

Random Variables I (Distribution and Expectation)

CS70: Discrete Mathematics and Probability Theory

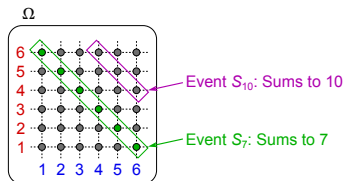
UC Berkeley – Summer 2026

Lecture 19

Ref: Note 16

Motivation: Rolling Two Dice

Often: Care about the sum, not individual rolls – used events for this earlier:



Rolling a 7 is important. Whether it's (1,6) or (2,5) or... isn't.

For outcome (r, b) , what's important is the value $r + b$

Think of as: Interested in some *measurement* of the outcome.

$\Rightarrow$ “Measurement” means numerical value, regardless of sample space.

Motivation: Flip a Coin 10 Times

Ten coin flips: Win a dollar for each “heads”

Outcomes:

$$\Omega = \{ \text{HHHHHHHHHH}, \text{HHHHHHHHHT}, \dots, \text{THHTTTHTHT}, \dots, \text{TTTTTTTTTT} \}$$

How big is Ω ? $|\Omega| = 2^{10} = 1024$

What we care about:

HHHHHHHHHH $\rightarrow$ 10 heads

HHHHHHHHHT $\rightarrow$ 9 heads

...

THHTTTHTHT $\rightarrow$ 4 heads

...

TTTTTTTTTT $\rightarrow$ 0 heads

1024 different outcomes, but we only care about which of 11 values results

Full set of outcomes is *too detailed (or fine-grained)* to be useful

Motivation: Fixed Point of a Permutation

A **permutation** is a bijection $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$

A *re-arrangement* of the values $1, \dots, n$

A **fixed point** in a permutation is an i such that $\sigma(i) = i$

Fixed point doesn't move in re-arrangement

Question: How many fixed points are there in a random permutation?

Example applications/settings:

Random routing with n processors and packets

Randomly returning exam papers to n students in class

Permutations with 3 students:

Permutation:	123	132	213	231	312	321
Num fixed points:	3	1	1	0	0	1

3 students: $|\Omega| = 3! = 6$

$\Rightarrow$ Number of fixed points is in $\{0, 1, 2, 3\}$; 4 possible values (really 3)

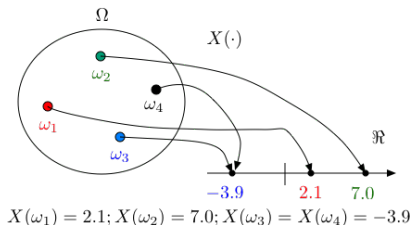
100 students: $|\Omega| = 100! \approx 9.33 \times 10^{157}$

$\Rightarrow$ Number of fixed points is in $\{0, 1, \dots, 100\}$; 101 possible values

Random Variables

Definition: A **random variable** for an experiment with sample space Ω is a function $X : \Omega \rightarrow \mathbb{R}$.

$X(\cdot)$ assigns a real number $X(\omega)$ to each $\omega \in \Omega$:



Function??? Not a variable??? **Yes!** (although variables can represent functions)

Any randomness in X ? **No!** Given outcome, value is deterministic.

$\Rightarrow$ Of course, ω is chosen randomly, but X itself has no “randomness”

So “random variable” X isn’t random, and isn’t a variable...

Random Variable: Sum of Two Dice

Experiment: Roll two dice

Sample Space: $\Omega = \{(1,1), (1,2), (1,3), \dots, (3,4), \dots, (6,6)\} = \{1, \dots, 6\}^2$

Random Variable X : sum of rolls.

$$\left. \begin{array}{l} X((1,1)) = 2 \\ X((1,2)) = 3 \\ \dots \\ X((6,6)) = 12 \end{array} \right\} \text{ For all } (a,b) \in \Omega, X((a,b)) = a + b$$

Since $X : \Omega \rightarrow \mathbb{R}$ is a function, we can consider the **preimage** of X :

$$X^{-1}(2) = \{(1,1)\}$$

$$X^{-1}(3) = \{(1,2), (2,1)\}$$

...

$$X^{-1}(7) = \{(1,6), (1,5), (3,4), (4,3), (5,2), (6,1)\}$$

...

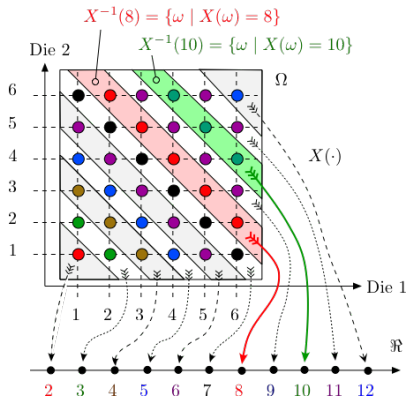
$X^{-1}(s)$ is a subset of outcomes, so is an *event*.

Notation: The event $X^{-1}(s)$ is often written as “ $X = s$ ”

Sets $X^{-1}(s)$ (or “ $X = s$ ”) for all s in range of X *partition* Ω

Random Variable: Sum of Two Dice

In illustration of two dice rolls, random variable values are diagonals:



$$X((2,4)) = 6$$

$$X^{-1}(8) = \{(2,6), (3,5), (4,4), (5,3), (6,2)\} \text{ (also written “} X = 8 \text{”)}$$

$$X^{-1}(10) = \{(4,6), (5,5), (6,4)\} \text{ (also written “} X = 10 \text{”)}$$

Random Variables: Flipping Until Heads

Experiment: Flip a coin until you get a “heads”

Random variable X = “how many flips were needed to get a heads”

Outcomes:

H: $X = 1$

TH: $X = 2$

TTH: $X = 3$

...

TTTTTTH: $X = 7$

...

Is there any upper bound for X ? **No!!**

$X = 1,000$ might be unlikely, but it's not impossible

Same for $X = 1,000,000!$ $\leftarrow$ *unexpected factorial – but that too!*

So:

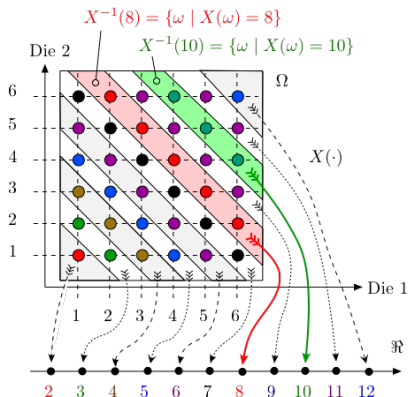
Range of a random variable can be an infinite set

For now: Range will be finite or countably infinite

Random Variables and Probabilities

Events have probabilities $\implies$ Random variable values have probabilities

$$\mathbb{P}[X = a] = \sum_{\omega \in X^{-1}(a)} \mathbb{P}[\omega] = \sum_{\substack{\omega \in \Omega \\ X(\omega) = a}} \mathbb{P}[\omega]$$



$$X^{-1}(8) = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$$

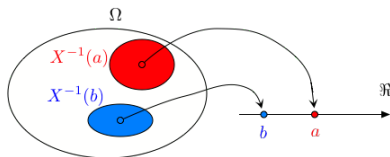
$$\implies \mathbb{P}[X = 8] = \frac{5}{36}$$

$$X^{-1}(10) = \{(4,6), (5,5), (6,4)\}$$

$$\implies \mathbb{P}[X = 10] = \frac{3}{36}$$

Distribution

Definition: The **distribution** of a random variable X , is $\{(a, \mathbb{P}[X = a]) : a \in \mathcal{A}\}$, where $\mathcal{A}$ is the range of X .



The “distribution view” is a probability space.

(a) New sample space: $\mathcal{A}$

(b) For all $a \in \mathcal{A}$, $\mathbb{P}[a]$ (in $\mathcal{A}$) = $\mathbb{P}[X = a]$ (in Ω)

Sanity check! Do probabilities sum to 1 in the “distribution view”?

Events “ $X = a$ ” for $a \in \mathcal{A}$ partition Ω , so

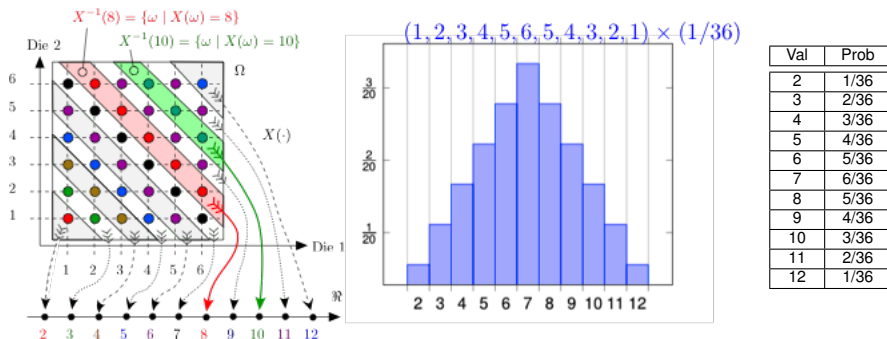
$$\sum_{a \in \mathcal{A}} \mathbb{P}[X = a] = \sum_{a \in \mathcal{A}} \sum_{\substack{\omega \in \Omega \\ X(\omega) = a}} \mathbb{P}[\omega] = \sum_{\omega \in \Omega} \mathbb{P}[\omega] = 1$$

Distribution: Sum of Two Dice

Experiment: roll two dice Random variable $X = \text{"sum of dice"}$

Distribution: Probability for every possible sum $(2, \dots, 12)$

Various equivalent views:



Distribution as pairs: $(2, \frac{1}{36}), (3, \frac{2}{36}), \dots, (7, \frac{6}{36}), \dots, (12, \frac{1}{36})$

Distribution: Flipping a Coin 10 Times

Ten coin flips: Win a dollar for each “heads”

Outcomes:

$$\Omega = \{\text{HHHHHHHHHH}, \text{HHHHHHHHHT}, \dots, \text{THHTTTHTHT}, \dots, \text{TTTTTTTTTT}\}$$

Number of outcomes: $|\Omega| = 2^{10} = 1024$

Random variable X = “Number of heads”

$\Rightarrow$ Possible values: $0, 1, 2, \dots, 10$

Number of ways to get k heads: $\binom{10}{k}$

$$\Rightarrow \mathbb{P}[X = k] = \binom{10}{k} \frac{1}{1024}$$

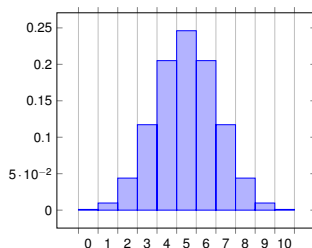
$$\mathbb{P}[X = 0] = \frac{1}{1024}$$

$$\mathbb{P}[X = 1] = \frac{10}{1024}$$

$$\mathbb{P}[X = 2] = \frac{45}{1024}$$

$$\mathbb{P}[X = 3] = \frac{120}{1024}$$

...



Distribution: Handing Back Assignments (3 students)

Experiment: Hand back assignments randomly to 3 students.

Permutation:	123	132	213	231	312	321
Num fixed points:	3	1	1	0	0	1

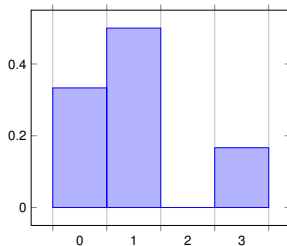
Distribution:

$$\mathbb{P}[X = 0] = \frac{2}{6} = \frac{1}{3}$$

$$\mathbb{P}[X = 1] = \frac{3}{6} = \frac{1}{2}$$

$$\mathbb{P}[X = 2] = 0$$

$$\mathbb{P}[X = 3] = \frac{1}{6}$$



Note: 2 isn't actually in the range of X , so pair $(2, 0)$ can be omitted:

$$\{(0, \frac{1}{3}), (1, \frac{1}{2}), (3, \frac{1}{6})\}$$

Distribution: Flipping Until Heads

Experiment: Flip a coin until you get a “heads”

Random variable X = “how many flips were needed to get a heads”

Outcomes:

H: $X = 1$

TH: $X = 2$

TTH: $X = 3$

...

TTTTTTH: $X = 7$

...

Let event T_i = “tails on flip number i ”

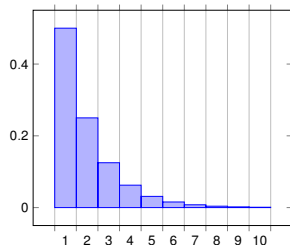
$$\mathbb{P}[X = 1] = \mathbb{P}[\overline{T_1}] = \frac{1}{2}$$

$$\mathbb{P}[X = 2] = \mathbb{P}[T_1 \cap \overline{T_2}] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\mathbb{P}[X = 3] = \mathbb{P}[T_1 \cap T_2 \cap \overline{T_3}] = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

...

$$\mathbb{P}[X = k] = \frac{1}{2^k}$$



Continues infinitely to the right...

Flipping a Coin 4 Times – What Do You Expect?

Four coin flips: Win a dollar for each “heads”

Outcomes: $\Omega = \{HHHH, HHHT, \dots, THHT, \dots, TTTT\}$

Random variable X = “Number of heads”

X	Outcomes	# outcomes	Probability
0	TTTT	1	1/16
1	HTTT THTT TTHT TTTH	4	4/16
2	HHTT HTHT HTTH THHT THTH TTHH	6	6/16
3	THHH HTHH HHTH HHHT	4	4/16
4	HHHH	1	1/16

With everyday language: How many dollars do you expect to win?

Intuitively: About half of the flips are heads, so expect to win \$2.

Let's try computing “weighted average” for X

$$\frac{1}{16} \cdot 0 + \frac{4}{16} \cdot 1 + \frac{6}{16} \cdot 2 + \frac{4}{16} \cdot 3 + \frac{1}{16} \cdot 4 = \frac{0+4+12+12+4}{16} = \frac{32}{16} = 2$$

Sensible! Let's formalize this...

Expectation

Definition: The **expected value** of a random variable X with range $\mathcal{A}$ is

$$\mathbb{E}[X] = \sum_{a \in \mathcal{A}} a \cdot \mathbb{P}[X = a].$$

The expected value is also called the **mean**.

More Intuition: If we repeat an experiment a large number of times and if $X_1, \dots, X_N$ are the successive values of the random variable.

What's the *average* of our trials? $\frac{X_1 + \dots + X_N}{N}$

Is this related to expected value? *Yes!*

$$\frac{X_1 + \dots + X_N}{N} \approx \mathbb{E}[X]$$

This is formalized in the (nontrivial) result called the **Law of Large Numbers**.

We will get back to this in a future lecture.

Expectation: Alternative View

Instead of summing over range of X , sum over underlying outcomes.

Theorem:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}[\omega].$$

Proof: Remember that “ $X = a$ ” events partition Ω . Then

$$\begin{aligned}\mathbb{E}[X] &= \sum_{a \in \mathcal{A}} a \cdot \mathbb{P}[X = a] \\ &= \sum_{a \in \mathcal{A}} a \sum_{\substack{\omega \in \Omega \\ X(\omega) = a}} \mathbb{P}[\omega] \\ &= \sum_{a \in \mathcal{A}} \sum_{\substack{\omega \in \Omega \\ X(\omega) = a}} X(\omega) \cdot \mathbb{P}[\omega] \\ &= \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}[\omega]\end{aligned}$$



Expectation: Flipping a Coin 3 Times

Three coin flips: Win a dollar for each “heads”

Random variable X = “Number of heads”

X	Outcomes	# outcomes	Probability
0	TTT	1	1/8
1	HTT THT TTH	3	3/8
2	HHT HTH THH	3	3/8
3	HHH	1	1/8

What is the expected value of X ?

$$\mathbb{E}[X] = \frac{1}{8} \cdot 0 + \frac{3}{8} \cdot 1 + \frac{3}{8} \cdot 2 + \frac{1}{8} \cdot 3 = \frac{0+3+6+3}{8} = \frac{12}{8} = \frac{3}{2}$$

Difference between mathematical expectation and English-language “expect”:

Do you expect 1.5 heads?

Even with integer-valued random variables, expectation may be a non-integer!

Expectation and Average

“Expectation” is from probability theory ; “Average” is from statistics

There are n students in the class; exam scores $s_1, s_2, \dots, s_n$

“Average score” of the n students: add scores and divide by n :

$$\text{Average} = \frac{s_1 + s_2 + \dots + s_n}{n}.$$

Experiment: choose a student uniformly at random.

⇒ Uniform sample space over $\Omega = \{1, 2, \dots, n\}$, so $\mathbb{P}[\omega] = 1/n$, for all ω .

⇒ Random Variable: $X(\omega) = s_\omega$ (midterm score of student ω)

Expectation:

$$\mathbb{E}[X] = \sum_{\omega \in \{1, \dots, n\}} X(\omega) \cdot \mathbb{P}[\omega] = \sum_{\omega \in \{1, \dots, n\}} s_\omega \cdot \frac{1}{n} = \text{Average}.$$

Expected value of uniformly-selected outcome = average value of outcomes

⇒ Uniform distribution of students, not random distribution of scores

Common Distributions: Bernoulli

Random variable X with two values: $\{0, 1\}$ (or $\{\text{failure}, \text{success}\}$)

$$\mathbb{P}[X = 1] = p$$

$$\mathbb{P}[X = 0] = 1 - p$$

Completely determined by parameter p

$$\mathbb{E}[X] = 0 \cdot (1 - p) + 1 \cdot p = p$$

We say that X is distributed as Bernoulli random variable with parameter p

$\Rightarrow$ *Notation:* $X \sim \text{Bernoulli}(p)$ or $X \sim \text{Ber}(p)$

General distribution appears in many real-world problems:

- Coin flip (potentially biased): Heads or tails
- Equipment: Working or broken
- Testing: Pass or fail
- Medicine: Sick or not

Counting Coin Flips: Poll

Experiment: Flip n coins with heads probability p .

Which of these is true?

(A) Number of outcomes is 2^n .

Yes!

(B) Number of possibilities with k heads $\binom{n}{pk}$.

No. pk may not even be an integer...

(C) Probability of k heads is $p^k/2^n$.

No. This was kind of random...

(D) Number of possibilities with k heads $\binom{n}{k}$

Yes!

(E) The probability of every outcome is $1/2^n$.

No. But... true if $p = 1/2$.

Common Distributions: Binomial

n independent Bernoulli trials, each with success probability p

X = “number of successes”

Just like our many “how many heads” examples...

One experimental outcome might look like:

Trial number	1	2	3	4	5	...	$n-1$	n
Trial outcomes	0	1	1	0	1	...	1	0

How many outcomes have i successes out of n trials? $\binom{n}{i}$

Consider outcome: fail, success, success, fail, ...

$\Rightarrow$ Independent, so probability: $(1-p) \cdot p \cdot p \cdot (1-p) \dots$

Probability of *a specific outcome* with i success and $n-i$ fails: $p^i(1-p)^{n-i}$

So probability of *any* sequence with i successes: $\mathbb{P}[X=i] = \binom{n}{i} p^i (1-p)^{n-i}$

Common Distributions: Binomial

n independent Bernoulli trials, each with success probability p

Probability exactly i successes: $\mathbb{P}[X = i] = \binom{n}{i} p^i (1 - p)^{n-i}$

Sanity-check: Do the probabilities add up to 1?

Cute trick: Consider Binomial theorem for $(x + y)^n$ (with $n \in \mathbb{N}$):

$$(x + y)^n = \sum_{i=0}^n \binom{n}{i} x^i y^{n-i}.$$

With $x = p$ and $y = (1 - p)$, RHS is $\sum_{i=0}^n \mathbb{P}[X = i]$

$\Rightarrow$ LHS is $(p + (1 - p))^n = 1^n = 1$

$\Rightarrow$ *So yes, this is a good distribution!*

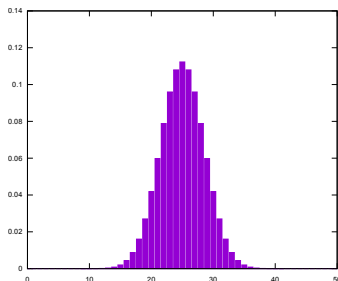
We say X is distributed as Binomial random variable with parameters n and p

$\Rightarrow$ *Notation:* $X \sim \text{Binomial}(n, p)$ or $X \sim \text{Bin}(n, p)$

$\mathbb{E}[X] = np$ **Proof:** “Trust me.” $\square$ (OK, we’ll see a simple proof in a few slides)

Common Distributions: Binomial

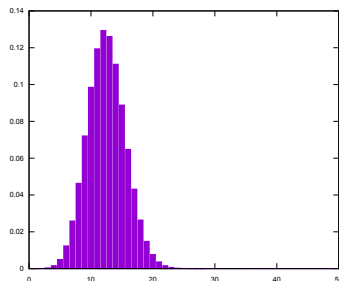
What does distribution look like?



Binomial(50, 1/2)

For $p = 1/2$:

- ⇒ Centered at $n/2$
- ⇒ Symmetric



Binomial(50, 1/4)

For $p = 1/4$:

- ⇒ Centered at $n/4$
- ⇒ No longer symmetric

Multiple Random Variables

Experiment: Toss two coins. $\Omega = \{HH, TH, HT, TT\}$.

$$X_1(\omega) = \begin{cases} 1, & \text{if coin 1 is heads} \\ 0, & \text{otherwise} \end{cases}$$

$$X_2(\omega) = \begin{cases} 1, & \text{if coin 2 is heads} \\ 0, & \text{otherwise} \end{cases}$$

$$X_s(\omega) = X_1(\omega) + X_2(\omega) = \begin{cases} 2, & \text{if both coins are heads} \\ 1, & \text{if exactly one coin is heads} \\ 0, & \text{otherwise} \end{cases}$$

If events “ $X_1 = 1$ ” and “ $X_2 = 1$ ” are independent:

$$\mathbb{P}[X_s = 2]? \quad \mathbb{P}[(X_1 = 1) \cap (X_2 = 1)] = \mathbb{P}[X_1 = 1] \cdot \mathbb{P}[X_2 = 1] = 1/4$$

$$\mathbb{P}[X_s = 1]? \quad 1/2$$

$$\mathbb{P}[X_s = 0]? \quad 1/4$$

Notation: $\mathbb{P}[X_1 = a, X_2 = b] = \mathbb{P}[(X_1 = a) \cap (X_2 = b)]$

$\Rightarrow$ Read as “Probability that X_1 is a and X_2 is b ”

Multiple Random Variables: Joint/Marginal Distribution

Definition: The **joint distribution** of random variables X and Y is $\{(a, b), \mathbb{P}[X = a, Y = b]\} : a \in \mathcal{A}, b \in \mathcal{B}\}$, where $\mathcal{A}$ and $\mathcal{B}$ are the possible values of X and Y , respectively.

Probability of pairs... must add up to 1:

$$\sum_{a \in \mathcal{A}, b \in \mathcal{B}} \mathbb{P}[X = a, Y = b] = 1$$

Definition: The **marginal distribution** for X , given the joint distribution of X and Y , sums over all values of Y for each value of X (basically, ignores the effect of Y).

$$\Rightarrow \text{Marginal for } X: \mathbb{P}[X = a] = \sum_{b \in \mathcal{B}} \mathbb{P}[X = a, Y = b].$$

$$\Rightarrow \text{Marginal for } Y: \mathbb{P}[Y = b] = \sum_{a \in \mathcal{A}} \mathbb{P}[X = a, Y = b].$$

Multiple Random Variables: Joint/Marginal Distribution

Example joint and marginal distributions:

$X \backslash Y$	1	2	3	X
1	.2	.1	.1	.4
2	0	0	.3	.3
3	.1	0	.2	.3
Y	.3	.1	.6	

Black: Joint distribution of X and Y .

$\Rightarrow$ Add them up! Sum of all is 1.

Blue: Marginal distributions (sum across row or down column).

Can also define conditional probabilities (same as for events):

$$\Rightarrow \mathbb{P}[X = a | Y = b] = \frac{\mathbb{P}[X=a, Y=b]}{\mathbb{P}[Y=b]} \quad (\text{assuming } \mathbb{P}[Y = b] > 0)$$

$$\text{Example: } \mathbb{P}[X = 2 | Y = 3] = \frac{\mathbb{P}[X=2, Y=3]}{\mathbb{P}[Y=3]} = \frac{.3}{.6} = \frac{1}{2}$$

Independent Random Variables

Recall: Events A and B are independent iff $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$.

Definition: Random variables X and Y are **independent** if and only if

$$\mathbb{P}[X = a, Y = b] = \mathbb{P}[X = a] \cdot \mathbb{P}[Y = b], \text{ for all } a \text{ and } b.$$

$X = a$ and $Y = b$ are events, and all such events must be independent.

If $\mathbb{P}[X = a] > 0$ for all a and $\mathbb{P}[Y = b] > 0$ for all b , this is equivalent to:

$$\mathbb{P}[X = a | Y = b] = \mathbb{P}[X = a] \text{ for all } a$$

Note: Which is true if and only if $\mathbb{P}[Y = b | X = a] = \mathbb{P}[Y = b]$ for all b

For more variables, can define *pairwise independent*, *mutually independent*, and *k-wise independent* like we did for events.

Independence 1: Two Coin Flips

Experiment: Toss two coins. $\Omega = \{HH, TH, HT, TT\}$.

$$X_1(\omega) = \begin{cases} 1, & \text{if coin 1 is heads} \\ 0, & \text{otherwise} \end{cases}$$

$$X_2(\omega) = \begin{cases} 1, & \text{if coin 2 is heads} \\ 0, & \text{otherwise} \end{cases}$$

$X_1 \backslash X_2$	0	1	X_1
0	.25	.25	.5
1	.25	.25	.5
X_2	.5	.5	

Question: Are X_1 and X_2 independent?

Yes. Intuition: Nothing linking two separate coin tosses.

Math, not intuition – check all values in joint distribution:

$$\mathbb{P}[X_1 = 0, X_2 = 0] = .25 \quad \mathbb{P}[X_1 = 0] \cdot \mathbb{P}[X_2 = 0] = .5 \cdot .5 = .25$$

$$\mathbb{P}[X_1 = 0, X_2 = 1] = .25 \quad \mathbb{P}[X_1 = 0] \cdot \mathbb{P}[X_2 = 1] = .5 \cdot .5 = .25$$

$$\mathbb{P}[X_1 = 1, X_2 = 0] = .25 \quad \mathbb{P}[X_1 = 1] \cdot \mathbb{P}[X_2 = 0] = .5 \cdot .5 = .25$$

$$\mathbb{P}[X_1 = 1, X_2 = 1] = .25 \quad \mathbb{P}[X_1 = 1] \cdot \mathbb{P}[X_2 = 1] = .5 \cdot .5 = .25$$

Independence 2: Two Coin Flips and Sum

Experiment: Toss two coins. $\Omega = \{HH, TH, HT, TT\}$.

$$X_1(\omega) = \begin{cases} 1, & \text{if coin 1 is heads} \\ 0, & \text{otherwise} \end{cases}$$

$$X_2(\omega) = \begin{cases} 1, & \text{if coin 2 is heads} \\ 0, & \text{otherwise} \end{cases}$$

$$X_s(\omega) = X_1(\omega) + X_2(\omega) = \begin{cases} 2, & \text{if both coins are heads} \\ 1, & \text{if exactly one coin is heads} \\ 0, & \text{otherwise} \end{cases}$$

$X_1 \backslash X_s$	0	1	2	X_1
0	.25	.25	0	.5
1	0	.25	.25	.5
X_s	.25	.5	.25	

Question: Are X_1 and X_s independent?

Tip: Look for zero probability entries in a non-zero row and column.

Here: $\mathbb{P}[X_1 = 1, X_s = 0] = 0$, but $\mathbb{P}[X_1 = 1] = 0.5$ and $\mathbb{P}[X_s = 0] = 0.25$.

So $\mathbb{P}[X_1 = 1] \cdot \mathbb{P}[X_s = 0] = 0.5 \cdot 0.25 = 0.125 \dots$ which is $\neq \mathbb{P}[x_1 = 1, X_s = 0] = 0$

Linearity of Expectation: Simple Form

Theorem: For any random variables X and Y ,

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

Important: Any random variables – no requirement that they are independent!

Proof:

$$\begin{aligned}\mathbb{E}[X + Y] &= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega))\mathbb{P}[\omega] \\ &= \sum_{\omega \in \Omega} (X(\omega)\mathbb{P}[\omega] + Y(\omega)\mathbb{P}[\omega]) \\ &= \sum_{\omega \in \Omega} X(\omega)\mathbb{P}[\omega] + \sum_{\omega \in \Omega} Y(\omega)\mathbb{P}[\omega] \\ &= \mathbb{E}[X] + \mathbb{E}[Y]\end{aligned}$$



Linearity of Expectation: Generalized

Theorem: For any constants $a_1, a_2, \dots, a_n \in \mathbb{R}$ and random variables $X_1, X_2, \dots, X_n$,

$$\mathbb{E}[a_1 X_1 + \dots + a_n X_n] = a_1 \mathbb{E}[X_1] + \dots + a_n \mathbb{E}[X_n] .$$

Proof:

$$\begin{aligned}\mathbb{E}[a_1 X_1 + \dots + a_n X_n] &= \sum_{\omega \in \Omega} (a_1 X_1 + \dots + a_n X_n)(\omega) \mathbb{P}[\omega] \\ &= \sum_{\omega \in \Omega} (a_1 X_1(\omega) + \dots + a_n X_n(\omega)) \mathbb{P}[\omega] \\ &= a_1 \sum_{\omega \in \Omega} X_1(\omega) \mathbb{P}[\omega] + \dots + a_n \sum_{\omega \in \Omega} X_n(\omega) \mathbb{P}[\omega] \\ &= a_1 \mathbb{E}[X_1] + \dots + a_n \mathbb{E}[X_n]\end{aligned}$$

□

A specific form of interest has just *one* random variable: $\mathbb{E}[cX] = c\mathbb{E}[X]$

Linearity Example 1: Sum of Dice Rolls

Experiment: Roll a 6-sided die n times.

X_i = value on roll i .

$X = X_1 + \cdots + X_n$ = sum of n rolls.

$$\begin{aligned}\mathbb{E}[X] &= \text{??} \mathbb{E}[X_1 + \cdots + X_n] && \text{(this is not } \mathbb{E}[n \cdot X_1]) \\ &= \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] && \text{(by linearity of expectation)} \\ &= n\mathbb{E}[X_1] && \text{(identically distributed)}\end{aligned}$$

One die roll is easy: $\mathbb{E}[X_1] = 1 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{6 \times 7}{2} \times \frac{1}{6} = \frac{7}{2}$.

Therefore: $\mathbb{E}[X] = \frac{7n}{2}$.

Note: Computing $\sum_x x \cdot \mathbb{P}[X = x]$ directly is not easy!

Linearity Example 2: Binomial Distribution

Recall: $X \sim \text{Bin}(n, p) \implies X$ is the number of successes in n Bernoulli trials

Expectation given earlier with the proof “Trust me.” No need to trust me:

Binomial is number of Bernoulli successes, so define $X_i \sim \text{Bernoulli}(p)$:

$$X_i = \begin{cases} 1 & \text{if success (probability } p); \\ 0 & \text{otherwise} \end{cases}$$

Then: $X = X_1 + X_2 + \dots + X_n$

$$\mathbb{E}[X] = \mathbb{E}[X_1 + X_2 + \dots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n]$$

Each $X_i \sim \text{Bernoulli}(p)$ so $\mathbb{E}[X_i] = p$ for all $i = 1, \dots, n$.

Therefore $\mathbb{E}[X] = np$

Wow... that was easy. Compare to trying to compute

$$\sum_{i=0}^n i \cdot \mathbb{P}[X = i] = \sum_{i=0}^n i \binom{n}{i} p^i (1-p)^{n-i}$$

Indicator Random Variables

Definition: Let A be an event. The random variable I_A defined by

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A \end{cases}$$

is called the **indicator variable** of event A .

For events $A_1, A_2, \dots$ we often use $I_1, I_2, \dots$ rather than $I_{A_1}, I_{A_2}, \dots$

Note that $\mathbb{P}[I_A = 1] = \mathbb{P}[A]$ and $\mathbb{P}[I_A = 0] = 1 - \mathbb{P}[A]$.

Hence,

$$\mathbb{E}[I_A] = 1 \cdot \mathbb{P}[I_A = 1] + 0 \cdot \mathbb{P}[I_A = 0] = \mathbb{P}[I_A = 1] = \mathbb{P}[A].$$

Indicator variables + linearity of expectation is *very powerful*.

$\Rightarrow$ *Especially* when counting random events...

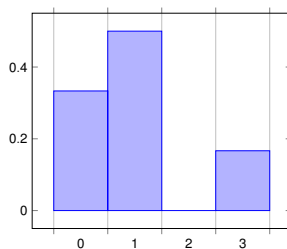
Linearity Example 3: Number of Fixed Points

Recall this problem: Hand back assignments randomly to 3 students.

Permutation:	123	132	213	231	312	321
Num fixed points:	3	1	1	0	0	1

Distribution:

$$X = \begin{cases} 0, & \text{probability } \frac{1}{3} \\ 1, & \text{probability } \frac{1}{2} \\ 2, & \text{probability } 0 \\ 3, & \text{probability } \frac{1}{6} \end{cases}$$



Distribution skips over 2?!?! Looks like a complicated distribution...

⇒ Even moreso for n students!

Indicator variables and linearity of expectation to the rescue!

Linearity Example 3: Number of Fixed Points

Experiment: Hand back assignments randomly to n students.

X = number of students that get their own assignment back.

Event A_m = student m gets their own assignment back

$\Rightarrow I_m$ = indicator variable for event A_m

$$I_m(\omega) = \begin{cases} 1 & \text{if } \omega \in A_m; \\ 0 & \text{otherwise.} \end{cases}$$

So $X(\omega) = I_1(\omega) + \cdots + I_n(\omega)$ (number of fixed points in ω)

$\mathbb{P}[I_m = 1] = \mathbb{P}[A_m] = \frac{1}{n}$. Student is equally likely to get any assignment.

$\Rightarrow$ For all m , $\mathbb{E}[I_m] = \frac{1}{n}$

$$\mathbb{E}[X] = \mathbb{E}[I_1 + \cdots + I_n] = \mathbb{E}[I_1] + \cdots + \mathbb{E}[I_n] = n \cdot \frac{1}{n} = 1.$$

Linearity of expectation applied in second step – are I_m 's independent?

Irrelevant!

Conclusion: The expected number of fixed points is 1.

$\Rightarrow$ Regardless of n . $n = 3$, $n = 150$, or $n = 300,000,000\ldots$ always $\mathbb{E}[X] = 1$

Summary – Random Variables

- A random variable X is a function $X : \Omega \rightarrow \mathbb{R}$
- Preimages $X^{-1}(a)$ are just events
$$\mathbb{P}[X = a] := \mathbb{P}[X^{-1}(a)] = \mathbb{P}[\{\omega \mid X(\omega) = a\}]$$
- Distribution of X is the set of pairs: $\{(a, \mathbb{P}[X = a]), a \in \mathcal{A}\}$.
- Expected value: $\mathbb{E}[X] := \sum_a a \cdot \mathbb{P}[X = a]$
- Common distributions:
 - Bernoulli (success in one trial): $X \sim \text{Ber}(p)$
 - Binomial (success in n trials): $X \sim \text{Bin}(n, p)$
- Multiple random variables:
 - Joint distributions and marginal distributions
 - Independence
- Linearity of expectation: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- Examples: Rolling dice, flipping coins, fixed points of permutation, ...