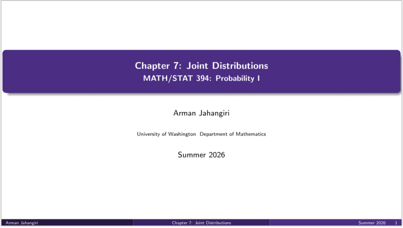
chapter7_joint_distributions (1)

Wednesday, August 5, 2026 12:40 PM

chapter7_joint_distribu...



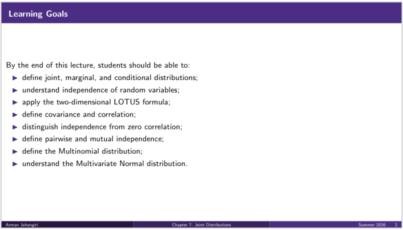
Chapter 7: Joint Distributions

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington Department of Mathematics

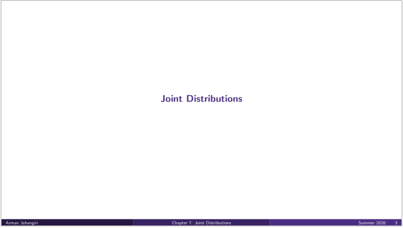
Summer 2026



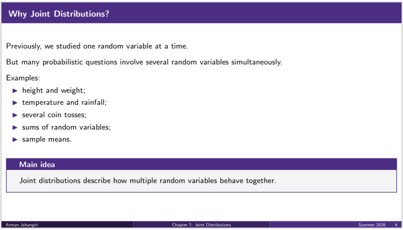
Learning Goals

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

$A, B \in \mathcal{F}$ $\quad \mathbb{P}(A|B) = \boxed{\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}}$ ←

$\underbrace{Y \,|\, \boxed{X=x}} \rightarrow \{X=x\}$ ←

Definition

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

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

$\left\{ \begin{array}{l} Z = X+Y \quad \boxed{X+Y=3} \\ \mathbb{P}(Y=1 \,|\, Z=3) \end{array} \right.$ ⇓ $Y=3-X$

$\mathbb{P}(Y=1 \,|\, Z=3)$

$= \mathbb{P}(3-X=1 \,|\, Z=3)$

$= \mathbb{P}(X=2 \,|\, X+Y=3)$

Conditional distributions describe one variable given information about another.

Conditional PMF $\rightarrow \underbrace{P_{Y|X}}_{\text{a new distribution}}(y) := \mathbb{P}\Big(\underbrace{\{Y=y\}}_{A} \,\Big|\, \underbrace{\{X=x\}}_{B}\Big)$

$= \mathbb{P}(Y=y, X=x)/\mathbb{P}(X=x) \triangleq \frac{P_{X,Y}(x,y)}{P_X(x)}$

Joint PDFs

Definition

For continuous random variables X, Y, the joint PDF is a function
$$f_{X,Y}(x, y)$$
such that
$$\mathbb{P}((X, Y) \in A) = \iint_A f_{X,Y}(x, y)\,dx\,dy.$$

$A \subseteq \mathbb{R}^2$

Joint PMFs
Joint CDFs
← Joint PDFs.

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

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

Joint PMF (Discrete case)

⇓

Marginal PMF

Marginal densities are obtained by integrating out the other variable.

$$\left\{ \begin{array}{l} \underbrace{f_X(x)}_{\text{marginal pdf}} = \int_{-\infty}^{\infty} \underbrace{f_{X,Y}(x, y)}_{\text{Joint pdf}}\,dy \\ \text{and} \\ f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx. \end{array} \right.$$

$$\left\{ \begin{array}{l} P_X(x) = \sum_y P_{X,Y}(x,y) \\ P_Y(y) = \sum_x P_{X,Y}(x,y) \end{array} \right.$$

$f_X(x) = \iint_{\mathbb{R}\mathbb{R}} f_{X,Y,Z}(x,y,z)\,dy\,dz$

$X_1 \ldots X_i \quad X_n$

$X \rightarrow P(X=x) \quad E(X)$
$P_X(x) \quad Var(X)$
$f_X(x) \quad M_X(t) = \mathbb{E}(e^{tX})$
$F_X(x) \quad - - -$

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

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

Conditional PDFs

$P(A|B) = \frac{P(A\cap B)}{P(B)} \longleftrightarrow P(B|A) = \frac{P(B\cap A)}{P(A)}$

Disc.

Definition

If Cont. case $f_X(x) > 0$, ✓

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

≥ 0 ≥ 0 ≥ 0

$$\left\{ \begin{array}{l} \boxed{P_{Y|X}(y) = \frac{P_{X,Y}(x,y)}{P_X(x)}} \\ \boxed{P_{X|Y}(x) = \frac{P_{X,Y}(x,y)}{P_Y(y)}} \end{array} \right.$$

$$\int_{\mathbb{R}} f_{Y|X}(y)\,dy = \int_{\mathbb{R}} \frac{f_{X,Y}(x,y)}{f_X(x)}\,dy = \frac{1}{\cancel{f_X(x)}} \underbrace{\int_{\mathbb{R}} f_{X,Y}(x,y)\,dy}_{\cancel{f_X(x)}} = \boxed{1}$$

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

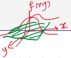
$F(x) = P(X \leq x)$

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

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



Since

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

we may equivalently write

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

$F_X(x) = \mathbb{P}(X \le x)$

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

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

Geometrically, $\{X \le x\} \cap \{Y \le y\} \equiv \{X \in (-\infty, x]\} \cap \{Y \in (-\infty, y]\}$

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

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

$$\mathbb{R}^n \ni \mathbf{X} = (X_1, \dots, X_n)$$ a n-dim. random vector

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

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

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.

$X_1, \dots, X_n$

Basic Properties of a Joint CDF

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

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

1. **Bounds:**

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

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

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

   If $y_1 \le y_2$, then

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

$\{Y \le y\} \cap \{X \le x_1\} \subseteq \{X \le x_2\} \cap \{Y \le y\}$

$P[\{X \le x_1, Y \le y\}] \le P[\{X \le x_2, Y \le y\}]$

3. **Right-continuity:**
   If

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

   then

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

$F_X(x)$

$f: A \to B$

$x_n \downarrow x \Rightarrow f(x_n) \downarrow f(x)$

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

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

$\phi = \underbrace{\{X < -\infty\}}_{\phi} \cap \{\dots\}$

for every fixed y, and

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

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

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

$\underbrace{\{X < \infty\}}_{S} \cap \underbrace{\{Y < \infty\}}_{S}$

Indeed,

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

$= S \cap S = S$

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

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

$P(S) = 1$

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

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

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

$\downarrow ?$

$P_X(x)$

$P_X(x) = \sum_y P_{X,Y}(x,y)$

For example,

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

$\lim P(X \le x, \underbrace{Y < \infty}_{S})$

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 $+\infty$:

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

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

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

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

important

(HW)

Probabilities of Rectangles from the Joint CDF

Let

$$a < b \quad\text{and}\quad c < d.$$

Then

$$\boxed{\begin{aligned}&\mathbb{P}(a < X \le b,\ c < Y \le d)\\ &\quad = F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c).\end{aligned}}$$

$$\boxed{p(a < X \le b) = F(b) - F(a)}$$

This is the two-dimensional inclusion–exclusion formula.

Since probabilities are nonnegative,

$$\boxed{F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c) \ge 0.}$$

Two-increasing property

Every valid bivariate CDF must assign a nonnegative probability to every rectangle.

$\mathrm{Bin}(n,p)$ Poisson(λ)

conditions:

$n \to \infty$ $\quad n \times p_n \to \lambda > 0$

$p_n \to 0$

$\binom{n}{k} p_n^k (1-p_n)^{n-k} \to \dfrac{e^{-\lambda}\lambda^k}{k!}$

$n \to \infty$

$p_n \to 0$

$n \cdot p_n \to \lambda$

$\mathrm{Bin}\left(n, \frac{1}{n}\right)$

$p = \frac{1}{n}$

$n \to \infty$

$1/n \to 0$

$\frac{1}{n} \times n \to 1$ $\quad \lambda = 1$

$\mathrm{Bin}\left(n, \frac{1}{\sqrt{n}}\right)$

$n \to \infty$

$1/\sqrt{n} \to$

Deriving the Rectangle Formula

Consider the rectangle

$$(a,b] \times (c,d].$$

Start with the larger southwest rectangle:

$$\mathbb{P}(X \le b,\ Y \le d) = F_{X,Y}(b,d).$$

Subtract the two unwanted regions:

$$F_{X,Y}(a,d) \quad\text{and}\quad F_{X,Y}(b,c).$$

The region

$$\{X \le a,\ Y \le c\}$$

was subtracted twice, so it must be added back once.
Therefore,

$$\begin{aligned}\mathbb{P}(a < X \le b,\ c < Y \le d) &= F_{X,Y}(b,d) - F_{X,Y}(a,d)\\ &\quad - F_{X,Y}(b,c) + F_{X,Y}(a,c).\end{aligned}$$

This shows that the joint CDF determines probabilities of all rectangles, and hence determines the entire joint distribution.

Joint CDFs Versus Marginal CDFs

The joint CDF always determines the marginal CDFs:

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

and

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

However, knowing only F_X and F_Y does not generally determine $F_{X,Y}$.

Important

Different joint distributions can have exactly the same marginal distributions but different dependence structures.

If X and Y are independent, then the marginals do determine the joint CDF:

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

Indeed,

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

X, Y, Z

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

Properties of an n-Dimensional Joint CDF

(skip)

at home

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 $$\boxed{F_{X_1,\dots,X_n}(x_1,\dots,x_n) \longrightarrow 0.}$$
5. If every coordinate tends to $+\infty$, then $$F_{X_1,\dots,X_n}(\overset{\infty}{x_1},\dots,\overset{\infty}{x_n}) \longrightarrow 1.$$
6. It assigns nonnegative probability to every n-dimensional rectangle.

~~The final property is called the~~

~~n-increasing property.~~

$$F_{X_1,\dots,X_n}(x_1,\dots,x_n) = \mathbb{P}(\underset{-\infty}{X_1 \le x_1}, X_2 \le x_2, \dots, X_n \le x_n)$$

ϕ

ϕ $= 0$

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

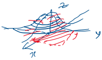
$$\boxed{F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du.}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dX$$

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

Thus the joint CDF accumulates the density over the southwest rectangle

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



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

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

Connection

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

Example: Uniform on the Unit Square

$A, B \in \mathcal{F}$ are indep.

$P(A \cap B) = P(A)P(B)$

Suppose

$$(X,Y) \sim \mathrm{Unif}([0,1]^2).$$

Then:

$$f_{X,Y}(x,y) = 1, \quad 0 < x < 1,\ 0 < y < 1.$$

Marginals:

$$f_X(x) = 1, \qquad f_Y(y) = 1.$$

Thus:

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

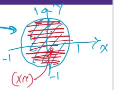
$$\forall x \in (0,1):\ f_X(x) = \int_0^1 f_{X,Y}(x,y)\,dy = \int_0^1 1\,dy = y\Big]_0^1 = 1 - 0 = 1$$

$$f_Y(y) = \int_0^1 f_{X,Y}(x,y)\,dx = \int_0^1 1\,dx = x\Big]_0^1 = 1 - 0 = 1$$

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

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

Indeed,

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

$$\iint_D 1\,dx\,dy = \mathrm{Area}(D) = \pi$$

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

$$-\sqrt{1-x^2} \le Y \le \sqrt{1-x^2}. \quad \Leftarrow x^2 + y^2 \le 1$$

Important

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

$P(A|B)$

$= P(A)$

$$-\sqrt{1-(0.3)^2} < Y < +\sqrt{1-(0.3)^2}$$

Independence

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

Theorem

The following are equivalent for random variables X and Y:

1. X and Y are independent.
2. $$F_{X,Y}(x,y) = F_X(x)F_Y(y).$$
3. If discrete:
   $$p_{X,Y}(x,y) = p_X(x)p_Y(y).$$
4. If continuous:
   $$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$
5. $$f_{Y|X}(y|x) = f_Y(y).$$
6. $$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$
7. If MGFs exist:
   $$M_{X,Y}(s,t) = M_X(s)M_Y(t).$$

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.

Independence of Several Random Variables

Definition

Random variables

$$X_1, \dots, X_n$$

are mutually independent if for every subset

$$i_1, \dots, i_k,$$

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

Pairwise Independence

Definition

Random variables

$$X_1, \dots, X_n$$

are pairwise independent if every pair

$$(X_i, X_j)$$

is independent.

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

But the converse is false.

Pairwise Independent But Not Mutually Independent

Let

$$X, Y$$

be independent Bernoulli$(1/2)$ random variables.
Define

$$Z = X \oplus Y,$$

where $\oplus$ denotes XOR.

Then:

- every pair is independent;
- but

  $$Z$$

  is completely determined by

  $$X, Y.$$

Hence:

$$X, Y, Z$$

are not mutually independent.

2D LOTUS

2D LOTUS

Theorem

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

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

Meaning

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

Linearity of Expectation Revisited

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

Covariance

Definition

The covariance of X, Y is

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

Equivalent formula:

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

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

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.

Correlation

Definition

The correlation coefficient is

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

Properties:

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

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

Variance of Sums

Theorem

For random variables X, Y,

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

If X, Y are independent:

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

Independence Implies Zero Correlation

Theorem

If X, Y are independent, then

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

Hence:

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

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:

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

But:

$$Y$$

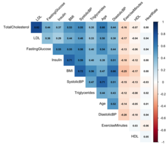
is completely determined by

$$X.$$

Thus X, Y are dependent.

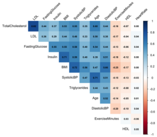
Heatmaps and Correlation Visualization

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



	LDL	FastingGlucose	Insulin	BMI	SystolicBP	Triglycerides	Age	DiastolicBP	ExerciseMinutes	HDL	HeartRate
TotalCholesterol	0.82	0.44	0.37	0.53	0.55	0.35	0.60	0.44	-0.16	-0.07	0.00
LDL		0.36	0.29	0.44	0.48	0.33	0.50	0.38	-0.17	-0.04	0.04
FastingGlucose			0.55	0.55	0.58	0.40	0.54	0.44	-0.13	-0.06	0.04
Insulin				0.71	0.59	0.40	0.39	0.51	-0.18	-0.12	0.00
BMI					0.72	0.56	0.47	0.66	-0.25	-0.17	0.00
SystolicBP						0.47	0.71	0.51	-0.15	-0.13	-0.03
Triglycerides							0.44	0.43	-0.12	-0.08	0.02
Age								0.52	-0.14	-0.05	0.01
DiastolicBP									-0.26	-0.10	0.04
ExerciseMinutes										0.03	-0.06
HDL											0.05

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

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

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

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

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

$$E(X_i) = np_i,$$

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

and for

$$i \neq j,$$

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

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,

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

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.

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

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

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

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

- The covariance is

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

- Therefore,

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

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,

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

does not necessarily imply independence.

But for the bivariate normal distribution, it does.

Important Gaussian Property

If (X, Y) is bivariate normal, then

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

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

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

So in the bivariate normal case,

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