Random Variables I (Distribution and Expectation)

CS70: Discrete Mathematics and Probability Theory

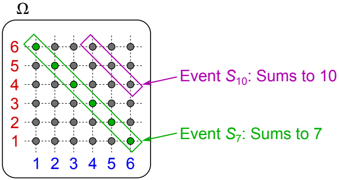
UC Berkeley – Summer 2026

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:

$$\Omega = \{\texttt{HHHHHHHHHH}, \texttt{HHHHHHHHHT}, \ldots, \texttt{THHTTTHTHT}, \ldots, \texttt{TTTTTTTTTT}\}$$

How big it Ω? $|\Omega| =$

What we care about:

$\texttt{HHHHHHHHHH} \rightarrow$ 10 heads
$\texttt{HHHHHHHHHT} \rightarrow$ 9 heads
$\ldots$
$\texttt{THHTTTHTHT} \rightarrow$ 4 heads
$\ldots$
$\texttt{TTTTTTTTTT} \rightarrow$ 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 $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$
A *re-arrangement* of the values $1, \ldots, n$

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

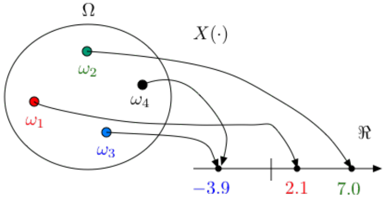
3 students: $|\Omega| = 3! = 6$
$\Rightarrow$ Number of fixed points is in $\{0, 1, 2, 3\}$; 4 possible values (really 3)

100 students: $|\Omega| =$
$\Rightarrow$ Number of fixed points is in $\{0, 1, \ldots, 100\}$; 101 possible values

Random Variables

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

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



$X(\omega_1) = 2.1; X(\omega_2) = 7.0; X(\omega_3) = X(\omega_4) = -3.9$

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

Random Variable X: sum of rolls.

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

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

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

$X^{-1}(3) =$

$\cdots$

$X^{-1}(7) =$

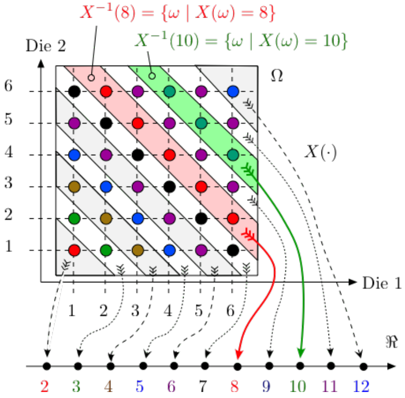
$\cdots$

$X^{-1}(s)$ 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

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



$X((2,4)) =$
$X^{-1}(8) =$
$X^{-1}(10) =$

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`: $X = 3$
$\ldots$
`TTTTTTH`: $X = 7$
$\ldots$

Is there any upper bound for X?

$X = 1{,}000$ might be unlikely, but it's not impossible
Same for $X = 1{,}000{,}000!$ $\leftarrow$ *unexpected factorial – but that too!*

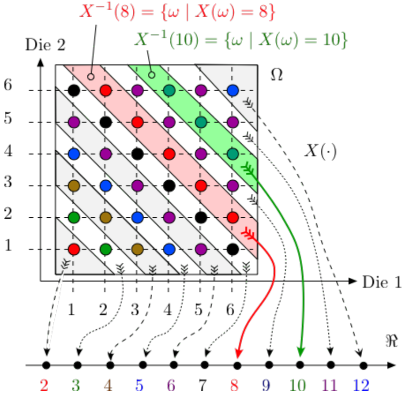
So:

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

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



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

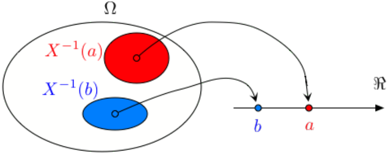
$\Rightarrow \quad \mathbb{P}[X=8] =$

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

$\Rightarrow \quad \mathbb{P}[X=10] =$

Distribution

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



The "distribution view" is a probability space.

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

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

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

Events "$X=a$" for $a \in \mathscr{A}$ partition Ω, so

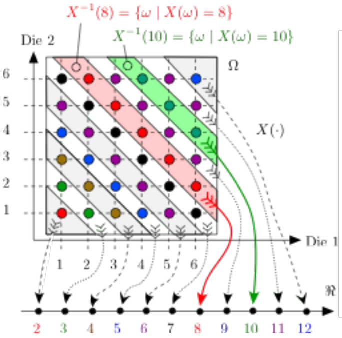
$$\sum_{a \in \mathscr{A}} \mathbb{P}[X=a] = \sum_{a \in \mathscr{A}} \sum_{\substack{\omega \in \Omega \\ X(\omega)=a}} \mathbb{P}[\omega] = \sum_{\omega \in \Omega} \mathbb{P}[\omega] = 1$$

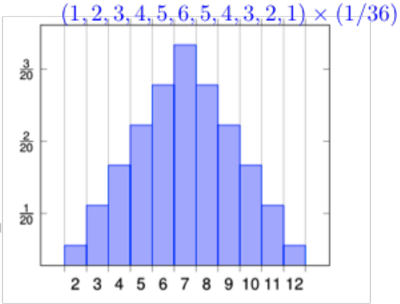
Distribution: Sum of Two Dice

Experiment: roll two dice   Random variable $X =$ "sum of dice"

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

Various equivalent views:





Val	Prob
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

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

Distribution: Flipping a Coin 10 Times

Ten coin flips: Win a dollar for each "heads"

Outcomes:

$\Omega = \{\mathtt{HHHHHHHHHH}, \mathtt{HHHHHHHHHT}, \ldots, \mathtt{THHTTTHTHT}, \ldots, \mathtt{TTTTTTTTTT}\}$

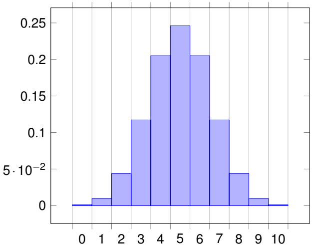
Number of outcomes: $|\Omega| = 2^{10} = 1024$

Random variable $X =$ "Number of heads"

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

Number of ways to get k heads:

$\Rightarrow \quad \mathbb{P}[X = k] =$



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

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

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

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

$\ldots$

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

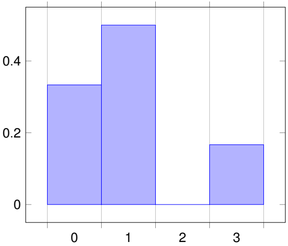
Distribution:

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

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

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

$\mathbb{P}[X = 3] =$



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"

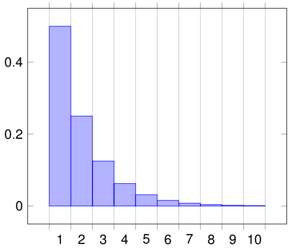
Random variable $X =$ "how many flips were needed to get a heads"

Outcomes:

$\texttt{H}$: $X = 1$
$\texttt{TH}$: $X = 2$
$\texttt{TTH}$: $X = 3$
$\cdots$
$\texttt{TTTTTTH}$: $X = 7$
$\cdots$

Let event $T_i =$ "tails on flip number i"

$\mathbb{P}[X = 1] =$
$\mathbb{P}[X = 2] =$
$\mathbb{P}[X = 3] =$
$\cdots$
$\mathbb{P}[X = k] =$



Continues infinitely to the right...

Flipping a Coin 4 Times – What Do You Expect?

Four coin flips: Win a dollar for each "heads"

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

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 $\mathscr{A}$ is

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

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

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

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

Is this related to expected value? *Yes!*

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

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

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

□

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?

Expectation and Average

"Expectation" is from probability theory ; "Average" is from statistics

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

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

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

Experiment: choose a student uniformly at random.

- $\Rightarrow$ Uniform sample space over $\Omega = \{1, 2, \cdots, n\}$, so $\mathbb{P}[\omega] = 1/n$, for all ω.
- $\Rightarrow$ Random Variable: $X(\omega) = s_\omega$ (midterm score of student ω)

Expectation:

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

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

- $\Rightarrow$ Uniform distribution of students, not random distribution of scores

Common Distributions: Bernoulli

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

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

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

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

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

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?

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

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

Probability of *a specific outcome* with i success and $n-i$ fails:

So probability of *any* sequence with i successes:

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!*

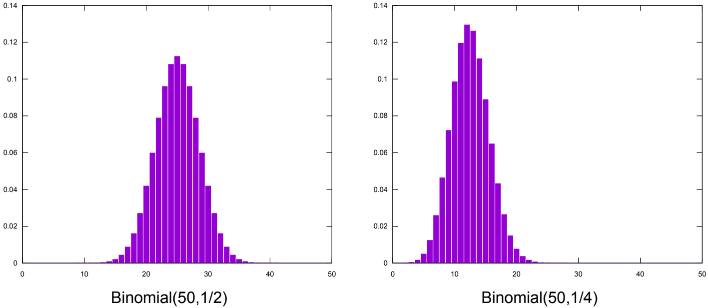
We say X is distributed as Binomial random variable with parameters n and p

$\Rightarrow$ *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?



For $p = 1/2$:

- $\Rightarrow$ Centered at $n/2$
- $\Rightarrow$ Symmetric

For $p = 1/4$:

- $\Rightarrow$ Centered at $n/4$
- $\Rightarrow$ No longer symmetric

Multiple Random Variables

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

$$X_1(\omega) = \begin{cases} 1, & \text{if coin 1 is heads} \\ 0, & \text{otherwise} \end{cases} \qquad X_2(\omega) = \begin{cases} 1, & \text{if coin 2 is heads} \\ 0, & \text{otherwise} \end{cases}$$

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

If events "$X_1 = 1$" and "$X_2 = 1$" are independent:

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

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

$\mathbb{P}[X_s = 0]$?

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 <u>and</u> X_2 is b"

Multiple Random Variables: Joint/Marginal Distribution

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

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

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

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).

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

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

Multiple Random Variables: Joint/Marginal Distribution

Example joint and marginal distributions:

X \ 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):

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

Example: $\mathbb{P}[X=2|Y=3] =$

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:

$\mathbb{P}[X = a | Y = b] = \mathbb{P}[X = a]$ 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. $\Omega = \{HH, TH, HT, TT\}$.

$$X_1(\omega) = \begin{cases} 1, & \text{if coin 1 is heads} \\ 0, & \text{otherwise} \end{cases} \qquad X_2(\omega) = \begin{cases} 1, & \text{if coin 2 is heads} \\ 0, & \text{otherwise} \end{cases}$$

X_1 \ 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?

Independence 2: Two Coin Flips and Sum

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

$$X_1(\omega) = \begin{cases} 1, & \text{if coin 1 is heads} \\ 0, & \text{otherwise} \end{cases} \qquad X_2(\omega) = \begin{cases} 1, & \text{if coin 2 is heads} \\ 0, & \text{otherwise} \end{cases}$$

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

X_1 \ 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?

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:

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

Linearity of Expectation: Generalized

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

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

Proof:

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

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.

$$\mathbb{E}[X] \quad =$$

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

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 $X_i \sim \text{Bernoulli}(p)$:

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

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

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

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

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 $A_1, A_2, \cdots$ we often use $I_1, I_2, \cdots$ rather than $I_{A_1}, I_{A_2}, \cdots$

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...

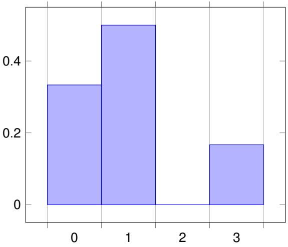
Linearity Example 3: Number of Fixed Points

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...

$\Rightarrow$ 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*)

$\mathbb{P}[I_m = 1] = \mathbb{P}[A_m] =$

$\Rightarrow$ 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?

Conclusion: The expected number of fixed points is 1.
$\Rightarrow$ Regardless of n. $n = 3$, $n = 150$, or $n = 300,000,000$... always $\mathbb{E}[X] = 1$

Summary – Random Variables

- A random variable X is a function $X : \Omega \to \mathbb{R}$
- 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: $\{(a, \mathbb{P}[X = a]), a \in \mathscr{A}\}$.
- 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, ...