STAT 131 – Lecture Review Week 8

Functions of random variables, order statistics, and expected value

Antonio Aguirre

Spring 2026

Based on Lectures 7.1 and 7.2, with final-style examples.

How to use this review

Goal of this review

This deck is a study guide for three recurring tasks: **transform a random variable**, **find a distribution of a sum/max/min**, and **compute an expectation efficiently**.

Before choosing a formula, identify the task

1. What is the new random variable? Examples: $Y = r(X)$, $Y = X_1 + X_2$, $M = \max_i X_i$, $L = \min_i X_i$.
2. What is known about the original variable(s): density, CDF, support, independence?
3. Do we need the **full distribution** of the new variable, or only an **expected value**?
4. What is the support of the new variable? Write it before doing calculus.

Main habit

State the random variable, state its support, then choose the workflow.

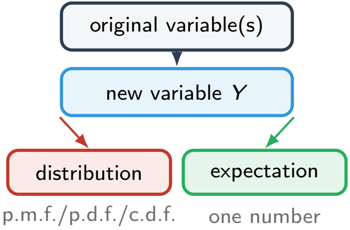
Three workflows for this week

Match the problem to the workflow

Task	Main move
$Y = r(X)$, discrete	group all x values with the same output y
$Y = r(X)$, continuous	f_X, $\mathrm{Supp}(X) \to F_X \to \mathrm{Supp}(Y) \to F_Y \to f_Y$
$Y = X_1 + X_2$	use the line $x_1 + x_2 = y$ and integrate over the valid segment
$M = \max_i X_i$	use $\{M \le y\} = \{X_1 \le y, \ldots, X_n \le y\}$
$L = \min_i X_i$	use the complement: $\{L > y\} = \{X_1 > y, \ldots, X_n > y\}$

Efficiency principle

If the question asks only for $\mathbb{E}[r(X)]$, use LOTUS. Do not find the distribution of $r(X)$ unless needed.



Study cue

Most problems are not about memorizing a formula; they are about recognizing the right workflow.

Discrete functions: probability mass can merge

Recipe

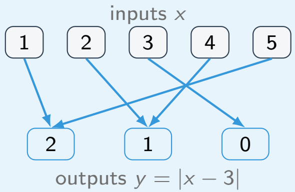
If X is discrete and $Y = r(X)$, then

$$f_Y(y) = \Pr(Y = y) = \sum_{x:r(x)=y} f_X(x).$$

1. List the possible x values.
2. Compute the output $y = r(x)$ for each input.
3. Add probabilities of inputs with the same output.

Intuition: many-to-one maps merge probability mass.

$X \sim \mathrm{Unif}\{1, 2, 3, 4, 5\}$ and $Y = |X - 3|$



$$\Pr(Y = 0) = \frac{1}{5},\ \Pr(Y = 1) = \Pr(Y = 2) = \frac{2}{5}.$$

Continuous transformations: the CDF recipe

Use this when $Y = r(X)$ and X has density f_X

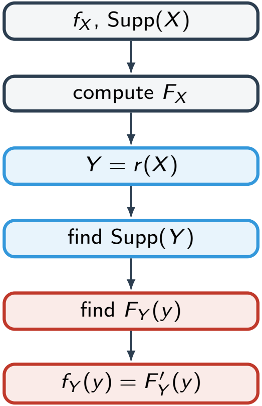
1. **Start with X.** Find $\mathrm{Supp}(X)$ and compute

$$F_X(x) = \Pr(X \le x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t.$$

2. **Transform the support.** Find $\mathrm{Supp}(Y)$ from the range of $r(x)$ on $\mathrm{Supp}(X)$.
3. **Write the CDF of Y.**

$$F_Y(y) = \Pr(Y \le y) = \Pr(r(X) \le y).$$

4. **Translate back to X** and express the event using F_X.
5. **Differentiate:** $f_Y(y) = F_Y'(y)$ on $\mathrm{Supp}(Y)$.



Important: this works for one-to-one and many-to-one maps.

CDF recipe example: $X \sim \mathrm{Unif}(-1,1)$ and $Y = X^2$

Step 1: compute the original CDF

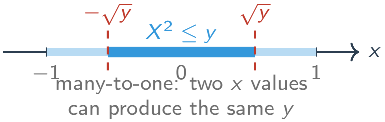
$$\mathrm{Supp}(X) = (-1,1), \;\; f_X(x) = \frac{1}{2}.$$

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

Step 2: transform the support

Since $Y = X^2$ and $-1 < X < 1$,

$$\mathrm{Supp}(Y) = (0,1).$$



Important

The shortcut theorem does not apply directly on $(-1,1)$ because $x \mapsto x^2$ is not one-to-one.

CDF recipe example: compute F_Y and differentiate

Steps 3–5

For $0 < y < 1$,

$$F_Y(y) = \Pr(Y \le y) = \Pr(X^2 \le y).$$

Translate the event:

$$X^2 \le y \iff -\sqrt{y} \le X \le \sqrt{y}.$$

Then use the CDF of X:

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$= \frac{\sqrt{y}+1}{2} - \frac{-\sqrt{y}+1}{2} = \sqrt{y}.$$

Differentiate:

$$f_Y(y) = F_Y'(y) = \frac{1}{2\sqrt{y}}, \; 0 < y < 1.$$

Final answer

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

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

Interpretation: squaring compresses values toward 0, so the density is larger near $y = 0$.

Shortcut theorem: use it only when the map is one-to-one

One-to-one transformation theorem

If $Y = r(X)$, r is one-to-one on $\text{Supp}(X)$, and

$$s(y) = r^{-1}(y),$$

then

$$f_Y(y) = f_X(s(y))\,|s'(y)|, \; y \in \text{Supp}(Y).$$

Shortcut checklist

1. Check r is one-to-one on $\text{Supp}(X)$.
2. Find $s(y) = r^{-1}(y)$.
3. Find $\text{Supp}(Y)$.
4. Multiply by the Jacobian factor $|s'(y)|$.

$X \sim \text{Unif}(0,1)$ and $Y = 10e^{5X}$

Here $r(x) = 10e^{5x}$ is increasing on $(0,1)$, so it is one-to-one.

$$s(y) = \frac{1}{5}\log\left(\frac{y}{10}\right), \quad |s'(y)| = \frac{1}{5y}.$$

$$\text{Supp}(Y) = (10, 10e^5).$$

Since $f_X(x) = 1$ on $(0,1)$,

$$f_Y(y) = \frac{1}{5y}, \; 10 < y < 10e^5.$$

The CDF method would give the same answer, but the theorem is shorter when its assumptions hold.

Final-style example: marginal first, then transform

Start from the joint density

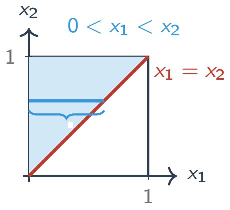
$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 2, & 0 < x_1 < x_2 < 1, \\ 0, & \text{otherwise.} \end{cases}$$

To transform X_2, first find the marginal density of X_2. For $0 < x_2 < 1$,

$$f_{X_2}(x_2) = \int_0^{x_2} 2\,\mathrm{d}x_1 = 2x_2.$$

Why this step matters

The transformation is applied to X_2, not to the full pair (X_1, X_2).



Final-style example: apply the recipe to $Y = X_2^2$

After marginalizing, the original variable is X_2 with

$$f_{X_2}(x) = 2x,\ 0 < x < 1.$$

CDF recipe

Step 1. Compute the original CDF:

$$F_{X_2}(x) = \int_0^x 2t\,\mathrm{d}t = x^2,\ 0 < x < 1.$$

Step 2. Transform the support:

$$Y = X_2^2,\ 0 < X_2 < 1 \Rightarrow \mathrm{Supp}(Y) = (0,1).$$

Steps 3–5.

$$F_Y(y) = \Pr(X_2^2 \le y) = F_{X_2}(\sqrt{y}) = y.$$

Thus $f_Y(y) = 1$ on $(0,1)$.

Shortcut check

On $\mathrm{Supp}(X_2) = (0,1)$, $r(x) = x^2$ is one-to-one.

$$s(y) = \sqrt{y},\ |s'(y)| = \frac{1}{2\sqrt{y}}.$$

$$f_Y(y) = f_{X_2}(\sqrt{y})\frac{1}{2\sqrt{y}} = 1.$$

$$\boxed{Y \sim \mathrm{Unif}(0,1)}$$

Support decides the tool: x^2 is one-to-one on $(0,1)$ but not on $(-1,1)$.

Sums of independent continuous variables: convolution

General form

If $Y = X_1 + X_2$ and $X_1 \perp\!\!\!\perp X_2$, then

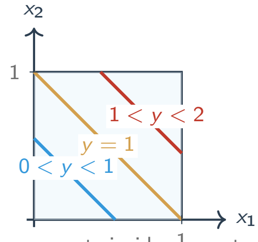
$$f_Y(y) = \int_{-\infty}^{\infty} f_1(x) f_2(y - x)\,\mathrm{d}x.$$

The integrand is nonzero only when both

$$x \in \operatorname{Supp}(X_1), \;\; y - x \in \operatorname{Supp}(X_2).$$

How to find the limits

Draw the line $x_1 + x_2 = y$. The bounds are the part of that line inside the joint support.



line segments inside support

Example: sum of two independent Uniform$(0,1)$ variables

Let $X_1, X_2 \overset{\text{iid}}{\sim} \mathrm{Unif}(0,1)$ and $Y = X_1 + X_2$.

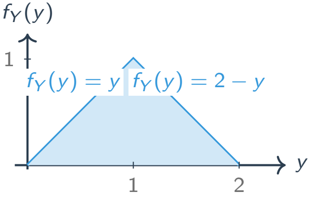
The limits change at $y = 1$

For $0 < y < 1$, the line segment runs from $x = 0$ to $x = y$:

$$f_Y(y) = \int_0^y 1\,\mathrm{d}x = y.$$

For $1 \le y < 2$, it runs from $x = y - 1$ to $x = 1$:

$$f_Y(y) = \int_{y-1}^1 1\,\mathrm{d}x = 2 - y.$$



Density

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

Order statistics: maximum and minimum of an i.i.d. sample

Let $X_1, \ldots, X_n$ be i.i.d. with CDF F and density f.

$$M_n = \max\{X_1, \ldots, X_n\}, \; L_n = \min\{X_1, \ldots, X_n\}.$$

Maximum: all observations below y

$$\{M_n \le y\} = \{X_1 \le y, \ldots, X_n \le y\}.$$

By independence,

$$F_M(y) = \Pr(M_n \le y) = [F(y)]^n.$$

If F is differentiable,

$$f_M(y) = n[F(y)]^{n-1} f(y).$$

Minimum: use the complement

$$\{L_n > y\} = \{X_1 > y, \ldots, X_n > y\}.$$

By independence,

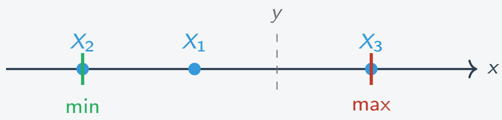
$$\Pr(L_n > y) = [1 - F(y)]^n.$$

Thus

$$F_L(y) = 1 - [1 - F(y)]^n,$$
$$f_L(y) = n[1 - F(y)]^{n-1} f(y).$$

Visual cue for max and min

Example with three observations



Translate events first

$$M_3 \le y \iff X_1 \le y, X_2 \le y, X_3 \le y.$$

This is an intersection, so independence lets probabilities multiply.

$$L_3 > y \iff X_1 > y, X_2 > y, X_3 > y.$$

Then convert back to $F_L(y) = \Pr(L_3 \le y)$ using a complement.

Common error

For max/min, do not add probabilities. Translate to an event involving all observations.

Example: max/min of three Uniform$(0, 1)$ variables

Let $X_1, X_2, X_3 \overset{\text{iid}}{\sim} \mathrm{Unif}(0, 1)$, so

$$F(y) = y,\ f(y) = 1,\ 0 < y < 1.$$

Maximum $M = \max X_i$

$$F_M(y) = [F(y)]^3 = y^3.$$

$$f_M(y) = 3[F(y)]^2 f(y) = 3y^2,\ 0 < y < 1.$$

The maximum is more likely to be near 1.

Minimum $L = \min X_i$

$$F_L(y) = 1 - [1 - F(y)]^3 = 1 - (1 - y)^3.$$

$$f_L(y) = 3[1 - F(y)]^2 f(y) = 3(1 - y)^2,\ 0 < y < 1.$$

The minimum is more likely to be near 0.

Final-style example: max of two normal bottle fills

Suppose two independently selected bottles have fill amounts

$$X_1, X_2 \stackrel{\text{iid}}{\sim} N(16, 0.09), \ \sigma = 0.3.$$

Let $Y_2 = \max\{X_1, X_2\}$.

Use the maximum recipe

$$G_{Y_2}(y) = \Pr(Y_2 \le y)$$
$$= \Pr(X_1 \le y, X_2 \le y).$$

Because X_1 and X_2 are independent,

$$G_{Y_2}(y) = \Pr(X_1 \le y)\Pr(X_2 \le y).$$

Use the standard normal CDF

$$\Pr(X_i \le y) = \Phi\left(\frac{y-16}{0.3}\right).$$

Thus

$$G_{Y_2}(y) = \left[\Phi\left(\frac{y-16}{0.3}\right)\right]^2.$$

Exam cue: for maxima, start with $\Pr(\max \le y)$.

Expected value: probability-weighted average

Definitions

If X is discrete,

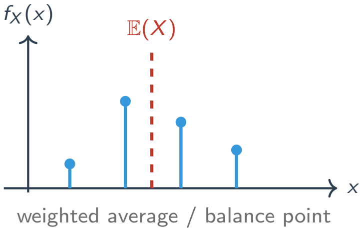
$$\mathbb{E}(X) = \sum_x x f_X(x).$$

If X is continuous,

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x.$$

Interpretation

$\mathbb{E}(X)$ is a number: the long-run average or balance point of the distribution.



Expectation of a function: use LOTUS

Law of the unconscious statistician

If $Y = r(X)$, compute the expected value of Y directly from the distribution of X:

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

Discrete case:

$$\mathbb{E}[r(X)] = \sum_x r(x) f_X(x).$$

Continuous case:

$$\mathbb{E}[r(X)] = \int_{-\infty}^{\infty} r(x) f_X(x)\,\mathrm{d}x.$$

Example

If $Y = X^2$, then

$$\mathbb{E}(Y) = \mathbb{E}(X^2) = \int x^2 f_X(x)\,\mathrm{d}x.$$

Main advantage

You often do not need to find the p.d.f. of Y to compute $\mathbb{E}(Y)$.

Linearity of expectation

Most useful property

For constants a, b,

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b.$$

For random variables $X_1, \ldots, X_n$,

$$\mathbb{E}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i).$$

No independence needed

Linearity works even when the variables are dependent.

Bernoulli sum

If $X_i \sim \mathrm{Bern}(p)$ and $Y = \sum_{i=1}^{n} X_i$, then

$$\mathbb{E}(Y) = \sum_{i=1}^{n} \mathbb{E}(X_i) = np.$$

Product rule for expectations: independence matters

Correct statement

If $X_1, \ldots, X_n$ are independent and have finite expectations, then

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i).$$

For two variables,

$$X \perp\!\!\!\perp Y \implies \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

True

Independence lets products factor.

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

False in general

Knowing only $\mathbb{E}(X) = 1$ and $\mathbb{E}(Y) = 2$ does not imply $\mathbb{E}(XY) = 2$.

Final-style example: normal linear combinations

Let

$$X_1, X_2 \overset{\text{ind}}{\sim} N(16, 0.09).$$

Define

$$Y = X_1 - X_2, \;\; W = X_1 + X_2.$$

Difference

$$\mathbb{E}(Y) = 16 - 16 = 0,$$

$$\mathrm{Var}(Y) = 0.09 + 0.09 = 0.18.$$

So

$$Y \sim N(0, 0.18).$$

Sum

$$\mathbb{E}(W) = 16 + 16 = 32,$$

$$\mathrm{Var}(W) = 0.09 + 0.09 = 0.18.$$

So

$$W \sim N(32, 0.18).$$

Reason

Independent normal variables stay normal under linear combinations.

Expected values of common distributions

Know what X counts before using the formula

Distribution	**Meaning of X**	$\mathbb{E}(X)$
$\mathrm{Bern}(p)$	one success/failure indicator	p
$\mathrm{Binom}(n, p)$	successes in n independent trials	np
$\mathrm{Poisson}(\lambda)$	count in a fixed interval	λ
$\mathrm{Geom}(p)$	failures before first success	$\frac{1-p}{p}$
$\mathrm{NegBin}(r, p)$	failures before r successes	$\frac{r(1-p)}{p}$
$\mathrm{Hypergeom}(n, A, B)$	red items in sample of size n	$\frac{nA}{A+B}$

Convention warning

Here geometric counts failures before the first success, so $X = 0$ is possible.

Final-style example: Poisson counts and linearity

Customers arrive at a coffee shop according to a Poisson model. Let

$$X = \text{number of customers in 10 minutes}, \ X \sim \mathrm{Poisson}(3).$$

At least one customer

$$\Pr(X \geq 1) = 1 - \Pr(X = 0) = 1 - e^{-3}.$$

Expected value

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

If $Y = 2X + 1$, then

$$\mathbb{E}(Y) = 2\mathbb{E}(X) + 1 = 7.$$

Change the time window

One hour has six 10-minute intervals. If the rate is stable,

$$W = \text{customers in 60 minutes}$$

has

$$W \sim \mathrm{Poisson}(6 \cdot 3) = \mathrm{Poisson}(18).$$

Common mistakes to avoid

Transformations

- Forgetting the support of the new variable.
- Using the one-to-one formula when the map is not one-to-one.
- Treating a density value as a probability.
- Missing one branch of a many-to-one transformation.

Expectations

- Finding the whole distribution when LOTUS is enough.
- Thinking linearity requires independence.
- Factoring $\mathbb{E}(XY)$ without checking independence.
- Confusing $\mathbb{E}(X)$ with a random variable.

Reliable habit

State the random variable, state the support, then choose the tool.

Final checklist

For distributions of new random variables

- Discrete $Y = r(X)$: group inputs with the same output and add their probabilities.
- Continuous $Y = r(X)$: use the CDF recipe $f_X, \text{Supp}(X) \to F_X \to \text{Supp}(Y) \to F_Y \to f_Y$.
- One-to-one shortcut: use $f_Y(y) = f_X(s(y))|s'(y)|$ only after checking the support.
- Sum $Y = X_1 + X_2$: identify the valid segment of $x_1 + x_2 = y$ inside the support.
- Max/min of i.i.d. sample: use $F_M(y) = [F(y)]^n$ and $F_L(y) = 1 - [1 - F(y)]^n$.

For expected values

- Use LOTUS for $\mathbb{E}[r(X)]$ when the full distribution of $r(X)$ is not needed.
- Use linearity freely: independence is not required.
- Factor $\mathbb{E}(XY)$ only when independence has been given or proved.

STAT 131 – Lecture Review Week 9

Variance, covariance, conditional expectation, and a first normal-table workflow

Antonio Aguirre

Spring 2026

Based on Lectures 8.1 and 8.2, with final-style examples.

How to use this review

Goal of this review

This deck is a study guide for four tasks: **measure spread**, **measure linear association**, **average after conditioning**, and **use the standard normal table**.

Match the question to the tool

Question	**Tool**
Spread of one variable	$\mathrm{Var}(X)$, $\mathrm{SD}(X)$
Linear movement of two variables	$\mathrm{Cov}(X, Y)$, $\rho(X, Y)$
Average after information	$\mathbb{E}(Y \mid X = x)$, $\mathbb{E}(Y \mid X)$
Combine conditional pieces	total expectation / total variance
Normal-table probability	standardize and use Φ

Main habit

Write the random variable first, then choose the formula.

Check before using shortcuts

- $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ needs independence.
- Variances of sums may need covariance terms.
- $\mathbb{E}(Y \mid X)$ is random before X is observed.

Variance: average squared distance from the mean

Definition

If $\mu = \mathbb{E}(X)$ is finite, then

$$\mathrm{Var}(X) = \mathbb{E}\big[(X-\mu)^2\big], \qquad \mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}.$$

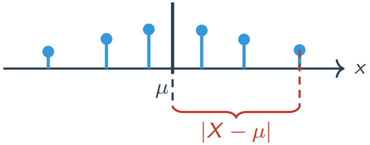
Equivalent computation

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

This is usually the faster way to compute variance.

Interpretation

Variance measures spread around the mean. Squaring makes distances nonnegative and gives more weight to large deviations.



variance averages squared distances from the mean

Reading cue

Same mean does not imply same spread.

Computing variance without extra work

Two equivalent forms

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2]$$
$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

Common mistake

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

not $a\mathrm{Var}(X) + b$.

Discrete vs. continuous

Case	**Second moment**
Discrete	$\mathbb{E}(X^2) = \sum_x x^2 f(x)$
Continuous	$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x$

Workflow

Find $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$, then subtract:

$$\boxed{\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.}$$

Variance of sums: where independence matters

General two-variable rule

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

If X and Y are independent

Independence implies $\mathrm{Cov}(X, Y) = 0$, so

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

For independent $X_1, \ldots, X_n$,

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i).$$

Known variances: use structure, not long sums

Binomial as a sum of Bernoulli variables

If

$$X \sim \mathrm{Binomial}(n, p), \qquad X = Y_1 + \cdots + Y_n,$$

where $Y_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bernoulli}(p)$, then

$$\mathbb{E}(X) = np, \qquad \mathrm{Var}(X) = np(1-p).$$

Poisson

If

$$X \sim \mathrm{Poisson}(\lambda),$$

then

$$\mathbb{E}(X) = \lambda, \qquad \mathrm{Var}(X) = \lambda.$$

Why

Independent Bernoulli variances add.

Interpretation

For a Poisson count, the mean and variance are the same number.

Exam habit

If the distribution is named, use its known mean and variance unless the problem asks you to derive them.

Final-style example: coffee shop arrivals

Setup

Customers arrive at a coffee shop according to

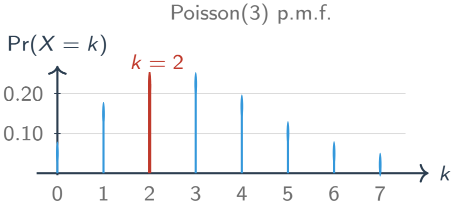
$$X \sim \mathrm{Poisson}(3),$$

where X is the number of arrivals in a 10-minute period.

Use the p.m.f.

$$\Pr(X = k) = e^{-3}\frac{3^k}{k!}.$$

$$\Pr(X = 2) = e^{-3}\frac{3^2}{2!} \approx 0.224.$$



At least one arrival

$$\Pr(X \geq 1) = 1 - \Pr(X = 0) = 1 - e^{-3} \approx 0.950.$$

Same example: transform and rescale the count

Mean and variance

Since $X \sim \text{Poisson}(3)$,

$$\mathbb{E}(X) = 3, \qquad \text{Var}(X) = 3.$$

If

$$Y = 2X + 1,$$

then

$$\mathbb{E}(Y) = 2\mathbb{E}(X) + 1 = 7,$$

$$\text{Var}(Y) = 2^2\text{Var}(X) = 12.$$

One-hour count

One hour has six 10-minute periods. If the rate is constant,

$$W \sim \text{Poisson}(6 \cdot 3) = \text{Poisson}(18).$$

$$\mathbb{E}(W) = 18, \qquad \text{Var}(W) = 18.$$

Reading the problem

Changing the time window changes the Poisson mean.

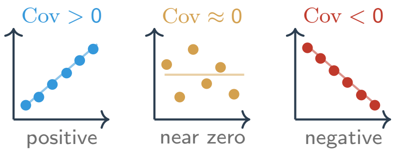
Covariance: one number for joint movement

Definition

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

Sign

- $\mathrm{Cov}(X, Y) > 0$: large values tend to occur together.
- $\mathrm{Cov}(X, Y) < 0$: one tends to be large when the other is small.
- $\mathrm{Cov}(X, Y) = 0$: no linear association detected.



Caution

Covariance depends on units. Correlation standardizes it.

Correlation: standardized covariance

Definition

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

Range

By Cauchy–Schwarz,

$$-1 \le \rho(X,Y) \le 1.$$

Meaning

Value	Interpretation
$\rho \approx 1$	strong positive linear association
$\rho \approx 0$	weak/no linear association
$\rho \approx -1$	strong negative linear association

Useful facts

If $X \perp\!\!\!\perp Y$, then $\rho(X,Y) = 0$ if the variances are positive. If $Y = aX + b$, then $\rho(X,Y) = 1$ for $a > 0$ and -1 for $a < 0$.

Variance of a linear combination

General two-variable formula

If X and Y have finite variances and covariance, then

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

If independent

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

so

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

Exam warning

Do not erase the covariance term unless independence or zero covariance is given or proven.

Special case

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

Final-style trap: when can we multiply expectations?

Claim

Suppose $\mathbb{E}(X) = 1$ and $\mathbb{E}(Y) = 2$. Then

$$\mathbb{E}(XY) = 2.$$

Answer

False. The marginal means alone do not determine the joint mean $\mathbb{E}(XY)$.

Safe statement

If $X \perp\!\!\!\perp Y$, then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Without independence, this need not hold.

Concrete counterexample

Let X take values 0 and 2 with probability 1/2 each, and let $Y = 2X$.

	X	$Y = 2X$
case 1	0	0
case 2	2	4

$$\mathbb{E}(X) = 1, \qquad \mathbb{E}(Y) = 2,$$

$$\mathbb{E}(XY) = \mathbb{E}(2X^2) = 4 \neq 2.$$

Lesson

Dependence lives in the joint distribution, not in the two marginal means.

Continuous covariance: organize the integrals

Example from the notes

Let

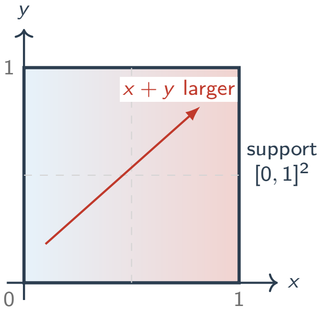
$$f(x,y) = \begin{cases} x+y, & 0 \le x \le 1,\ 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

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

Plan

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

Compute all three expectations over the same support.



Result

$$\mathbb{E}(X) = \mathbb{E}(Y) = \frac{7}{12}, \qquad \mathbb{E}(XY) = \frac{1}{3},$$

$$\mathrm{Cov}(X,Y) = \frac{1}{3} - \left(\frac{7}{12}\right)^2 = -\frac{1}{144}.$$

Conditional expectation: average after information

Definition

For a conditional distribution of Y given $X = x$,

$$\mathbb{E}(Y \mid X = x) = \sum_y y \, p_{Y|X}(y \mid x)$$

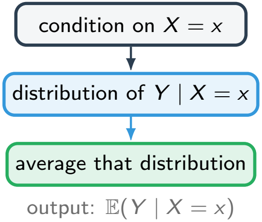
for discrete Y, and

$$\mathbb{E}(Y \mid X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x) \, \mathrm{d}y$$

for continuous Y.

Notation

$\mathbb{E}(Y \mid X = x)$ is a number once x is fixed. $\mathbb{E}(Y \mid X)$ is a random variable because it changes with X.



Interpretation

Conditioning changes the reference population before averaging.

A model with a random probability

Setup

Let P be an unknown disease probability, and assume

$$P \sim \mathrm{Uniform}(0,1).$$

Given $P = p$, let

$$Y \mid P = p \sim \mathrm{Binomial}(n, p),$$

where Y is the number of diseased patients in a sample of size n.

Conditional expectation

$$\mathbb{E}(Y \mid P = p) = np, \qquad \mathbb{E}(Y \mid P) = nP.$$

P — random risk

$\downarrow$

$Y \mid P \sim \mathrm{Bin}(n, P)$ — sampling variation

$\downarrow$

$\mathbb{E}(Y \mid P) = nP$ — conditional mean

Key distinction

np is fixed after p is known. nP is random before P is known.

Law of total expectation

Formula

$$\boxed{\mathbb{E}(Y) = \mathbb{E}\big[\mathbb{E}(Y \mid X)\big].}$$

Apply to the random-probability model

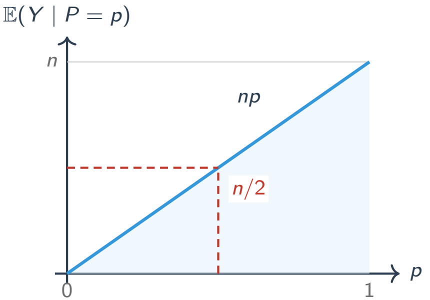
$$\mathbb{E}(Y \mid P) = nP.$$

Therefore,

$$\mathbb{E}(Y) = \mathbb{E}\big[\mathbb{E}(Y \mid P)\big] = \mathbb{E}(nP) = n\mathbb{E}(P).$$

Since $P \sim \mathrm{Uniform}(0,1)$,

$$\mathbb{E}(P) = \frac{1}{2}, \qquad \boxed{\mathbb{E}(Y) = \frac{n}{2}.}$$



average the conditional mean over $P \sim \mathrm{Unif}(0,1)$

Law of total variance: what it separates

Formula

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

Within-condition variation

After X is fixed, Y may still vary.

$$\mathbb{E}\big[\mathrm{Var}(Y \mid X)\big]$$

averages the remaining variability inside each condition.

Between-condition variation

The conditional mean may change as X changes.

$$\mathrm{Var}\big(\mathbb{E}(Y \mid X)\big)$$

measures variability among the conditional averages.

Law of total variance: patient model

Recall the model

$$P \sim \mathrm{Uniform}(0,1), \qquad Y \mid P = p \sim \mathrm{Binomial}(n,p).$$

Conditional pieces

$$\mathrm{Var}(Y \mid P) = nP(1-P), \qquad \mathbb{E}(Y \mid P) = nP.$$

Therefore,

$$\mathrm{Var}(Y) = \mathbb{E}[nP(1-P)] + \mathrm{Var}(nP).$$

Use moments of P

For $P \sim \mathrm{Uniform}(0,1)$,

$$\mathbb{E}(P) = \frac{1}{2}, \quad \mathbb{E}(P^2) = \frac{1}{3}, \quad \mathrm{Var}(P) = \frac{1}{12}.$$

$$\boxed{\mathrm{Var}(Y) = \frac{n}{6} + \frac{n^2}{12}.}$$

Interpretation

The $n^2/12$ term appears because the disease probability itself is uncertain.

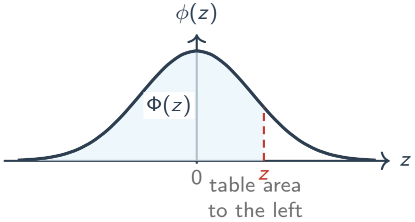
First normal-table workflow

Standard normal

If $Z \sim N(0,1)$, then

$$\Phi(z) = \Pr(Z \le z).$$

The normal table gives **left-tail probabilities**.



Core moves

For $z > 0$,

$$\Phi(-z) = 1 - \Phi(z).$$

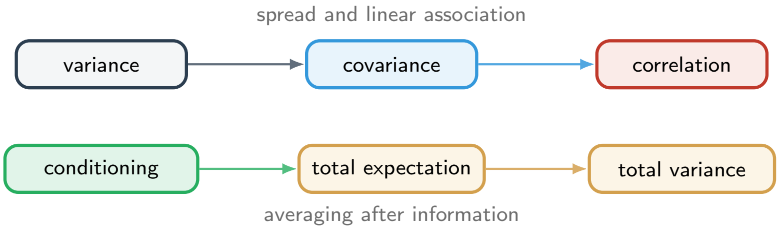
For $a < b$,

$$\Pr(a < Z < b) = \Phi(b) - \Phi(a).$$

Example

$$\Pr(Z \le -0.6) = 1 - \Phi(0.6) = 1 - 0.7257 = 0.2743.$$

How today's ideas connect



What each tool answers

Tool	Question it answers
Variance	How spread out is one variable?
Covariance	Do two variables move together or in opposite directions?
Correlation	How strong is the linear association on a unit-free scale?
Total expectation	What is the overall average after conditioning?
Total variance	How much variability comes from within conditions vs. between conditional means?

Checklist for practice problems

Before computing

1. Identify the random variable(s), support, and what is being requested.
2. Decide whether you need $\mathbb{E}(X)$, $\mathrm{Var}(X)$, $\mathrm{Cov}(X, Y)$, $\rho(X, Y)$, or a conditional quantity.
3. If variables are combined, check whether a covariance term is needed.
4. If the problem conditions on another variable, try total expectation or total variance.

Most common mistakes

- Treating $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ as always true.
- Forgetting the square in $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.
- Dropping covariance without independence or zero covariance.
- Confusing $\mathbb{E}(Y \mid X = x)$ with $\mathbb{E}(Y \mid X)$.