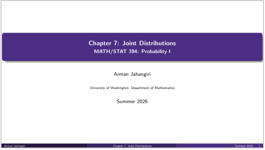
chapter7_joint_distributions (2)

Monday, August 10, 2026 11:55 AM

chapter7_joint_distribu...



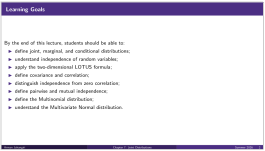
Chapter 7: Joint Distributions

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington Department of Mathematics

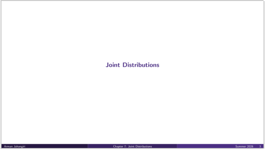
Summer 2026



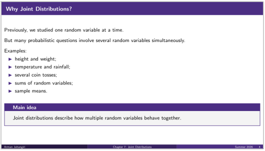
Learning Goals

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.

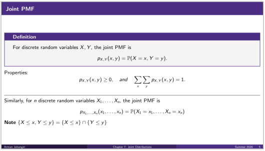
But many probabilistic questions involve several random variables simultaneously.

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.



Joint PMF

Definition

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

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

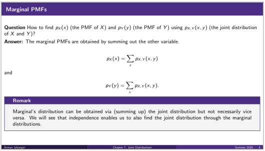
Properties:

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

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

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

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



Marginal PMFs

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.

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

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

For example,

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

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:

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

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

Conditional PMFs

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.

Joint PDFs

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:

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

Similarly, for n continuous random variables

$$X_1, \ldots, X_n,$$

the joint PDF is a function

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

such that

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

Marginal PDFs

Marginal densities are obtained by integrating out the other variable.

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

and

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

Conditional PDFs

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

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

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

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

Equivalently,

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

Geometrically, it gives the probability that

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

belongs to the lower-left n-dimensional rectangle

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

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

Basic Properties of a Joint CDF

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

1. **Bounds:**

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

2. **Nondecreasing in each coordinate:**
   If $x_1 \le x_2$, then

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

   If $y_1 \le y_2$, then

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

3. **Right-continuity:**
   If

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

   then

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

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

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

for every fixed y, and

$$\boxed{\lim_{\underline{y \to -\infty}} F_{X,Y}(x, y) = \enclose{circle}{0}}$$

for every fixed x.
If both coordinates approach $+\infty$, then

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

Indeed,

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

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

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

Obtaining Marginal CDFs from the Joint CDF

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

Marginal CDFs

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

and

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

For example,

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

This works because

$$\{Y \le y\} \uparrow \mathcal{S} \qquad \text{as } y \to \infty.$$

Similarly,

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

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 $+\infty$:

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

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

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

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. $$0 \le F_{X_1,\dots,X_n}(x_1, \dots, x_n) \le 1.$$
2. It is nondecreasing in each coordinate.
3. It is right-continuous in each coordinate.
4. If any one coordinate tends to $-\infty$, then

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

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

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

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

The final property is called the

$$\boxed{n\text{-increasing property.}}$$

August 10

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

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

Thus the joint CDF accumulates the density over the southwest rectangle

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

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

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

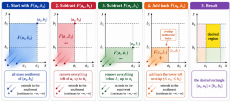
Connection

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

Probabilities of Rectangles from the Joint CDF

Let
$$a_1 < a_2 \quad \text{and} \quad b_1 < b_2.$$
For the rectangle
$$a_1 < X \le a_2, \qquad b_1 < Y \le b_2,$$
we have
$$\boxed{\begin{aligned}&\mathbb{P}(a_1 < X \le a_2,\ b_1 < Y \le b_2)\\ &\quad = F_{X,Y}(a_2,b_2) - F_{X,Y}(a_1,b_2) - F_{X,Y}(a_2,b_1) + F_{X,Y}(a_1,b_1).\end{aligned}}$$
This is the two-dimensional inclusion–exclusion formula.

Visualizing the Joint CDF


1. Start with $F(a_2,b_2)$
2. Subtract $F(a_1,b_2)$
3. Subtract $F(a_2,b_1)$
4. Add back $F(a_1,b_1)$
5. Result
y
x
b_2
b_1
a_1
a_2
(a_2,b_2)
$F(a_2,b_2)$
(a_1,b_2)
$F(a_1,b_2)$
−
(a_2,b_1)
$F(a_2,b_1)$
−
overlap subtracted twice
(a_1,b_1)
$F(a_1,b_1)$
+
desired region
all mass southwest of (a_2,b_2)
remove everything left of a_1 up to b_2
remove everything below b_1 up to a_2
add back the lower-left overlap ($\le a_1$, $\le b_1$)
the desired rectangle $[a_1,a_2]\times[b_1,b_2]$
extends to the southwest (continues to $-\infty,-\infty$)

$P(X\in[a,b])$

$\mathbb{P}((X,Y)\in[a_1,b_1]\times[a_2,b_2])$

$= \mathbb{P}(a_1\le X\le b_1,\ a_2\le Y\le b_2)$

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

★ Joint CDF definition: $F(c,d) = P(X\le c, Y\le d)$ = 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:
$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y), \qquad F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y).$$

Bounded in X, one-sided in Y:

$-\infty < Y \le b_2$

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

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

One-sided in X, bounded in Y:

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

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

exercise

Joint CDF: When Both Conditions Are One-Sided

There are four possible combinations.

Both are lower-tail events:
$$\boxed{\mathbb{P}(X \le a_2,\ Y \le b_2) = F_{X,Y}(a_2,b_2)}$$

X upper-tail, Y lower-tail:
$$\boxed{\mathbb{P}(X > a_1,\ Y \le b_2) = F_Y(b_2) - F_{X,Y}(a_1,b_2)}$$

X lower-tail, Y upper-tail:
$$\boxed{\mathbb{P}(X \le a_2,\ Y > b_1) = F_X(a_2) - F_{X,Y}(a_2,b_1)}$$

Both are upper-tail events:
$$\boxed{\begin{aligned}\mathbb{P}(X > a_1,\ Y > b_1) &= 1 - F_X(a_1) - F_Y(b_1)\\ &\quad + F_{X,Y}(a_1,b_1).\end{aligned}}$$

exercise

Deriving the Rectangle Formula

Let
$$a_1 < a_2, \qquad b_1 < b_2,$$
and consider the event
$$R = \{a_1 < X \le a_2,\ b_1 < Y \le b_2\}.$$ ✓

Define
$$E = \{X \le a_2,\ Y \le b_2\},$$
the large southwest region ending at (a_2,b_2).
Inside E, the points we do *not* want are
$$L = \{X \le a_1,\ Y \le b_2\}, \qquad B = \{X \le a_2,\ Y \le b_1\}.$$
Hence, as an equality of events,
$$\boxed{R = E \setminus (L \cup B).}$$
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,
$$L \cap B = \{X \le a_1,\ Y \le b_1\}.$$ ✓

Using
$$F_{X,Y}(x,y) = \mathbb{P}(X \le x,\ Y \le y),$$
we obtain
$$\begin{aligned}\mathbb{P}(a_1 < X \le a_2,\ b_1 < Y \le b_2) &= F_{X,Y}(a_2,b_2) - F_{X,Y}(a_1,b_2)\\ &\quad - F_{X,Y}(a_2,b_1) + F_{X,Y}(a_1,b_1).\end{aligned}$$

Rectangle probability from the joint CDF

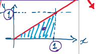
$$\boxed{\mathbb{P}(a_1 < X \le a_2,\ b_1 < Y \le b_2) = F_{X,Y}(a_2,b_2) - F_{X,Y}(a_1,b_2) - F_{X,Y}(a_2,b_1) + F_{X,Y}(a_1,b_1).}$$

Goal: find $f_X, f_Y, F_{XY}, F_X, F_Y$

Example

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

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

$x^2]_0^1$

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

Important

The inequalities defining the support determine the limits of integration:
$$0 < x < 1, \qquad 0 < y < x.$$

Deriving the Marginal Densities

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

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

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

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

$$\boxed{f_X(x) = \int_0^x 2\,dy = 2x \qquad 0<x<1} \tag{2}$$

Hence,

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases} \Rightarrow \int_0^1 2x = x^2\Big]_0^1 = 1$$

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

$$f_Y(y) = \int_y^1 2\,dx = 2x\Big]_y^1 = 2-2y = 2(1-y)$$

$$\boxed{0<y<1}$$

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

$$x,y \in (0,1) \Rightarrow f_X(x)\cdot f_Y(y) = (2x)(2\cdot(1-y)) = 4x(1-y) \neq 2 = f_{X,Y}(x,y)$$

$$0<x,y<1$$

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.

where $f(u,v) = 2\ (\neq 0)$

Deriving the Joint CDF from the Joint PDF

Recall the definition:

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

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt$$

For continuous random variables,

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

However, the limits must respect the support

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

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

Case 1: $x \le 0$ **or** $y \le 0$.

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

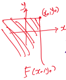
$$F_{X,Y}(x,y) = 0.$$

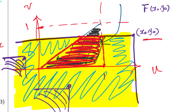
Case 2: $0 < x \le 1$ **and** $y \ge x$.

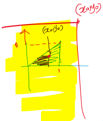
Since every supported point with $X \le x$ automatically satisfies $Y < X \le x \le y$,

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

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







Deriving the Joint CDF: The Nontrivial Case

Now suppose

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

The restriction $Y \le y$ 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,

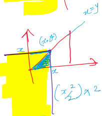
$$F_{X,Y}(x,y) = y^2 + 2y(x-y)$$
$$= 2xy - y^2.$$

Hence,

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

$$F_{X,Y}(x,y) = \begin{cases} 0 & x<0 \text{ or } y<0 \\ x^2 & 0<x,y<1,\ x\le y \\ 1 & x,y \ge 1 \end{cases}$$

$0<x,y<1$ → $x>y$, $x<y$



$$\left(\frac{x^2}{2}\right)\times 2 = x^2$$

$$F(x,y) = \iint 2\,du\,dv = 2\boxed{\iint 1\,du\,dv} = 2\times\left(\frac{x^2}{2}\right) = x^2$$

The Complete Joint CDF

Combining all cases gives

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

$0<x<1$ $y>1$

The term

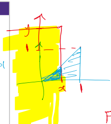
$$2y - y^2$$

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

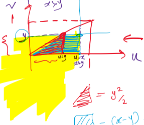
$$F_{X,Y}(x,y) = \mathbb{P}(Y \le y) = F_Y(y).$$

Similarly, whenever $y \ge 1$,

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



$$F(x,y) = \iint 2\,du\,dv$$

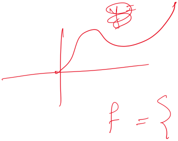


$$▲ = \frac{y^2}{2}$$

$$▨ = (x-y)\times y$$

$$\text{add} \Rightarrow \frac{y^2}{2} + xy - y^2 = \left(xy - \frac{y^2}{2}\right) \text{area}$$

$$\underset{\text{(case 2)}}{\Rightarrow \iint 2\,du\,dv} = 2\cdot\left(xy - \frac{y^2}{2}\right) = 2xy - y^2$$



$$f = \begin{cases} 2 \\ 0 \end{cases}$$

$\rightarrow f_X$ $\quad f_Y$ $\quad F_{X,Y}$

$f_{X,Y}$ given

$$F_X(x) \begin{cases} \text{①}\ \lim\limits_{y\to\infty} F_{X,Y}(x,y) \\ \boxed{\int_{-\infty}^{x} f_X(u)\,du} \end{cases}$$

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:

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

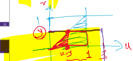
Then differentiate with respect to x:

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

Therefore,

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

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



$$▨ = (1-y)\times y$$

$$▲ = \frac{y^2}{2}$$

$$\Rightarrow y - y^2 + \frac{y^2}{2} = \underbrace{y - \frac{y^2}{2}}_{\text{area}}$$

$$2 \times \text{area} = 2\times\left(y - \frac{y^2}{2}\right) = 2y - y^2$$

Recovering the Marginal CDFs from the Joint CDF

For any joint CDF,

Method Ⅰ

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

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

$\boxed{x<0,\ y>1}$

For this example,

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

$0<x<1$ $\quad y\to\infty$ $\quad (y>1)$

Method Ⅱ

Indeed,

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

$+\ F_X(x) = 1$ $\quad x>1$

Similarly,

yourself $\Rightarrow$

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

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

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

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

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

✓

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

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

Important

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

Multivariate Moment Generating Function

$x \in \mathbb{R}^n \Rightarrow x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

MGF: $M_X(s) = \mathbb{E}[e^{sX}]$

Let $x^T = (x_1, \dots, x_n)$

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

two variable case

X, Y

s, t

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

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

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

for all

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

for which the expectation is finite.
In the two-dimensional case,

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

$t_1 X_1 + \cdots + t_n X_n = \underset{1 \times n}{t^T} \cdot \underset{n \times 1}{X}$

$\vec{t} = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$

$x, y \in \mathbb{R}^n \Rightarrow \langle \vec{x}, \vec{y} \rangle = \vec{x} \cdot \vec{y} = \sum_{i=1}^{n} x_i y_i = x^t \cdot y$

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

$\mathbb{E}\left[X_1^{r_1} \cdots X_n^{r_n}\right]$

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

for all

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

single variable

$M_X^{(n)}(0) = \mathbb{E}(X^n)$

for which the expectation is finite.
n=2: In the two-dimensional case,

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

Mixed moments can be recovered from derivatives:

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

Example: Joint MGF of Two Independent Normal Variables

Suppose

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

and assume

$$X \perp Y.$$

Skip

The marginal MGFs are

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

By independence,

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

Therefore,

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

$A, B \in \mathcal{F} \Rightarrow A \perp B$ when $\underbrace{\mathbb{P}(A|B) = \mathbb{P}(A)}_{\equiv}$

" " $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \leftarrow$

sample space

$Y, X : \underset{(\text{or } S^{\mathbb{R}})}{\Omega} \longrightarrow \mathbb{R}$

Independence

$\mathbb{P}(\quad) \; \in \mathcal{F}$

$X \perp Y$ (Discrete case) $\mathbb{P}(\{X=x\} \cap \{Y=y\}) = \mathbb{P}(\{X=x\}) \cdot \mathbb{P}(\{Y=y\})$

$\forall x, y \in \mathbb{R}$ $\boxed{P_{X,Y}(x,y) = P_X(x) \cdot P_Y(y)}$

Def: $\forall x, y \in \mathbb{R}$

Independence

$\forall x, y \in \mathbb{R} \quad f_{X,Y}(x,y) = f_X(x) f_Y(y)$

Definition

Random variables X, Y are independent if for all sets $A, B \subseteq \mathbb{R}$

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

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

or $\mathbb{P}(X \in A \mid Y \in B) = \mathbb{P}(X \in A)$ $\boxed{\forall A, B \subseteq \mathbb{R}}$

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 $x, y \in \mathbb{R}$,
PMF $$p_{X,Y}(x, y) = p_X(x) p_Y(y).$$
Equivalently, it is enough to verify this for
$$x \in \text{supp}(X), \qquad y \in \text{supp}(Y).$$
3b. **PDF factorization (continuous case):**
If (X, Y) has a joint density, then
PDF $$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$
for almost every $(x, y) \in \mathbb{R}^2$.

Equivalent Definitions of Independence II

4. **Conditional distribution does not depend on the other variable.**
In general, for every Borel set $B \subseteq \mathbb{R}$,

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

4a **Discrete case:** for every x with $p_X(x) > 0$ and every y,

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

Equivalently,

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

4b **Continuous case:** for f_X-almost every x and almost every y,

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

whenever the conditional density is defined.

$P(A|B) = P(A)$

5. **Factorization of expectations:**
For every pair of bounded Borel-measurable functions

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

skip

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

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

$$\underbrace{\iint_{\mathbb{R}\mathbb{R}} g(x)h(y) f_{X,Y}(x,y)\,dx\,dy}_{\mathbb{E}(g(X)h(Y))} = \underbrace{\left(\int g(x) f_X(x)\,dx\right)}_{\mathbb{E}(g(X))} \underbrace{\left(\int h(y) f_Y(y)\,dy\right)}_{\mathbb{E}(h(Y))}$$

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

$g(X) = e^{sX}$, $g(Y) = e^{tY}$

Equivalent Definitions of Independence III

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

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

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

$$X \perp Y.$$

$$M_{X,Y}(s,t) = \mathbb{E}\left(e^{sX+tY}\right) = M_X(s)\cdot M_Y(t) = \mathbb{E}(\underbrace{e^{sX}}_{g(X)})\cdot\mathbb{E}(\underbrace{e^{tY}}_{g(Y)})$$

$$f_{X,Y}(x) = \begin{cases} x^2+y^2+2xy = g(x) \times g(y) \end{cases} \quad 0<x<y<1$$

Which Criterion Should We Use?

example) (indp. r.v.s.)

$$f_{X,Y}(x,y) = \begin{cases} 4xy & 0<x,y<1 \\ 0 & \text{o.w.} \end{cases} \qquad X, Y$$

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.

$$\Rightarrow f_X(x) = \int_0^1 4xy\,dy = 4x\underbrace{\int_0^1 y\,dy}_{1/2} = 2x \quad (0<x<1)$$

$$f_Y(y) = \int_0^1 4xy\,dx = 4y\int_0^1 x\,dx = 2y \quad (0<y<1)$$

$$\Rightarrow f_X(x) \times f_Y(y) = (2x)\cdot(2y) = 4xy \quad \text{when } 0<x,y<1$$

$$f_X(x)\cdot f_Y(y) = 0 \quad \text{if} \quad x \text{ or } y \notin (0,1)$$

$$\Rightarrow f_X(x)\cdot f_Y(y) = \begin{cases} 4xy & 0<x,y<1 \\ 0 & \text{o.w.} \end{cases} = f_{X,Y}(x,y)$$

X, Y independent $\forall x,y$

Independence of Several Random Variables

Definition

Random variables

$$X_1, \ldots, X_n$$

are mutually independent if for every subset

$$\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$$

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

$$\left(\text{or } P_{X_{i_1} \ldots X_{i_k}}(x_{i_1}, \ldots, x_{i_k}) = P_{X_{i_1}}(x_{i_1}) \times \cdots \times P_{X_{i_k}}(x_{i_k})\right)$$

e.g.) three variables X_1, X_2, X_3 are mutually independent if:

$X_1 \perp X_2 \perp X_3$

$\{1,2,3\}$ ① $P_{X_1,X_2,X_3}(x_1,x_2,x_3) = P_{X_1}(x_1)P_{X_2}(x_2)P_{X_3}(x_3)$ $\forall x_1, x_2, x_3 \in \mathbb{R}$

$\{1,2\}$ and ② $P_{X_1,X_2}(x_1,x_2) = P_{X_1}(x_1)P_{X_2}(x_2)$ $\forall x_1, x_2 \in \mathbb{R}$ ← $X_1 \perp X_2$

$\{1,3\}$ and ③ $P_{X_1,X_3}(x_1,x_3) = P_{X_1}(x_1)P_{X_3}(x_3)$ $\forall x_1, x_3 \in \mathbb{R}$ ← $X_1 \perp X_3$

$\{2,3\}$ and ④ $P_{X_2,X_3}(x_2,x_3) = P_{X_2}(x_2)P_{X_3}(x_3)$ $\forall x_2, x_3 \in \mathbb{R}$ ← $X_2 \perp X_3$

Pairwise Independence

Definition

Random variables

$$X_1, \ldots, X_n$$

are pairwise independent if every pair

$$(X_i, X_j) \quad i \neq j$$

is independent.

e.g.) X_1, X_2, X_3 are pairwise independent if:

① $P_{X_1,X_2}(x_1,x_2) = P_{X_1}(x_1)P_{X_2}(x_2)$ $(\forall x_1, x_2 \in \mathbb{R})$, i.e., $X_1 \perp X_2$

② $P_{X_1,X_3}(x_1,x_3) = P_{X_1}(x_1)P_{X_3}(x_3)$ $(\forall x_1, x_3 \in \mathbb{R})$, i.e., $X_1 \perp X_3$

③ $P_{X_2,X_3}(x_2,x_3) = P_{X_2}(x_2)P_{X_3}(x_3)$ $(\forall x_2, x_3 \in \mathbb{R})$, i.e., $X_2 \perp X_3$

$$\text{mutual independence} \implies \text{pairwise independence}.$$

But the converse is false.

Pairwise Independent but Not Mutually Independent

Let X and Y be independent random variables with $X \perp Y$

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

and define

$$Z = XY.$$

$Z = 1 \iff X\cdot Y = 1 \iff X=Y=1$ ← 1/4 or $X=Y=-1$ ← 1/4

Then Z also takes values ± 1 with equal probability:

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

$\{Y = xz\} \equiv \{Y = z\}$?? since $x, z \in \{-1, 1\}$ & Y symmetric ....

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

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

$= P(X=x)\cdot P(Y=xz) \Rightarrow X \perp Z$

By the same argument,

$$Y \perp Z.$$

$x\cdot Y = z$, $Y = z/x = z\cdot x$

Thus X, Y, Z are **pairwise independent.**
But they are not mutually independent.
Since $Z = XY$,

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

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

1/2 1/2 1/2

1, 1
−1, −1
1, −1
−1, 1

$$= P(X=1, Y=1, XY=1) = P(X=1, Y=1) = P(X=1) \times P(Y=1) = 1/2 \times 1/2 = 1/4$$

$$\{X=1, Y=1, XY=1\} = \{X=1, Y=1\}$$

2D LOTUS

2D LOTUS

E

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).

sigle variable:

$$\mathbb{E}(X) = \begin{cases} \sum_{x_i} x_i p(x_i) & \text{Disc.} \\ \int_{\mathbb{R}} x f(x) dx & \text{cont.} \end{cases}$$

$$\mathbb{E}(g(X)) = \begin{cases} \sum_{x_i} g(x_i) P(X=x_i) & \text{Disc.} \\ \int_{\mathbb{R}} g(x) f(x) dx & \text{cont.} \end{cases}$$

$$\mathbb{E}(\underbrace{X \cdot Y}_{g(X,Y)}) = \begin{cases} \sum_i \sum_j \underbrace{x_i y_j}_{g(x_i, y_j)} \mathbb{P}(X=x_i, Y=y_j) & \text{disc.} \\ \int_{\mathbb{R}} \int_{\mathbb{R}} \underbrace{x \cdot y}_{g(x,y)} \cdot f_{X,Y}(x,y)\, dx dy & \text{cont.} \end{cases}$$

$$\mathbb{E}(X^2 + Y^2) = \sum_{j=1}^{6} \sum_{i=1}^{6} (x_i^2 + y_j^2) \frac{1}{36}$$

$P(X=i, Y=j)$ $\forall i,j \in \{1,\dots,6\}$

X : first roll die

Y : second die roll

$$= \mathbb{E}(X^2) + \mathbb{E}(Y^2)$$

$$= \sum x_i^2 \underbrace{p(x_i)}_{1/6} + \sum y_j^2 \underbrace{p(y_j)}_{1/6}$$

Linearity of Expectation Revisited

(linearity of expectation in the single variable case) $\mathbb{E}(aX+b) = a\mathbb{E}(X) + b$ $\forall a, b \in \mathbb{R}$

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.

X , Y $X\uparrow \Rightarrow Y$?

Covariance and Correlation

Covariance

Definition

The covariance of X, Y is

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

$(X - EX)$

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

Equivalent formula:

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

$X = Y$

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

$$\mathrm{Var}(X) = \mathbb{E}((X - EX)^2) = \mathbb{E}((X - EX)\cdot(X - EX))$$

$$= E(X^2) - (EX)^2$$

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).

if $X \perp Y \Rightarrow \mathbb{E}(XY) = (EX)(E(Y))$

$$\Rightarrow \underbrace{EXY - (EX)(E(Y))}_{\mathrm{Cov}(X,Y)} = 0$$

Important

Covariance measures linear dependence only.

$\mathrm{cov}(X,Y) = \mathbb{E}(XY)$

$\mathrm{cov}(X,Y) = 1 \Rightarrow X = aY + b$

Correlation

Definition

The correlation coefficient is

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

Properties:

$$-1 \le \mathrm{Corr}(X, Y) \le 1.$$

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

Variance of Sums

Theorem

For random variables X, Y,

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

If X, Y are independent:

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

$\mathrm{var}(X+Y)$

$$= \mathbb{E}\left[\left[(X+Y) - \mathbb{E}(X+Y)\right]^2\right]$$

$$= \mathbb{E}\left[\left[(X-\mu_X) + (Y-\mu_Y)\right]^2\right] = \mathbb{E}\left[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)\right] = \underbrace{\mathbb{E}((X-\mu_X)^2)}_{\mathrm{Var}(X)} + \underbrace{\mathbb{E}((Y-\mu_Y)^2)}_{\mathrm{Var}(Y)} + \underbrace{2\,\mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]}_{2\cdot\mathrm{Cov}(X,Y)}$$

Generalization: $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_j \sum_i \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum\sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$

Independence Implies Zero Correlation

Theorem

If X, Y are independent, then

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

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:

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

But:

$$Y$$

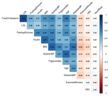
is completely determined by

$$X.$$

Thus X, Y are dependent.

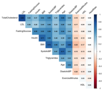
Heatmaps and Correlation Visualization

Now, suppose that instead of 2 variables, we have n variables $X_1, \dots, X_n$. If we consider all the correlations, i.e., $\{\mathrm{Corr}(X_i, X_j) \ : \ 1 \le i \ne j \le n\}$, then there is $\binom{n}{2}$ distinct correlations, so we can define a matrix whose ij entry is $\mathrm{Corr}(X_i, X_j)$. This matrix is symmetric since $\mathrm{Corr}(X_i, X_j) = \mathrm{Corr}(X_j, X_i)$. 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 $|\mathrm{Corr}(X,Y)|$ 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:

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

with probabilities

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

After n independent trials,

$$X_1 := X \qquad \textit{and} \qquad X_2 := n - X$$

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

$$p_1, \dots, p_k, \qquad \textit{where} \qquad p_1 + \cdots + p_k = 1$$

After n independent trials,

$$X_1, \dots X_k$$

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

Multinomial PMF

Definition

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

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

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

subject to:

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

Note. The Multinomial distribution is a multivariate distribution $k-\textit{dimensional}$, while the Binomial is a univariable distribution $1-\textit{dimensional}$. 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.

Moments of the Multinomial

For

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

$$E(X_i) = np_i,$$

$$\mathrm{Var}(X_i) = np_i(1-p_i),$$

and for

$$i \ne j,$$

$$\boxed{\mathrm{Cov}(X_i, X_j) = -np_ip_j.}$$

Multivariate Normal Distribution

Bivariate 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,

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

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.

Bivariate Normal Distribution: Density

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 $-1 < \rho < 1$.

Bivariate Normal Distribution: Main Properties

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
$$\mathbb{E}(X) = \mu_X, \qquad \mathbb{E}(Y) = \mu_Y.$$
- The variances are
$$\operatorname{Var}(X) = \sigma_X^2, \qquad \operatorname{Var}(Y) = \sigma_Y^2.$$
- The covariance is
$$\operatorname{Cov}(X, Y) = \rho\sigma_X\sigma_Y.$$
- Therefore,
$$\operatorname{Corr}(X, Y) = \rho.$$

Marginal Distributions

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

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

and

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

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

Conditional Distributions

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

Conditionals

If

$$(X, Y)$$

is bivariate normal, then

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

and

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

The conditional mean is linear in the observed value.

Independence and Zero Correlation

For general random variables,

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

does not necessarily imply independence.

But for the bivariate normal distribution, it does.

Important Gaussian Property

If (X, Y) is bivariate normal, then

$$\boxed{X \text{ and } Y \quad \textit{independent} \quad \Longleftrightarrow \quad \rho = 0.}$$

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

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

So in the bivariate normal case,

$$\text{uncorrelated} \quad \Longleftrightarrow \quad \text{independent}.$$