Cheatsheet

Contents

Probability

What's the point?

- Data science/AI/ML/… make predictions and conclusions
- Conclusions can be wrong
- The language of probability assigns **confidence** to our conclusions (not always correctly)
- Properly understanding probability is critical to living and working in the modern world, where confidence in data determines the decisions of people, companies and governments. You're *vulnerable to misinformation* if you don't understand it.

https://callingbullshit.org/

We will be describing lots of objects mathematically. We use standard notation, which is on the introductory cheatsheet. So have that handy!

Warmup

The Marimo notebooks on the *Learn probability* section of this link are good for getting to grips with the material. Choose them for learning, and the cheatsheet for reviewing.

Experiments

The notion of probability *relies* on the notion of an experiment. Whenever anybody says "this has probability 75%", they mean that if you repeated the experiment 100 times, *this* would be expected to happen on 75 outcomes.

It's often hard to figure out the underlying experiment. When you do, a probabilistic statement can sound a lot less convincing.

Uncertainty can depend on the observer. I don't know what's in your bag. You do.

An experiment is a process with *uncertain* outcomes.

Person	Process	Outcome
Life insurer	Health of 1000 people over the next 20 years	Health data, e.g. cancer, broken leg, death.
Climate modeller	Tomorrow's weather	Weather data, e.g. mm of rain, hours of sun

Probability space

This is the mathematical setting for experiments. It consists of **three objects**:

Ingredients of probability spaces	
Ω: The sample/ outcome space	A *set* containing all possible outcomes
$\mathcal{F}$: The event space	A *power set* (set of sets) containing all possible *sets* of outcomes
Probability function $\mathbb{P}$	$\mathbb{P} : \mathcal{F} \to [0, 1]$. IE $\mathbb{P}$ takes in events, and assigns them a *probability*: a number between 0 and 1 inclusive

Example

- I'm going to measure the weather tomorrow. Specifically, the millimetres of rainfall and hours of sunshine.
- Then the outcome space Ω is the set of possible values I could get. A given outcome, $\omega \in \Omega$ would be a tuple of two numbers representing these: e.g. $\omega = (2.4\text{mm}, 6.5 \text{ hours})$
- More than 6 hours of sunshine is an *event*. It's a set of all possible outcomes $\omega \in \Omega$ where the second value of ω is greater than 6. The event space $\mathcal{F}$ is the set of all possible events.
- We subjectively build a probability function that gives the probability of different events. If $\ell \in \mathcal{F}$ is the event that there are at least 8 hours of sunshine, then maybe $\mathbb{P}(\ell) = 0.2$ (i.e. 20%)

Event calculus

EVENT CALCULUS

Independent events	$A \in \mathcal{F}$ and $B \in \mathcal{F}$ satisfy $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
Mutually exclusive events	$A \in \mathcal{F}$ and $B \in \mathcal{F}$ satisfy $A \cap B = \emptyset$
Conditional probability	Take two events A and B. $\mathbb{P}[A \mid B]$ represents the probability of event A *given* that B happened. Note that $A \mid B$ isn't an event itself.

Random Variables

It's really important to properly understand what these mean

Random variables are **not** random. They are **not** variables.

RANDOM VARIABLES

Description (english)	**Description** (maths)
Quantitative questions about an experiment	Functions that map outcomes to numbers

Example:
- If you flip a coin, the outcome space is $\Omega = \{H, T\}$. So $\omega \in \Omega$ means $\omega = H$ or $\omega = T$.
- An example random variable is

$$X(\omega) = \begin{cases} 4 \text{ if } \omega = H \\ 2 \text{ if } \omega = T \end{cases}$$

Notice that the random variable itself is *not* random. The randomness is in the *outcome* of the experiment.

If we run the experiment and see what value the random variable takes, we say that we have *sampled* the random variable.

We will use the word 'sample' a lot. Remember it.

Random variable terminology

Support of a random variable: $\text{supp}(X)$	Set of plausible values it can take: $\text{supp}(X) = \{4, 2\}$ in the example above	
Continuous random variable	Support is an *infinite* set	e.g. a person's height can take infinite values
Discrete random variable	Support is a *finite* set	e.g. the example random variable above can only take two values

Probability mass/density functions (PMF/PDF)

RV type	Relevant function	Input	Output
Discrete	Probability *mass* function	$x \in \text{supp}(X)$ *single* value that RV can take	Probability of RV taking that value x
Continuous	Probability *density* function	$x \in \text{supp}(X)$ *single* value that RV can take	Probability *density* of RV taking that value x

- Let f_X be the probability mass function of a discrete random variable X. Then $f_{X(x)}$ is the probability of X taking the value x

Example:
- Recall the random variable from the previous example:

$X(\omega) = \begin{cases} 4 \text{ if } \omega = H \\ 2 \text{ if } \omega = T \end{cases}$

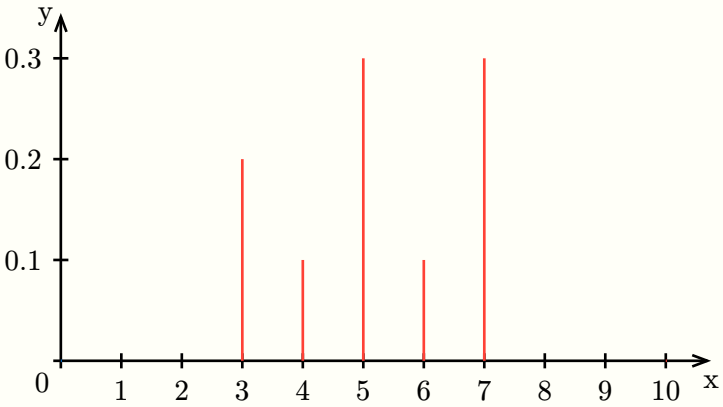
- The outcome space is finite (only two possible outcomes) so it's discrete and has a probability mass function $f_{X(x)}$.
- $f_X(4) = 0.5$ (50% probability it returns 4)
- $f_X(2) = 0.5$ (50% probability it returns 2)

Note that the PMF will change if the coin is biased…

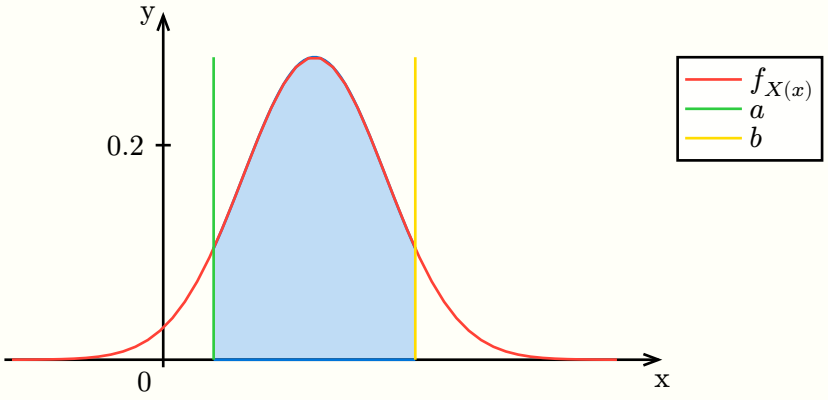
- Let f_X be the probability density function of a continuous random variable X. Then the probability of X taking any *individual* value x is always 0.

what's the probability a person is *exactly* 180.000… cm?

- Instead, let A be a *range* of values x can take. EG $A = [179, 181]$. Then the integral $\int_A f_{X(x)}dx$ is the probability that X takes a value falling in A



For a PMF the support of the RV is finite (there are 5 values it can take in this example). The height of each value is the probability of the RV taking that value. So heights sum to 1



For a PDF, the probability the random variable is between a and b is the shaded area. Mathematically, this is the integral $\int_a^b f_X(x)dx$. Take $b = a$ to see that the probability of a single value is 0. Total area under the curve is 1. But the height of the curve can be greater than 1

Flavours of random variable

Every experiment is different. But remember that random variables ask a quantitative question about the outcome of an experiment? Some quantitative questions have the same flavour, even if the experiment is different. Meaning what? They have the same PMF/PDF.

The PMF/PDF map the support of the random variable (values it can take) to numbers (probabilities/probability densities). Hence neither the inputs nor the outputs are experiment-specific quantities.

Bernoulli random variable:

Bernoullis are questions with binary (yes/no or 1/0) answers. So the support is $\{1, 0\}$.

use this information to write the PMF of a Bernoulli in mathematical notation.

We say a Bernoulli random variable belongs to the Bernoulli *distribution*. This tells us what its PMF will be, regardless of the experiment. Mathematically:

$$X \sim \text{Bern}(p) \tag{1}$$

- X is the name of the random variable.
- The $\sim$ means *is distributed as*.
- p is the probability of a "yes"/1 outcome.

Uniform random variable:

- Different uniform RVs can have different supports.
- can be continuous or discrete, depending on the support.
- They assign an equal probability to every outcome in the support. Used where you have very little knowledge of which outcomes are likely. Mathematically

$$X \sim U(\Omega) \tag{2}$$

Here, Ω is the support (and the sample space!). For instance, suppose the experiment is about how much rain (in millimetres) there is tomorrow, and the support is $\Omega = [0, 100]$.

- Then X is continuous as it can take infinite values
- The probability it takes *some* value is 1. So $\int_0^{100} f_{X(x)} dx = 1$.
- Uniformity means $f_{X(x)}$ is the same value for every x, so it's just a number.
- This means that $f_{X(x)} = \frac{1}{100}$, as $\int_0^{100} \frac{1}{100} dx = 1$

Exercises

1. Do notebooks 1-6 on the probability section of https://marimo-team.github.io/learn/.
2. Is Ω, the sample space, an event? What would you call it in English?
3. Give an example of events that are mutually exclusive but not independent.
4. Give an example of events that are independent but not mutually exclusive .
5. Give an example of events that are independent *and* mutually exclusive .
6. How many days will it rain next year? Describe the experiment, outcome space, event space, and the random variable used to ask the question. *Hint: no single correct answer*

Statistics

What's a statistic (intuitively)?

- Random variables ask *quantitative questions* about an experiment. How much rain will there be tomorrow?
- Sometimes the same experiment happens many times. Each time, the random variable will have a different output.
- Statistics are *summaries* of a random variable: they summarise based on all the different possible outputs and how probable they are.

Remember to always figure out the underlying experiment when somebody invokes a probability

Motivating example: measuring heights

We measure the height of people at Sussex university . In this case:

- The experiment is picking a random person from a sample space Ω consisting of all people at Sussex
- The random variable maps from any given outcome (i.e. person) $\omega \in \Omega$ to a number (their height)
- If we sampled the random variable for every outcome, we would get a list of lots of heights. Confusing!
- A statistic is a *summary* of the random variable. EG the *mean* (average) height. Or the *variance* of the height.

This type of *sampling experiment* is really common. So make sure you understand this box.

Statistics sacrifice *truthfulness* for *interpretability*. If I just tell you the average height at Sussex, I haven't given you the full story about

heights. But the average is much easier to understand and use than a list of thousands of heights!

Sample statistics vs population statistics

A population statistic is the *true* value of a statistic. We rarely know what this is, as we'd need to sample the random variable for *all* possible outcomes of an experiment.

A sample statistic is a summary of a random variable derived from sampling it on a *limited* set of outcomes. We often use sample statistics to *estimate* population statistics. In this case, they are called *estimators*.

Motivating example continued: measuring heights

- To find the average height at Sussex university, we would need to sample height for *every* person. Getting this population statistic is difficult.
- Instead we can sample the variable e.g. 100 times (measure 100 people's heights). The average of these 100 heights is a *sample* statistic. We can use it to estimate the population statistic.

Statistic 1: The expectation

Discrete random variables

Also known as the mean or average

Any statistic is a *function* that maps a random variable (input) to a number (output). We usually call the expectation μ. Here is the formula for μ, if X is a *discrete* random variable.

$$\mu(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega) \tag{3}$$

So…

1. We're taking an outcome ω in the sample space of the experiment.
2. We're calculating its probability $\mathbb{P}(\omega)$ and the value $X(\omega)$ that the random variable takes for that outcome.
3. We're multiplying them together: $X(\omega)\mathbb{P}(\omega)$
4. We're doing steps 1-3 for *every* possible outcome in the sample space, and adding all the results together.

Expectation of a die

- There are 6 outcomes for a standard die. Any outcome is equally likely so $\mathbb{P}(\omega) = \frac{1}{6}, \forall \omega \in \Omega$
- We have a random variable X that maps these outcomes to the numbers 1 to 6.
- Our sum is thus $\mu(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}$
- Or, more succinctly: $\mu(X) = \sum_{i=1}^{6} \frac{1}{6} i = 3.5$.

Equation 3 depends on knowing the outcome space of the experiment. It's *experiment specific*. We're going to give an equivalent formula for expectation that depends *only* on the PMF $f_{X(x)}$ of the random variable:

$$\mu(X) = \sum_{x \in \text{ supp}(X)} x f_x(x) \tag{4}$$

So…

1. We're taking a possible *value* x that the random variable can take.
2. We're calculating the probability $f_{X(x)}$ of the random variable taking that value.
3. We're multiplying the value with the probability: $x f_{X(s)}$.
4. We're doing steps 1-3 for *every* possible value the RV can take, and adding all the results together.

i.e. we are summing over the *support* of the RV.

Question

An RV X takes the value 100 if a die rolls 4,5, or 6. It takes the value 20 if the RV rolls 1,2 or 3.

- What is the support $\text{supp}(X)$?
- Calculate the expectation *twice*, using both formulas of Equation 3 and Equation 4.

Continuous random variables

You should have learnt in maths that integration is like taking a sum over infinite possible values. Similarly, for a continuous random variable, we replace the sum in Equation 4 with an integral:

$$\mu(X) = \int_{x \in \text{ supp}(X)} x f_X(x)$$

Understanding the discrete case properly is the most important.

Linearity of expectation

Definition of linearity

Let f be a function.

f is linear if $f(ax+by) = af(x) + bf(y)$ for *any* $a, b \in \mathbb{R}$

Try this for the linear function $f(x) = 2x + 4$

- Pick values for a, b e.g. $a = 5, b = 7$
- Pick values for x and y similarly
- Manually check that $f(ax+by) = af(x) + bf(y)$
- Now try for $f(x) = x^2$ which is *not* linear.

What's the maths symbol for 'for any'?

Note that linearity can apply to more than functions, see here

Expectation is linear:	$\mathbb{E}[aX + bY] = a\mathbb{E}[x] + b\mathbb{E}[y]$ for any random variables X, Y and any constants $a, b \in \mathbb{R}$
Expectation isn't multiplicative:	$\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$ in general. (Although possibly in individual case)

Statistic 2: The variance

Try 1: How big (in expectation) is the distance between a random variable and its actual value.

$$\mathbb{E}[X - \mathbb{E}[X]] = 0 \quad \text{always} \tag{5}$$

Prove this for yourself using the rules of expectations

Try 2: How big (in expectation) is the *squared* distance between a random variable and its actual value.

VARIANCE IDENTITIES

Formula 1:	$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$
Formula 2:	$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

careful about what is being squared!

Going between formulae

you should be able to do this yourself using rules of expectation

$$\begin{aligned}\mathbb{E}[(X-\mathbb{E}[X])^2] &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\ &\quad \text{(expanded quadratic)} \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[\mathbb{E}[X]^2]] \\ &\quad \text{(using linearity of expectation)} \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2]\end{aligned} \tag{6}$$

Question:

Prove the following formulae using linearity of expectation and formula 2 for the variance

$$\text{var}(cX) = c^2 \text{ var}(X) \tag{7}$$

Gaussian Random Variables

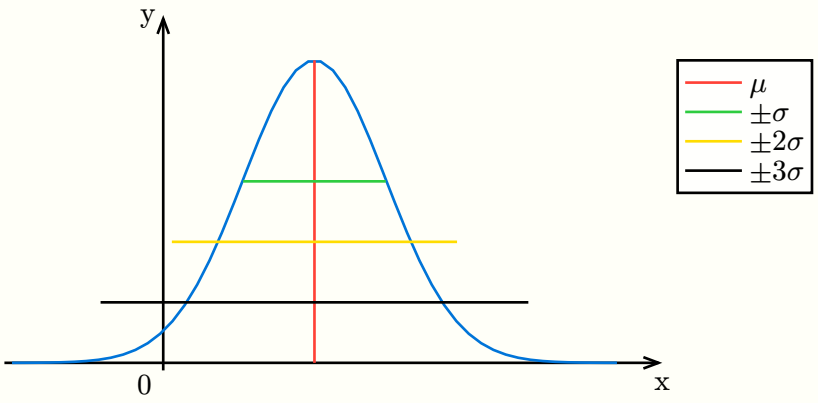
- These are *continuous* random variables
- Also called *normally* distributed random variables or bell curves.
- They are *special* and *ubiquitous* due to the Central Limit Theorem (coming later)

Maths	English
$X \sim \mathcal{N}(\mu, \sigma^2)$	X is distributed ($\sim$) as a Normal (Gaussian) random variable with mean μ and variance σ^2

Here is a plot of the probability density function of a Gaussian X:

don't need to memorise this formula

$f_{X(x)} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$



the smaller σ is, the thinner it looks

Eyeballing: it's important to be able to eyeball the width of the horizontal lines above:

$y\%$ of probability (area under curve)	Within x standard deviations σ from mean μ
$y = 68$	$x = 1$
$y = 95$	$x = 2$
$y = 99.7$	$x = 3$

Additivity

Gaussian variables have a nice mathematical property: the sum of two Gaussians is also Gaussian:

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

Then

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2) \tag{8}$$

Central Limit Theorem

Rough description:

The expectation of groups of random variables looks Gaussian, even if the individual random variables are not Gaussian.

If group size > 30, then statisticians just assume they are Gaussian, sometimes incorrectly.

Mathematical statement

- Let $X(\omega)$ be a random variable with mean μ and variance σ^2.
- Let $X_1, ...X_n$ be n samples of X
- Denote the expected value of n samples as $\bar{X}_n$. Note that $\bar{X}_n$ is itself a random variable

Ask yourself what the experiment is that generates $\bar{X}_n$, and what constitutes an outcome of the experiment

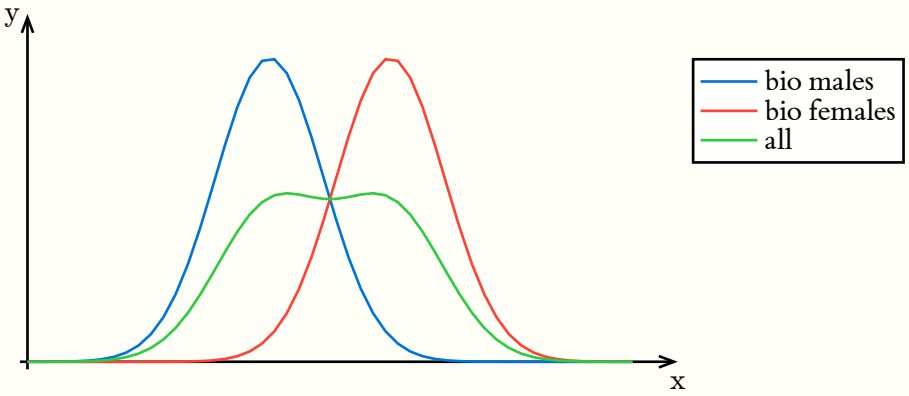
Then, as n gets large, the distribution of $\bar{X}_n$ looks *more and more* like a Gaussian, with mean μ and variance $\frac{\sigma^2}{n}$

Mathematically, we write this as

$$\bar{X}_n \underset{d}{\to} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \tag{9}$$

The symbol $\underset{d}{\to}$ means *tends in distribution*. We don't cover limits or limits of distribution in this course so don't worry about it.

What does this mean? Let's illustrate on a random variable that isn't Gaussian:



- Height does *not* have a Gaussian distribution due to two genders with different mean heights (see above)
- Let $X(\omega)$ be a random variable mapping from people in the UK to their heights. This means X is *not Gaussian*.

Outcome experiment

- We pick a human. Each human is an outcome ω. We measure their height $X(\omega)$

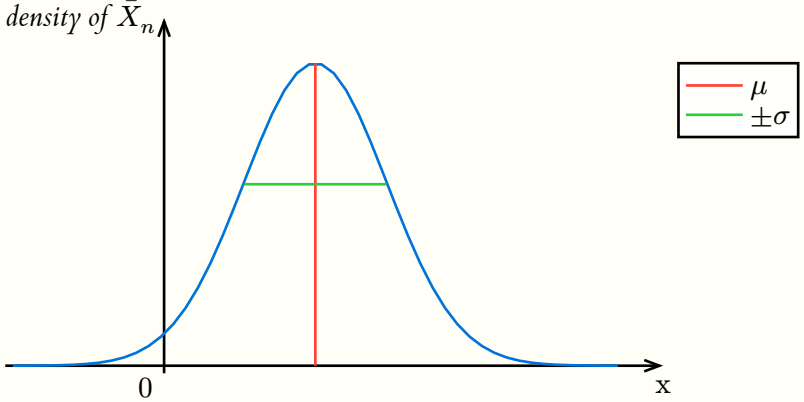
n-sample experiment

- We pick n humans (e.g. $n = 100$). Each set of 100 people is an outcome ω. We measure their average height: $\bar{X}_{n(\omega)}$

We could run the n −sample experiment many times to get many different samples of $\bar{X}_n$.

The probability density of $\bar{X}_n$ (*not* X!) will look *quite similar* to a Gaussian according to the CLT:

Probability density of $\bar{X}_n$



How similar is similar?

In practice, statisticians use $n = 30$ as a cutoff to apply CLT. If $n > 30$, then any sample variable $\bar{X}_n$ (the mean of n individual samples of a random variable X) is assumed to be Gaussian.

However statisticians are often wrong

Test your understanding

- Suppose the mean human height is 170cm and the variance is 20. What are the mean and variance of $\bar{X}_{100}$ according to CLT?

Important Terminology

If $X_1 ... X_n$ are samples from the *same* random variable, we say they are *independent and identically distributed*. This is shortened as i.i.d

Remember i.i.d!

Estimating distributions (from data)

Why bother?

- Sometimes we get data, (*e.g. on people's heights at Sussex uni*)
- We want to use this data to make quantitative predictions about unmeasured data (*e.g. on people's heights at Brighton uni*).
- Pretending data comes from a particular distribution (e.g. Gaussian) makes this easy to do mathematically.

"Based on measuring people at Sussex uni, I predict that fewer than 5% of people at Brighton Uni are more than 7 feet tall"

"Based on lots of historical credit card data, I predict that this person has a $> 95\%$ chance of paying back their loan, given their income, age, etc"

Here the loan to the person is the experiment, and

To do so, we assume that our old and new data come from the *same distribution*.

- In the first example: we assume people at Sussex and Brighton have no systematic height differences. The probability of a person being any given height is independent of whether they are at Sussex or Brighton.
- In the second example: we assume there haven't been events (e.g. a financial crisis) that make historical loan repayments probabilities different from current ones.

This assumption is never completely correct, and hence quantitative predictions are never completely reliable.

Learning from data

We rarely know the true distribution of real data. Heights probably aren't Gaussian. What are they? 😦. We have two options:

Parametric estimation	Pretend you know the distribution is e.g. Gaussian and estimate its statistics (mean/variance/…)
Nonparametric estimation	You know neither the distribution nor the statistics and learn both from data

Parametric estimation has an extra assumption. So it seems worse? Not necessarily!

Fundamental tradeoff in ALL of mathematical modelling	Techniques requiring *fewer* assumptions need *more* data.	Remember this tradeoff in everything you do!

Multivariate distributions

Name	What is it	What does it do
Random variable	Quantitative question about experiment	Maps an experiment outcome to *a single* number
Multivariate random variable	Multiple *related* questions about experiment	Maps an experiment outcome to *several* numbers

Suppose we have an experimental outcome ω.

- A random variable X maps ω to a number $X(\omega)$.

ω could be a randomly chosen person. X could be height and Y could be shoe size

- A random variable Y maps ω to a number $Y(\omega)$

(X, Y) is a *multivariate* random variable

Bivariate, since there are *two* numbers

The probability that X is a particular value is *related* to the probability that Y is a particular value. For instance, a 7 foot tall human is likely to have a large shoe size.

Therefore, we need to have a *joint* PMF/PDF to express the probabilities of X and Y taking different values.

Joint PMF (discrete)	$f_{XY}(x, y) = \mathbb{P}[X = x, Y = y]$
Joint PDF (continuous)	$\int_a^b \int_c^d f_{XY}(x, y) = \mathbb{P}[X \in (a, b), Y \in (c, d)] \ dx \ dy$

In the continuouse case, we have to apply an integral *twice* with respect to two variables to get a probability

As practice, integrate $\int_0^5 \int_2^6 xy^2 \ dx \ dy$. Then $\int_2^6 \int_0^5 xy^2 \ dy \ dx$. Swapping the order of integration doesn't change the outcome.

We've shown an example of a bivariate distribution since (X, Y) maps to two numbers. But we can have multivariate distributions that map to arbitrarily many numbers.

Statistic 3: The covariance

Univariate statistics we looked at before summarised a random variable. The covariance is a *multivariate* statistic: it summarises the *relationship* between random variables. We will look only at the bivariate case here.

Formula

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \tag{10}$$

$\text{cov}(X, Y) > 0$	Knowing that X is big *increases* how big we expect Y to be
$\text{cov}(X, Y) = 0$	Knowing that X is big *doesn't change* how big we expect Y to be
$\text{cov}(X, Y) < 0$	Knowing that X is big *decreases* how big we expect Y to be

Covariance is a *symmetric* measure. So you could swap X and Y in the descriptions of this table and it would still be correct.

Intuition for the covariance is visual and in the lecture notes.

Issue with covariance

Suppose $\text{cov}(X, Y) = 4$. What is $\text{cov}(3X, 4Y)$?

Covariance is a *dimensional* quantity. If two random variables are measured in cm, then their covariance is different than if they are measured in m.

Hence, we more frequently encounter a different statistic, the *correlation*.

$$\text{corr}(X, Y) = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\sqrt{\text{var(X) var(Y)}}} \tag{11}$$

- Variance is always positive, so the sign of the correlation is the same as the sign of the covariance.
- The correlation is *always* between -1 and 1.
- Technically, you could infer the correlation of two variables on a scatter plot without access to the values on the x or y axes. *Unlike* covariance.

Question

Prove the following formula by expanding the formula for variance:

$$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\ \text{cov}(X,Y) \tag{12}$$

Questions

We have samples $[x_1, ..., x_n]$ and $[y_1, ..., y_n]$ of random variables X and Y

1. Write a Python/Julia function that returns the *sample covariance* of X and Y. This should replace the true expected value with the sample mean in the covariance formula.
2. Write a Python/Julia function that returns the *sample correlation* of X and Y.
3. Explain the relationship between your sample correlation and the definition of correlation based on dot product, provided in the linear algebra notes.

The covariance matrix

Now let's do some linear algebra. Suppose we have a *vector* of random variables: $X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$. The covariance matrix is

$$\begin{aligned} &\text{cov}(X) \\ &\quad = \mathbb{E}[X - \mathbb{E}[X]][X - \mathbb{E}[X]]^T \\ &\quad = \begin{bmatrix} \text{cov}(X_1X_1) & \text{cov}(X_1X_2) & ... & \text{cov}(X_1X_n) \\ \text{cov}(X_2X_1) & \text{cov}(X_2X_2) & ... & \text{cov}(X_2X_n) \\ \vdots & \vdots & \vdots & \vdots \\ \text{cov}(X_nX_1) & \text{cov}(X_nX_2) & ... & \text{cov}(X_nX_n) \end{bmatrix} \end{aligned} \tag{13}$$

Question

As Linear Algebra practice, verify for yourself why the vectorised form of the equations above correspond to the matrix.

This is the *true* covariance matrix. Usually, you build the *sample* covariance matrix from data, as you don't know the true distributions of your random variables.

Questions

Consider a matrix of data where each row corresponds to a different random variable, and each column corresponds to a sample.

Most data tables are like this (example below)

In Python/Julia, build a function that returns the *sample* covariance matrix of the random variables using matrix operations where possible. I haven't given you a formula for sample covariance

Student Name	Assignment 1	Assignment 2	Mid-term	Final Exam
Jane Smith	78%	82%	75%	80%
Alex Johnson	90%	95%	94%	96%
John Doe	85%	90%	88%	92%
Maria Garcia	88%	84%	89%	85%
Zhang Wei	93%	89%	90%	91%
Marina Muster	96%	91%	74%	69%