \mathcal{S}) = \{X = x\} \cap \left(\bigcup_y \{Y = y\}\right) = \bigcup_y \{X = x\} \cap \{Y = y\} = \bigcup_y \{X = x, Y = y\}$$

Thus, applying the probability function to the above equation yields the result. Indeed, these events are pairwise disjoint, because Y cannot take two different values at the same time. Therefore, by the third axiom of probability (countable additivity),

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y).$$

Hence

Definition

If

$$\mathbb{P}(X = x) > 0,$$

then

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

Conditional distributions describe one variable given information about another.

Definition

For continuous random variables X, Y , the joint PDF is a function

$$f_{X,Y}(x,y)$$

such that

$$\mathbb{P}((X, Y) \in A) = \iint_A f_{X,Y}(x,y) \, dx \, dy.$$

Properties:

$$f_{X,Y}(x,y) \geq 0, \quad \iint_{\mathbb{R}^2} f_{X,Y}(x,y) \, dx \, dy = 1.$$

Similarly, for n continuous random variables

$$X_1, \dots, X_n,$$

the joint PDF is a function

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

such that

$$\mathbb{P}((X_1, \dots, X_n) \in A) = \int_A \cdots \int f_{X_1, \dots, X_n}(x_1, \dots, x_n) \, dx_1 \cdots dx_n.$$

Marginal densities are obtained by integrating out the other variable.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) \, dx.$$

Definition

If

$$f_X(x) > 0,$$

then

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Joint Cumulative Distribution Function

Definition: Two Random Variables

For any two random variables X and Y , their joint cumulative distribution function, or joint CDF, is

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y), \quad (x, y) \in \mathbb{R}^2.$$

Since

$$\{X \leq x, Y \leq y\} = \{X \leq x\} \cap \{Y \leq y\},$$

we may equivalently write

$$F_{X,Y}(x, y) = \mathbb{P}(\{X \leq x\} \cap \{Y \leq y\}).$$

Geometrically,

$$F_{X,Y}(x, y)$$

is the probability that the random point (X, Y) lies in the southwest rectangle

$$(-\infty, x] \times (-\infty, y].$$

Important

The joint CDF is defined for every pair of random variables: discrete, continuous, or mixed.

Joint CDF of n Random Variables

Let

$$\mathbf{X} = (X_1, \dots, X_n)$$

be an n -dimensional random vector.

Definition

The joint CDF of $X_1, \dots, X_n$ is

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Equivalently,

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \leq x_i\}\right).$$

Geometrically, it gives the probability that

$$(X_1, \dots, X_n)$$

belongs to the lower-left n -dimensional rectangle

$$(-\infty, x_1] \times \dots \times (-\infty, x_n].$$

When $n = 1$, this reduces to the ordinary CDF:

Basic Properties of a Joint CDF

For every $x, y \in \mathbb{R}$,

① **Bounds:**

$$0 \leq F_{X,Y}(x, y) \leq 1.$$

② **Nondecreasing in each coordinate:**

If $x_1 \leq x_2$, then

$$F_{X,Y}(x_1, y) \leq F_{X,Y}(x_2, y).$$

If $y_1 \leq y_2$, then

$$F_{X,Y}(x, y_1) \leq F_{X,Y}(x, y_2).$$

③ **Right-continuity:**

If

$$x_m \downarrow x \quad \text{and} \quad y_m \downarrow y,$$

then

$$F_{X,Y}(x_m, y_m) \rightarrow F_{X,Y}(x, y).$$

Interpretation

Increasing either threshold enlarges the event whose probability is being calculated, so the joint CDF cannot decrease.

Limits of a Joint CDF

The joint CDF satisfies the following boundary conditions.

If either coordinate approaches $-\infty$, then

$$\lim_{x \rightarrow -\infty} F_{X,Y}(x, y) = 0$$

for every fixed y , and

$$\lim_{y \rightarrow -\infty} F_{X,Y}(x, y) = 0$$

for every fixed x .

If both coordinates approach $+\infty$, then

$$\lim_{\substack{x \rightarrow \infty \\ y \rightarrow \infty}} F_{X,Y}(x, y) = 1.$$

Indeed,

$$\{X \leq x, Y \leq y\} \uparrow \mathcal{S}$$

as $x, y \rightarrow \infty$, so

$$\mathbb{P}(X \leq x, Y \leq y) \longrightarrow \mathbb{P}(\mathcal{S}) = 1.$$

Obtaining Marginal CDFs from the Joint CDF

The marginal CDFs can be obtained by sending the other threshold to $+\infty$.

Marginal CDFs

$$F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y)$$

and

$$F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y).$$

For example,

$$\begin{aligned} \lim_{y \rightarrow \infty} F_{X,Y}(x, y) &= \lim_{y \rightarrow \infty} \mathbb{P}(X \leq x, Y \leq y) \\ &= \mathbb{P}(X \leq x) \\ &= F_X(x). \end{aligned}$$

This works because

$$\{Y \leq y\} \uparrow \mathcal{S} \quad \text{as } y \rightarrow \infty.$$

Similarly,

$$\lim_{x \rightarrow \infty} F_{X,Y}(x, y) = F_Y(y).$$

Marginal CDFs in n Dimensions

Suppose

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

is known.

To obtain the CDF of X_i , send every other coordinate to $+\infty$:

$$F_{X_i}(x_i) = \lim_{\substack{x_j \rightarrow \infty \\ j \neq i}} F_{X_1, \dots, X_n}(x_1, \dots, x_n).$$

More generally, let

$$I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}.$$

The joint CDF of the subvector

$$(X_{i_1}, \dots, X_{i_k})$$

is

$$F_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}, \dots, x_{i_k}) = \lim_{\substack{x_j \rightarrow \infty \\ j \notin I}} F_{X_1, \dots, X_n}(x_1, \dots, x_n).$$

Thus every lower-dimensional marginal CDF can be derived from the full joint CDF.

Properties of an n -Dimensional Joint CDF

The joint CDF

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

has the following properties:

- 1
- 2 It is nondecreasing in each coordinate.
- 3 It is right-continuous in each coordinate.
- 4 If any one coordinate tends to $-\infty$, then

$$0 \leq F_{X_1, \dots, X_n}(x_1, \dots, x_n) \leq 1.$$

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) \rightarrow 0.$$

- 5 If every coordinate tends to $+\infty$, then

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) \rightarrow 1.$$

- 6 It assigns nonnegative probability to every n -dimensional rectangle.

The final property is called the

n -increasing property.

From the Joint CDF to a Joint PDF

Every random vector has a joint CDF, but not every random vector has a joint PDF.

If X and Y have a joint PDF $f_{X,Y}$, then

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) dv du.$$

Thus the joint CDF accumulates the density over the southwest rectangle

$$(-\infty, x] \times (-\infty, y].$$

If $F_{X,Y}$ is sufficiently differentiable, then

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

Connection

The PDF is obtained by differentiating the CDF, while the CDF is obtained by integrating the PDF.

Let

$$a_1 < a_2 \quad \text{and} \quad b_1 < b_2.$$

For the rectangle

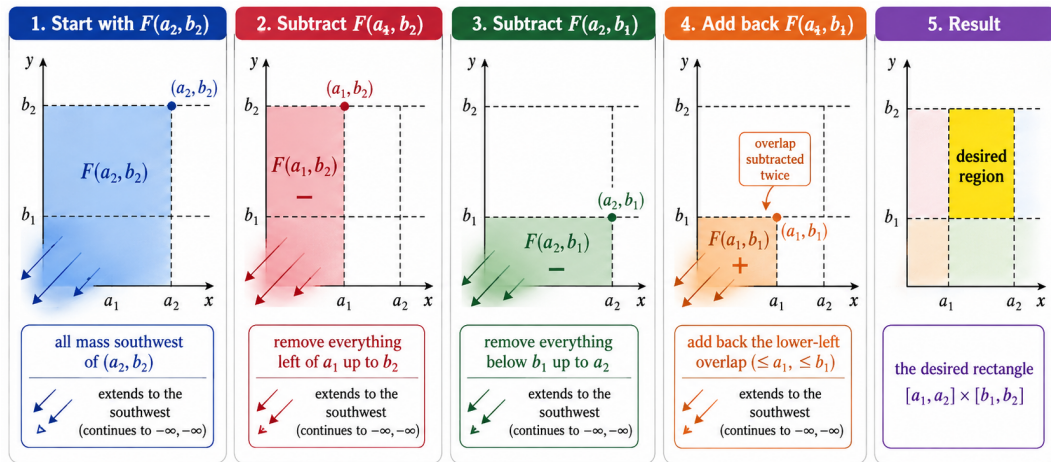
$$a_1 < X \leq a_2, \quad b_1 < Y \leq b_2,$$

we have

$$\begin{aligned} \mathbb{P}(a_1 < X \leq a_2, b_1 < Y \leq b_2) \\ = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1). \end{aligned}$$

This is the two-dimensional inclusion–exclusion formula.

Visualizing the Joint CDF



$$P((X, Y) \text{ in } [a_1, a_2] \times [b_1, b_2]) = F(a_2, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, b_1)$$

★ Joint CDF definition: $F(c, d) = P(X \leq c, Y \leq d)$ = all probability mass in the southwest region with corner (c, d) , extending to $-\infty$ in both x and y .

Joint CDF: One Bounded Interval and One One-Sided Interval

Recall that the marginal CDFs can be recovered from the joint CDF:

$$F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y), \quad F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y).$$

Bounded in X , one-sided in Y :

$$\mathbb{P}(a_1 < X \leq a_2, Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2)$$

$$\begin{aligned} \mathbb{P}(a_1 < X \leq a_2, Y > b_1) &= F_X(a_2) - F_X(a_1) \\ &\quad - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1). \end{aligned}$$

One-sided in X , bounded in Y :

$$\mathbb{P}(X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_2, b_1)$$

$$\begin{aligned} \mathbb{P}(X > a_1, b_1 < Y \leq b_2) &= F_Y(b_2) - F_Y(b_1) \\ &\quad - F_{X,Y}(a_1, b_2) + F_{X,Y}(a_1, b_1). \end{aligned}$$

Joint CDF: When Both Conditions Are One-Sided

There are four possible combinations.

Both are lower-tail events:

$$\mathbb{P}(X \leq a_2, Y \leq b_2) = F_{X,Y}(a_2, b_2)$$

X upper-tail, Y lower-tail:

$$\mathbb{P}(X > a_1, Y \leq b_2) = F_Y(b_2) - F_{X,Y}(a_1, b_2)$$

X lower-tail, Y upper-tail:

$$\mathbb{P}(X \leq a_2, Y > b_1) = F_X(a_2) - F_{X,Y}(a_2, b_1)$$

Both are upper-tail events:

$$\begin{aligned} \mathbb{P}(X > a_1, Y > b_1) &= 1 - F_X(a_1) - F_Y(b_1) \\ &\quad + F_{X,Y}(a_1, b_1). \end{aligned}$$

Deriving the Rectangle Formula

Let

$$a_1 < a_2, \quad b_1 < b_2,$$

and consider the event

$$R = \{a_1 < X \leq a_2, b_1 < Y \leq b_2\}.$$

Define

$$E = \{X \leq a_2, Y \leq b_2\},$$

the large southwest region ending at (a_2, b_2) .

Inside E , the points we do *not* want are

$$L = \{X \leq a_1, Y \leq b_2\}, \quad B = \{X \leq a_2, Y \leq b_1\}.$$

Hence, as an equality of events,

$$R = E \setminus (L \cup B).$$

Therefore,

$$\begin{aligned} \mathbb{P}(R) &= \mathbb{P}(E) - \mathbb{P}(L \cup B) \\ &= \mathbb{P}(E) - \mathbb{P}(L) - \mathbb{P}(B) + \mathbb{P}(L \cap B), \end{aligned}$$

where the second equality follows from inclusion–exclusion.

Continue.

Now,

$$L \cap B = \{X \leq a_1, Y \leq b_1\}.$$

Using

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y),$$

we obtain

$$\begin{aligned} \mathbb{P}(a_1 < X \leq a_2, b_1 < Y \leq b_2) &= F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) \\ &\quad - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1). \end{aligned}$$

Rectangle probability from the joint CDF

$$\mathbb{P}(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1).$$

Example

Suppose (X, Y) has joint PDF

$$f_{X,Y}(x, y) = \begin{cases} 2, & 0 < y < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

PDF verification:

Since $f_{X,Y}(x, y) \in \{0, 2\}$, we only need to check that it integrates to 1:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy dx = \int_0^1 \int_0^x 2 dy dx = \int_0^1 2x dx = \left[x^2 \right]_0^1 = 1. \quad (1)$$

Therefore $f_{X,Y}$ is a valid joint density.

Important

The inequalities defining the support determine the limits of integration:

$$0 < x < 1, \quad 0 < y < x.$$

Deriving the Marginal Densities

To obtain the marginal density of X , integrate out Y :

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy.$$

For a fixed x with $0 < x < 1$, the support requires $0 < y < x$. Therefore,

$$f_X(x) = \int_0^x 2 dy = 2x. \quad (2)$$

Hence,

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, to obtain the marginal density of Y , integrate out X :

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx.$$

For a fixed y with $0 < y < 1$, the support condition $y < x < 1$ gives

$$f_Y(y) = \begin{cases} 2(1 - y), & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Checking the Marginals

Each marginal density must itself integrate to 1.

For X ,

$$\int_0^1 f_X(x) dx = \int_0^1 2x dx = \left[x^2 \right]_0^1 = 1.$$

For Y ,

$$\int_0^1 f_Y(y) dy = \int_0^1 2(1-y) dy = \left[2y - y^2 \right]_0^1 = 1.$$

Therefore,

$$\boxed{\int_{-\infty}^{\infty} f_X(x) dx = \int_{-\infty}^{\infty} f_Y(y) dy = 1}$$

Thus, both f_X and f_Y are valid marginal probability density functions.

Deriving the Joint CDF from the Joint PDF

Recall the definition:

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y).$$

For continuous random variables,

$$F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u,v) dv du.$$

However, the limits must respect the support

$$0 < v < u < 1.$$

Consequently, the calculation depends on the location of (x,y) .

Case 1: $x \leq 0$ or $y \leq 0$.

There is no probability mass southwest of (x,y) , so

$$F_{X,Y}(x,y) = 0.$$

Case 2: $0 < x \leq 1$ and $y \geq x$.

Since every supported point with $X \leq x$ automatically satisfies $Y < X \leq x \leq y$,

$$F_{X,Y}(x,y) = \int_0^x \int_0^u 2 dv du = \int_0^x 2u du = x^2. \quad (3)$$

Thus, $F_{X,Y}(x,y) = x^2$.

Deriving the Joint CDF: The Nontrivial Case

Now suppose

$$0 < y < x \leq 1.$$

The restriction $Y \leq y$ cuts through the support, so the integral must be split at $u = y$:

$$F_{X,Y}(x, y) = \int_0^y \int_0^u 2 \, dv \, du + \int_y^x \int_0^y 2 \, dv \, du.$$

The first term gives

$$\int_0^y 2u \, du = y^2,$$

while the second gives

$$\int_y^x 2y \, du = 2y(x - y).$$

Therefore,

$$\begin{aligned} F_{X,Y}(x, y) &= y^2 + 2y(x - y) \\ &= 2xy - y^2. \end{aligned}$$

Hence,

$$F_{X,Y}(x, y) = 2xy - y^2, \quad 0 < y < x \leq 1.$$

The Complete Joint CDF

Combining all cases gives

$$F_{X,Y}(x,y) = \begin{cases} 0, & x \leq 0 \text{ or } y \leq 0, \\ x^2, & 0 < x \leq 1, \quad y \geq x, \\ 2xy - y^2, & 0 < y < x \leq 1, \\ 2y - y^2, & x \geq 1, \quad 0 < y < 1, \\ 1, & x \geq 1, \quad y \geq 1. \end{cases}$$

The term

$$2y - y^2$$

for $x \geq 1$ arises because once $x \geq 1$, the condition $X \leq x$ places no further restriction on the support:

$$F_{X,Y}(x,y) = \mathbb{P}(Y \leq y) = F_Y(y).$$

Similarly, whenever $y \geq 1$,

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x) = F_X(x).$$

Recovering the PDF from the Joint CDF

For an absolutely continuous random vector,

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

at points where the derivatives exist.

Inside the support, $0 < y < x < 1$, we found

$$F_{X,Y}(x,y) = 2xy - y^2.$$

Differentiate first with respect to y :

$$\frac{\partial}{\partial y} F_{X,Y}(x,y) = 2x - 2y.$$

Then differentiate with respect to x :

$$\frac{\partial}{\partial x} (2x - 2y) = 2.$$

Therefore,

$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) = 2 = f_{X,Y}(x,y), \quad 0 < y < x < 1.$$

Outside the support, the mixed derivative is 0 wherever it exists, which again agrees with the joint density.

Recovering the Marginal CDFs from the Joint CDF

For any joint CDF,

$$F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y)$$

For this example,

$$F_X(x) = \begin{cases} 0, & x \leq 0, \\ x^2, & 0 < x < 1, \\ 1, & x \geq 1. \end{cases}$$

Indeed,

$$F_X(x) = \int_{-\infty}^x f_X(u) du = \int_0^x 2u du = x^2, \quad 0 < x < 1.$$

Similarly,

$$F_Y(y) = \begin{cases} 0, & y \leq 0, \\ 2y - y^2, & 0 < y < 1, \\ 1, & y \geq 1. \end{cases}$$

since $F_Y(y) = \int_0^y 2(1 - v) dv = 2y - y^2$.

Example: Uniform on the Unit Disk

Suppose (X, Y) is uniformly distributed on the unit disk

$$D = \{(x, y) : x^2 + y^2 \leq 1\}.$$

Since the area of the unit disk is π , the joint density is

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Indeed,

$$\iint_{\mathbb{R}^2} f_{X,Y}(x, y) \, dx \, dy = \iint_D \frac{1}{\pi} \, dx \, dy = \frac{1}{\pi} \text{Area}(D) = 1.$$

For a fixed value of $X = x$, the possible values of Y satisfy

$$-\sqrt{1 - x^2} \leq Y \leq \sqrt{1 - x^2}.$$

Important

Knowing X restricts the possible values of Y , so X and Y are dependent.

Let

$$\mathbf{X} = (X_1, \dots, X_n)^\top$$

be a random vector. Its **joint moment generating function (MGF)** is defined by

$$M_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\left[e^{\mathbf{t}^\top \mathbf{X}}\right] = \mathbb{E}\left[e^{t_1 X_1 + \dots + t_n X_n}\right],$$

for all

$$\mathbf{t} = (t_1, \dots, t_n)^\top$$

for which the expectation is finite.

In the two-dimensional case,

$$M_{X,Y}(s, t) = \mathbb{E}\left[e^{sX + tY}\right].$$

Multivariate Moment Generating Function

Let

$$\mathbf{X} = (X_1, \dots, X_n)^\top$$

be a random vector. Its **joint moment generating function (MGF)** is defined by

$$M_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\left[e^{\mathbf{t}^\top \mathbf{X}}\right] = \mathbb{E}\left[e^{t_1 X_1 + \dots + t_n X_n}\right],$$

for all

$$\mathbf{t} = (t_1, \dots, t_n)^\top$$

for which the expectation is finite.

n=2: In the two-dimensional case,

$$M_{X,Y}(s, t) = \mathbb{E}\left[e^{sX + tY}\right].$$

Mixed moments can be recovered from derivatives:

$$\mathbb{E}[X^r Y^q] = \left. \frac{\partial^{r+q}}{\partial s^r \partial t^q} M_{X,Y}(s, t) \right|_{(s,t)=(0,0)}.$$

Example: Joint MGF of Two Independent Normal Variables

Suppose

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2), \quad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2),$$

and assume

$$X \perp Y.$$

The marginal MGFs are

$$M_X(s) = e^{\mu_X s + \frac{1}{2} \sigma_X^2 s^2}, \quad M_Y(t) = e^{\mu_Y t + \frac{1}{2} \sigma_Y^2 t^2}.$$

By independence,

$$M_{X,Y}(s, t) = \mathbb{E}[e^{sX+tY}] = \mathbb{E}[e^{sX} e^{tY}] = \mathbb{E}[e^{sX}] \mathbb{E}[e^{tY}] = M_X(s) M_Y(t).$$

Therefore,

$$\begin{aligned} M_{X,Y}(s, t) &= e^{\mu_X s + \frac{1}{2} \sigma_X^2 s^2} e^{\mu_Y t + \frac{1}{2} \sigma_Y^2 t^2} \\ &= e^{\mu_X s + \mu_Y t + \frac{1}{2} \sigma_X^2 s^2 + \frac{1}{2} \sigma_Y^2 t^2} \\ &= \exp \left\{ \begin{pmatrix} s \\ t \end{pmatrix}^T \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} s \\ t \end{pmatrix}^T \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} \right\}. \end{aligned}$$

Independence

Definition

Random variables X, Y are independent if for all sets A, B ,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Intuitively: knowing one variable gives no information about the other.

Equivalent Definitions of Independence I

For random variables X and Y , the following statements are equivalent.

1.

$$X \perp Y.$$

2. **CDF factorization:** For every $x, y \in \mathbb{R}$,

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$

3a. **PMF factorization (discrete case):**

If X and Y are discrete, then for every $x, y \in \mathbb{R}$,

$$p_{X,Y}(x, y) = p_X(x)p_Y(y).$$

Equivalently, it is enough to verify this for

$$x \in \text{supp}(X), \quad y \in \text{supp}(Y).$$

3b. **PDF factorization (continuous case):**

If (X, Y) has a joint density, then

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

for almost every $(x, y) \in \mathbb{R}^2$.

4. Conditional distribution does not depend on the other variable.

In general, for every Borel set $B \subseteq \mathbb{R}$,

$$\mathbb{P}(Y \in B \mid X) = \mathbb{P}(Y \in B) \quad \text{a.s.}$$

Discrete case: for every x with $p_X(x) > 0$ and every y ,

$$p_{Y|X}(y \mid x) = p_Y(y).$$

Equivalently,

$$\mathbb{P}(Y = y \mid X = x) = \mathbb{P}(Y = y).$$

Continuous case: for f_X -almost every x and almost every y ,

$$f_{Y|X}(y \mid x) = f_Y(y),$$

whenever the conditional density is defined.

5. Factorization of expectations:

For every pair of bounded Borel-measurable functions

$$g, h : \mathbb{R} \rightarrow \mathbb{R},$$

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

More generally, the same factorization holds whenever the relevant expectations exist and are finite.

6. Joint MGF factorization:

If the joint MGF exists on an open neighborhood of $(0, 0)$, then

$$M_{X,Y}(s, t) = M_X(s)M_Y(t)$$

for every (s, t) in that neighborhood, where

$$M_{X,Y}(s, t) = \mathbb{E}\left[e^{sX+tY}\right].$$

Recall that

$$M_X(s) = \mathbb{E}[e^{sX}], \quad M_Y(t) = \mathbb{E}[e^{tY}].$$

Therefore, under independence,

$$M_{X,Y}(s, t) = \mathbb{E}\left[e^{sX+tY}\right] = \mathbb{E}\left[e^{sX}e^{tY}\right] = \mathbb{E}[e^{sX}]\mathbb{E}[e^{tY}] = M_X(s)M_Y(t).$$

Conversely, if the joint MGF exists in a neighborhood of $(0, 0)$ and factorizes as

$$M_{X,Y}(s, t) = M_X(s)M_Y(t),$$

then uniqueness of MGFs implies that the joint distribution is the product of the marginal distributions, and hence

$$X \perp Y.$$

Which Criterion Should We Use?

Different characterizations are useful in different situations.

- ▶ PMF/PDF factorization: usually easiest computationally;
- ▶ conditional distributions: most conceptual;
- ▶ expectation factorization: useful theoretically;
- ▶ MGF factorization: useful for sums and CLT.

Definition

Random variables

$$X_1, \dots, X_n$$

are mutually independent if for every subset

$$i_1, \dots, i_k,$$

$$\mathbb{P}(X_{i_1} \in A_1, \dots, X_{i_k} \in A_k) = \prod_{j=1}^k \mathbb{P}(X_{i_j} \in A_j).$$

Definition

Random variables

$$X_1, \dots, X_n$$

are pairwise independent if every pair

$$(X_i, X_j)$$

is independent.

mutual independence $\implies$ pairwise independence.

But the converse is false.

Pairwise Independent but Not Mutually Independent

Let X and Y be independent random variables with

$$\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}, \quad \mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = \frac{1}{2},$$

and define

$$Z = XY.$$

Then Z also takes values ± 1 with equal probability:

$$\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \frac{1}{2}.$$

Pairwise independence. For $x, z \in \{-1, 1\}$,

$$\mathbb{P}(X = x, Z = z) = \mathbb{P}(X = x, XY = z) = \mathbb{P}(X = x, Y = xz) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = \mathbb{P}(X = x)\mathbb{P}(Z = z)$$

By the same argument,

$$Y \perp Z.$$

Thus X, Y, Z are **pairwise independent**.

But they are not mutually independent.

Since $Z = XY$,

$$XYZ = X Y (XY) = X^2 Y^2 = 1 \quad \text{always.}$$

For instance, $\mathbb{P}(X = 1, Y = 1, Z = 1) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(X = 1) \times \mathbb{P}(Y = 1) \times \mathbb{P}(Z = 1)$.

2D LOTUS

Theorem

For a function $g(X, Y)$,

$$\mathbb{E}[g(X, Y)] = \begin{cases} \sum_x \sum_y g(x, y) p_{X,Y}(x, y), & \text{if } X, Y \text{ are discrete,} \\ \iint g(x, y) f_{X,Y}(x, y) dx dy, & \text{if } X, Y \text{ are continuous.} \end{cases}$$

Meaning

To find the expected value of a function of two random variables, average $g(x, y)$ using the joint distribution of (X, Y) .

Using 2D LOTUS:

$$E(X + Y) = E(X) + E(Y).$$

More generally:

$$E(aX + bY + c) = aE(X) + bE(Y) + c.$$

Importantly: independence is NOT required.

Covariance and Correlation

Definition

The covariance of X, Y is

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

Equivalent formula:

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

Proposition: $\text{Cov}(X, X) = \text{Var}(X)$

Interpretation of Covariance

- ▶ Positive covariance: variables tend to move together;
- ▶ Negative covariance: one tends to increase while the other decreases;
- ▶ Zero covariance: no **linear** relationship (but there may be other forms of nonlinear relationships).

Important

Covariance measures linear dependence only.

Definition

The correlation coefficient is

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Properties:

$$-1 \leq \text{Corr}(X, Y) \leq 1.$$

Proof: Chapter 10a (inequalities) - needs Cauchy-Schwarz inequality.

Theorem

For random variables X, Y ,

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

If X, Y are independent:

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

Theorem

If X, Y are independent, then

$$\text{Cov}(X, Y) = 0.$$

This follows because:

$$E(XY) = E(X)E(Y).$$

Zero Correlation Does Not Imply Independence

Let

$$X \sim N(0, 1),$$

and define

$$Y = X^2.$$

Then:

$$E(X) = 0,$$

and

$$E(XY) = E(X^3) = 0.$$

Hence:

$$\text{Cov}(X, Y) = 0.$$

But:

$$Y$$

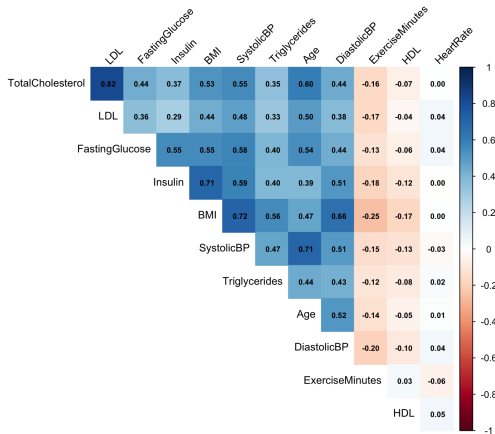
is completely determined by

$$X.$$

Thus X, Y are dependent.

Heatmaps and Correlation Visualization

Now, suppose that instead of 2 variables, we have n variables $X_1, \dots, X_n$. If we consider all the correlations, i.e., $\{\text{Corr}(X_i, X_j) : 1 \leq i \neq j \leq n\}$, then there is $\binom{n}{2}$ distinct correlations, so we can define a matrix whose ij entry is $\text{Corr}(X_i, X_j)$. This matrix is symmetric since $\text{Corr}(X_i, X_j) = \text{Corr}(X_j, X_i)$. There is an interesting way to visualize this matrix, because all entries are between -1 and 1, we can use a heatmap.



Correlation Heatmap

- ▶ A **heatmap** provides a visual summary of pairwise correlations between variables.

- ▶ Each cell shows the correlation coefficient

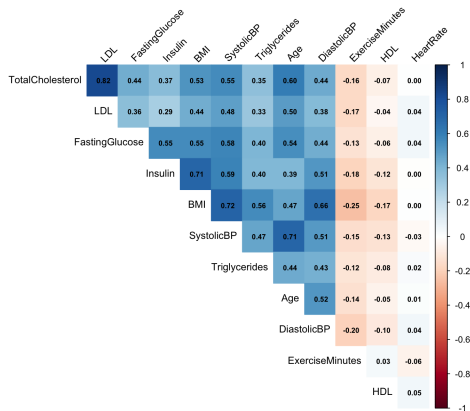
$$r(X, Y)$$

between two variables.

- ▶ Interpretation:

- Blue: positive correlation.
- Red: negative correlation.
- White/light colors: weak or near-zero correlation.

- ▶ The closer $|\text{Corr}(X, Y)|$ is to 1, the stronger the linear relationship.



Multinomial Distribution

Multinomial Distribution (generalization of Binomial Distribution)

In the **Binomial setting**, each trial has two possible outcomes:

$$o_1 : \text{Success} \quad \text{or} \quad o_2 : \text{Failure},$$

with probabilities

$$p_1 = p, \quad p_2 = 1 - p, \quad \text{where} \quad p_1 + p_2 = 1$$

After n independent trials,

$$X_1 := X \quad \text{and} \quad X_2 := n - X$$

count the number of trials resulting in category o_1 : Success and o_2 : Failure, respectively. Note that $X_1 + X_2 = X + n - X = n$

In the **multinomial setting**, each trial can result in one of k (instead of 2) possible categories

$$o_1, \dots, o_k,$$

with probabilities

$$p_1, \dots, p_k, \quad \text{where} \quad p_1 + \dots + p_k = 1$$

After n independent trials,

$$X_1, \dots, X_k$$

count the number of trials resulting in category $o_1, \dots, o_k$. Note that $X_1 + \dots + X_k = n$, since each trial results in exactly one category.

Definition

If $(X_1, \dots, X_k)$ counts the numbers of occurrences of the k categories in n independent trials, then

$$(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, \dots, p_k).$$

$$\mathbb{P}(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k},$$

subject to:

$$x_1 + \dots + x_k = n.$$

Note. The Multinomial distribution is a multivariate distribution $k - \text{dimensional}$, while the Binomial is a univariable distribution $1 - \text{dimensional}$. A multivariate random variable is also called a *random vector*. **Note.** It can be shown using the Multinomial Theorem (an extension of the Binomial Theorem) that the Multinomial PMF sums up to 1. [Click here](#) to see the Multinomial Theorem in Wikipedia.

Theorem

Suppose

$$(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, \dots, p_k), \quad \sum_{i=1}^k p_i = 1.$$

Then,

- ① For every i ,

$$X_i \sim \text{Bin}(n, p_i).$$

- ② More generally, if $A_1, \dots, A_r$ are disjoint subsets of $\{1, \dots, k\}$ and

$$Y_j = \sum_{i \in A_j} X_i, \quad q_j = \sum_{i \in A_j} p_i,$$

then

$$\left(Y_1, \dots, Y_r, n - \sum_{j=1}^r Y_j \right) \sim \text{Mult} \left(n, q_1, \dots, q_r, 1 - \sum_{j=1}^r q_j \right).$$

Multinomial Distribution: Trial Representation

A multinomial experiment consists of n independent trials.

For trial $\ell = 1, \dots, n$, define

$$I_{\ell i} = \begin{cases} 1, & \text{if trial } \ell \text{ results in category } i, \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$X_i = \sum_{\ell=1}^n I_{\ell i}.$$

For every fixed i ,

$$I_{1i}, \dots, I_{ni} \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p_i).$$

Therefore,

$$X_i = \sum_{\ell=1}^n I_{\ell i} \sim \text{Bin}(n, p_i).$$

Thus every individual component of a multinomial vector has a binomial marginal distribution.

Proof of the Grouping Property

Let $A_1, \dots, A_r$ be disjoint subsets of $\{1, \dots, k\}$, and define $Y_j = \sum_{i \in A_j} X_i$. For each trial ℓ , define

$$J_{\ell j} = \sum_{i \in A_j} I_{\ell i}.$$

Since the sets A_j are disjoint, $J_{\ell j} = 1$ exactly when trial ℓ falls into one of the categories in A_j . Hence

$$P(J_{\ell j} = 1) = \sum_{i \in A_j} p_i = q_j.$$

The remaining categories have probability

$$1 - \sum_{j=1}^r q_j.$$

Therefore each trial can be viewed as having the new categories $A_1, \dots, A_r$, remainder, with probabilities

$$q_1, \dots, q_r, 1 - \sum_{j=1}^r q_j.$$

Since the n trials are independent,

$$\left(Y_1, \dots, Y_r, n - \sum_{j=1}^r Y_j \right) \sim \text{Mult} \left(n, q_1, \dots, q_r, 1 - \sum_{j=1}^r q_j \right).$$

Subvectors of a Multinomial Vector

A useful special case is obtained by selecting some of the original coordinates.

Suppose

$$1 \leq i_1 < \dots < i_r \leq k$$

and define

$$X_{\text{rest}} = n - \sum_{j=1}^r X_{i_j}.$$

Then

$$(X_{i_1}, \dots, X_{i_r}, X_{\text{rest}}) \sim \text{Mult} \left(n, p_{i_1}, \dots, p_{i_r}, 1 - \sum_{j=1}^r p_{i_j} \right).$$

Thus a selected subvector is multinomial once all omitted categories are combined into one “remainder” category. In particular, for $i \neq j$,

$$(X_i, X_j, n - X_i - X_j) \sim \text{Mult}(n, p_i, p_j, 1 - p_i - p_j).$$

The grouping property also shows that multinomial categories may be combined. For example, if

$$(X_1, X_2, X_3, X_4) \sim \text{Mult}(n, p_1, p_2, p_3, p_4),$$

then $X_1 + X_2$ counts trials falling in either category 1 or category 2. Therefore,

$$X_1 + X_2 \sim \text{Bin}(n, p_1 + p_2).$$

More generally, for any

$$A \subseteq \{1, \dots, k\},$$

$$\sum_{i \in A} X_i \sim \text{Bin} \left(n, \sum_{i \in A} p_i \right).$$

This follows by treating all categories in A as one category and all categories outside A as another.

Corollary: Means and Variances

Corollary

If

$$(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, \dots, p_k),$$

then, for every i ,

$$E(X_i) = np_i$$

and

$$\text{Var}(X_i) = np_i(1 - p_i).$$

Proof.

Since

$$X_i \sim \text{Bin}(n, p_i),$$

the usual formulas for a binomial random variable immediately give

$$E(X_i) = np_i$$

and

$$\text{Var}(X_i) = np_i(1 - p_i).$$

Covariance Between Multinomial Counts

For $i \neq j$, use the trial representation

$$X_i = \sum_{\ell=1}^n I_{\ell i}, \quad X_j = \sum_{\ell=1}^n I_{\ell j}.$$

Therefore,

$$\text{Cov}(X_i, X_j) = \sum_{\ell=1}^n \sum_{m=1}^n \text{Cov}(I_{\ell i}, I_{mj}).$$

If $\ell \neq m$, the trials are independent, so

$$\text{Cov}(I_{\ell i}, I_{mj}) = 0.$$

Hence

$$\text{Cov}(X_i, X_j) = \sum_{\ell=1}^n \text{Cov}(I_{\ell i}, I_{\ell j}).$$

Covariance Between Multinomial Counts

For the same trial ℓ , categories i and j cannot occur simultaneously when $i \neq j$. Thus

$$I_{\ell i} I_{\ell j} = 0,$$

so

$$E(I_{\ell i} I_{\ell j}) = 0.$$

Also,

$$E(I_{\ell i}) = p_i, \quad E(I_{\ell j}) = p_j.$$

Therefore,

$$\begin{aligned} \text{Cov}(I_{\ell i}, I_{\ell j}) &= E(I_{\ell i} I_{\ell j}) - E(I_{\ell i})E(I_{\ell j}) \\ &= 0 - p_i p_j \\ &= -p_i p_j. \end{aligned}$$

Thus

$$\boxed{\text{Cov}(X_i, X_j) = -np_i p_j, \quad i \neq j.}$$

Why Is the Covariance Negative?

For $i \neq j$,

$$\text{Cov}(X_i, X_j) = -np_i p_j < 0$$

whenever $p_i, p_j > 0$.

This has a natural interpretation.

The total number of trials is fixed:

$$X_1 + \cdots + X_k = n.$$

Thus, if more observations fall into category i , fewer observations are available to fall into the other categories.

So the multinomial counts are generally *not independent*.

For example,

$$X_i \uparrow \implies \text{less room for } X_j \uparrow.$$

This competition for a fixed total produces the negative covariance.

Let

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}, \quad p = \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix}.$$

Then

$$E(X) = np.$$

The covariance matrix has diagonal entries

$$np_i(1 - p_i)$$

and off-diagonal entries

$$-np_i p_j.$$

Therefore,

$$\text{Cov}(X) = n \left[\text{diag}(p) - pp^T \right].$$

Multivariate Normal Distribution

Big Picture

The bivariate normal distribution is the two-dimensional version of the normal distribution. It describes two continuous random variables X and Y whose joint density has a bell-shaped surface over the xy -plane.

A bivariate normal distribution is determined by five parameters:

$$\mu_X, \quad \mu_Y, \quad \sigma_X^2, \quad \sigma_Y^2, \quad \rho.$$

Here,

$$\rho = \text{Corr}(X, Y)$$

controls the strength and direction of the linear relationship between X and Y .

► [Open interactive bivariate normal visualization](#)

Click this link for interactive bivariate normal visualization.

Definition

We say

$$(X, Y) \sim N_2 \left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} \right)$$

if the joint density of (X, Y) is

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\ \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \left(\frac{x-\mu_X}{\sigma_X} \right) \left(\frac{y-\mu_Y}{\sigma_Y} \right) + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\}.$$

where $\sigma_X > 0$, $\sigma_Y > 0$, and $-1 < \rho < 1$.

If

$$(X, Y) \sim N_2 \left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} \right),$$

then:

► The means are

$$\mathbb{E}(X) = \mu_X, \quad \mathbb{E}(Y) = \mu_Y.$$

► The variances are

$$\text{Var}(X) = \sigma_X^2, \quad \text{Var}(Y) = \sigma_Y^2.$$

► The covariance is

$$\text{Cov}(X, Y) = \rho\sigma_X\sigma_Y.$$

► Therefore,

$$\text{Corr}(X, Y) = \rho.$$

A beautiful property of the bivariate normal distribution is that the marginal distributions are normal.

Marginals

If

$$(X, Y) \sim N_2 \left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} \right),$$

then

$$X \sim N(\mu_X, \sigma_X^2)$$

and

$$Y \sim N(\mu_Y, \sigma_Y^2).$$

So the joint distribution is bivariate normal, but each marginal variable individually is ordinary normal.

Another beautiful property is that the conditional distributions are also normal.

Conditionals

If

$$(X, Y)$$

is bivariate normal, then

$$X \mid Y = y \sim N \left(\mu_X + \rho \frac{\sigma_X}{\sigma_Y} (y - \mu_Y), \sigma_X^2 (1 - \rho^2) \right)$$

and

$$Y \mid X = x \sim N \left(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X), \sigma_Y^2 (1 - \rho^2) \right).$$

The conditional mean is linear in the observed value.

Independence and Zero Correlation

For general random variables,

$$\text{Corr}(X, Y) = 0$$

does not necessarily imply independence.

But for the bivariate normal distribution, it does.

Important Gaussian Property

If (X, Y) is bivariate normal, then

$$X \text{ and } Y \text{ independent} \iff \rho = 0.$$

When $\rho = 0$, the joint density factors:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

So in the bivariate normal case,

$$\text{uncorrelated} \iff \text{independent}.$$

Temporary page!

$\text{\LaTeX}$ was unable to guess the total number of pages correctly. As there was some unprocessed data have been added to the final page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus page will go away, because $\text{\LaTeX}$ now knows the number of pages to expect for this document.