Discrete Probability

CS 330: Discrete Structures

"probability" : (noun) the extent to which an event is likely to occur, measured by the ratio of the favorable cases to the whole number of cases possible (New Oxford American Dictionary)

experiment: a procedure that yields one of a set of possible outcomes

e.g., rolling a six-sided die

sample space: the set of possible outcomes

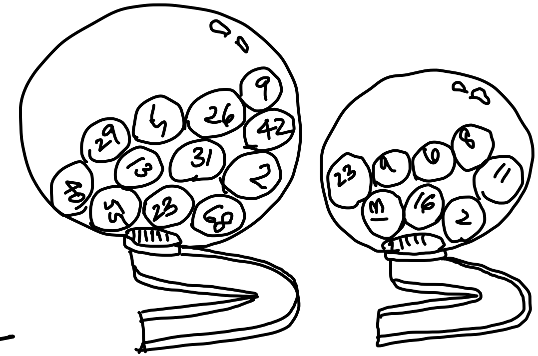
e.g., 1, 2, 3, 4, 5, 6

event: a subset of the sample space

e.g., rolling a 2, rolling an even #

probability of an event E: $p(E) = \frac{|E|}{|S|}$, given sample space S

e.g., $p(\text{rolling an even \# w/ a six-sided die}) = \frac{3}{6} = \frac{1}{2}$



e.g., odds of winning Mega-Millions jackpot

- five numbers in range 1–70 (no duplicates, order doesn't matter)
- 1 number in range 1–25 ("Mega ball")

all numbers must match.

$$|S| = \binom{70}{5} \cdot 25 = 302{,}575{,}350$$

$$\text{odds of winning} = \frac{1}{302{,}575{,}350}$$

e.g., probability of having a full house after being dealt five cards from a regular deck of playing cards?

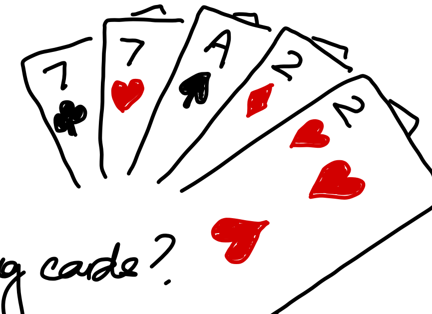
$$\begin{array}{r} 4 \text{ suits} \\ \times\ 13 \text{ ranks} \\ \hline 52 \text{ cards} \end{array}$$

$$|S| = \binom{52}{5} = 2{,}598{,}960$$

$$|\text{full house}| = \underbrace{13}_{\text{ranks for triple}} \cdot \underbrace{\binom{4}{3}}_{\text{suits for triple}} \cdot \underbrace{12}_{\text{ranks for pair}} \cdot \underbrace{\binom{4}{2}}_{\text{suits for pair}} = 3{,}744$$

$$P(\text{full house}) = \frac{3{,}744}{2{,}598{,}960} \approx 0.00144 \approx 0.144\,\%$$

e.g., probability of having two pairs after being dealt five cards from a regular deck of playing cards?



$$|\text{two pairs}| = \frac{13\cdot\binom{4}{2}\cdot 12\cdot\binom{4}{2}\cdot 11\cdot\binom{4}{1}}{2} = 123{,}552$$

← because order of pairs doesn't matter! (division rule)

$$P(\text{two pairs}) = \frac{123{,}552}{2{,}598{,}960} \approx 4.75\%$$

sum rule:

if E_1 and E_2 are disjoint events, $p(E_1 \cup E_2) = p(E_1) + p(E_2)$,

and generally, for pairwise disjoint events $E_1, E_2, \ldots, E_n$,

$$p\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} p(E_i)$$

Complement rule:

if E is an event in the sample space S, $P(\bar{E}) = 1 - P(E)$

e.g., p(rolling anything other than a 7 w/ two 6-sided dice)?

$$1 - p(\text{rolling a } 7)$$

$$= 1 - \frac{|\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}|}{6 \cdot 6}$$

$$= 1 - \frac{1}{6}$$

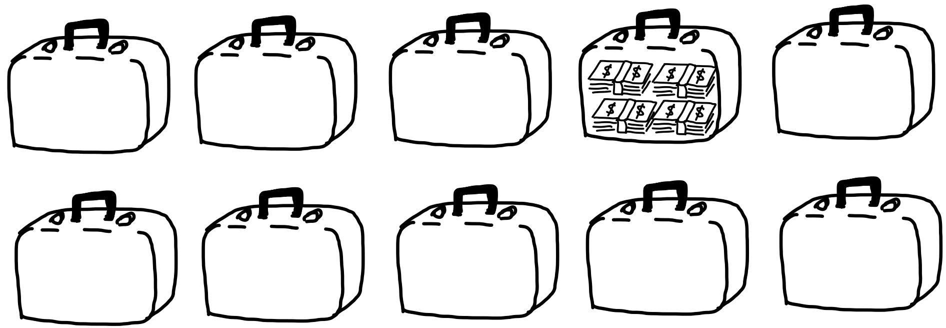
$$= \frac{5}{6}$$

e.g., probability of being dealt a 5-card hand that contains at least one Ace?

$$|\text{hands w/o Ace}| = \binom{48}{5}$$

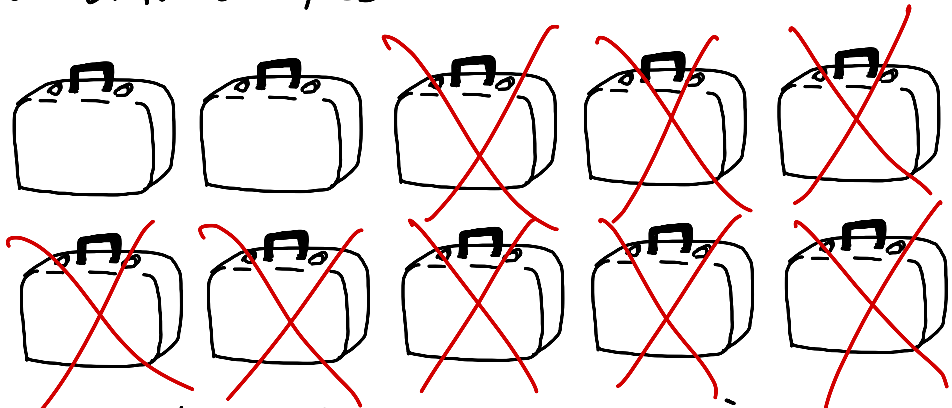
$$p(\text{hand w/ at least one Ace}) = 1 - \frac{\binom{48}{5}}{\binom{52}{5}}$$

e.g., "deal or no deal, CS 330 edition":



- 9 empty suitcases, 1 w/ $1,000,000.
- you pick a suitcase at the start of class
- every 10 minutes I reveal one of the suitcases (other than the one you picked) to be empty, until there are 2 left.
- at any point you can choose a suitcase and leave w/ its contents.

e.g., "deal or no deal, CS 330 edition":

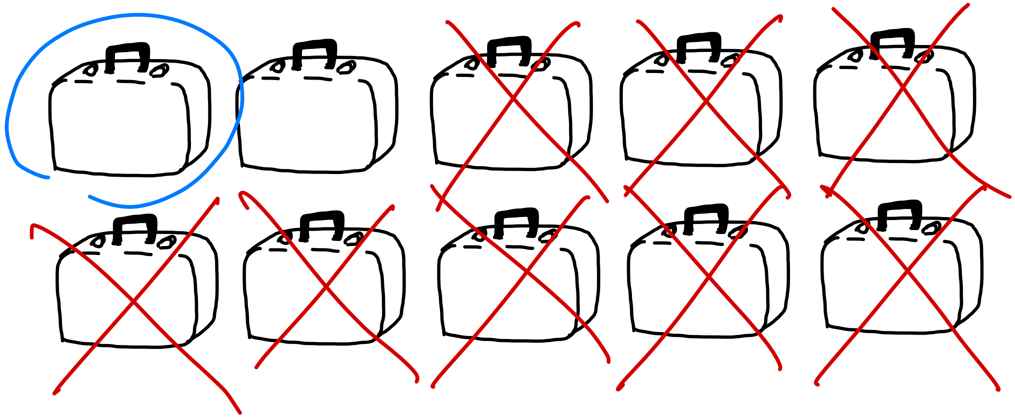


is it worth waiting for me to reveal 8 suitcases to be empty, or do you have the same odds of leaving w/ $1M if you choose early?

definitely worth waiting! 10 vs. 2 suitcases to pick from!

e.g., "deal or no deal, CS 330 edition":

(Monty Hall problem)



after I've revealed the 8^{th} empty suitcase, do you persist in opening the suitcase you picked initially, or do you switch? (are the odds of leaving w/ \$1M any different?)

switch! $P(\text{original pick} = \$1M) = \frac{1}{10}$ $P(\text{switch pick} = \$1M) = \frac{9}{10}$

Conditional probability:

if E and F are events w/ $P(F) > 0$, the probability of E given that F has already occurred (i.e., probability of E conditioned on F) is:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

e.g., In a best-of-three tournament, the IIT women's soccer team wins the first game w/ probability $\frac{1}{2}$. The probability of winning any following game is:

$\frac{2}{3}$, if the preceding game was won, and

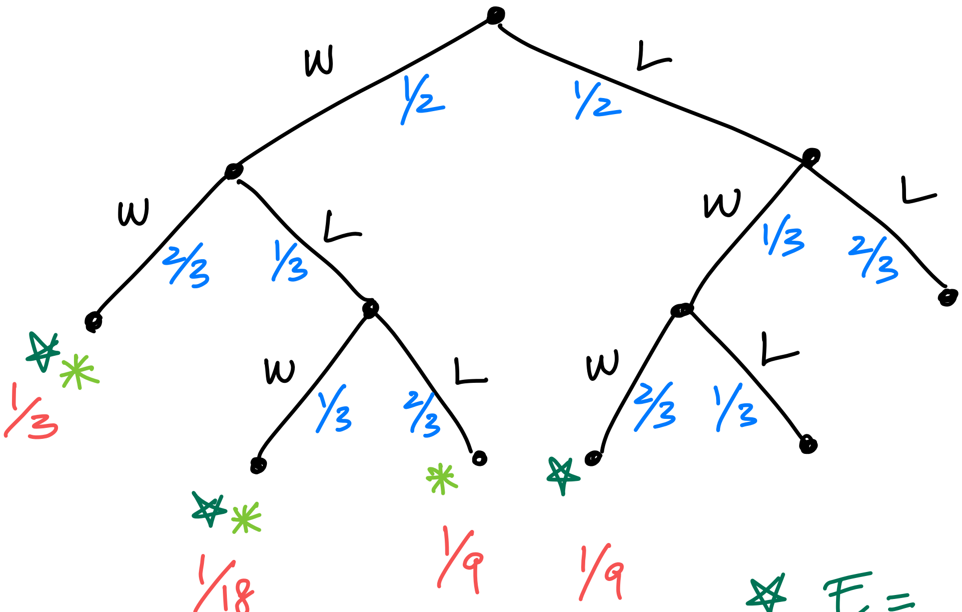
$\frac{1}{3}$, if the preceding game was lost.

What is the probability that we won the tournament, given that we won the first game?

E = we won the tournament, F = we won the first game

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

tree diagram:



$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{1}{3} + \frac{1}{18}}{\frac{1}{3} + \frac{1}{18} + \frac{1}{9}} = \frac{\frac{7}{18}}{\frac{9}{18}} = \frac{7}{9}$$

E = win tournament

F = win first game

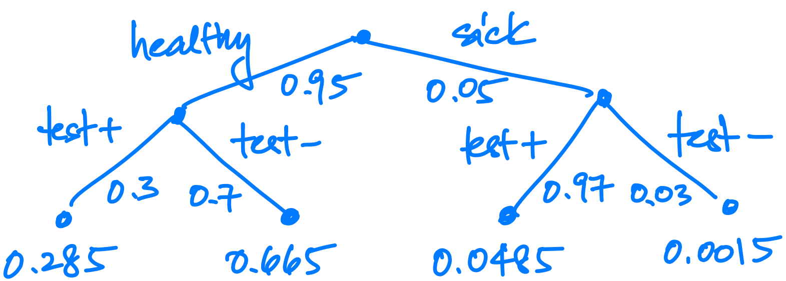
S = {WW, WLW, WLL, LWW, LWL, LL}

e.g., Consider a covid-19 test with a 3% false negative rate, and a 30% false positive rate. i.e.,

- if you have covid-19, there is a 3% chance the test says you don't
- if you don't have covid-19, there is a 30% chance the test says you do

assuming an infection rate of 5%, how accurate is the test?

i.e., if E is the event that someone has covid-19, and F is the event that the test is positive, what is $P(E|F)$?



$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

$$= \frac{0.0485}{0.0485 + 0.285} \approx 14.5\%$$

e.g., based on the preceding example, if the women's soccer team won the tournament, what is the likelihood that they won the first game?

from before, E = won tournament, $P(E|F) = \frac{7}{9}$
F = won first game,

now we want $P(F|E) \leftarrow$ "a posteriori" probability (E occurs after F!)

conditional probability $P(E|F) = \frac{P(E \cap F)}{P(F)}$

$$P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{P(E|F)P(F)}{P(E)} = \frac{\frac{7}{9}\,\frac{1}{2}}{\frac{1}{2}} = \frac{7}{9}$$

Bayes' Rule:

if E and F are events where $p(E) > 0$ and $p(F) > 0$,

$$p(E|F) = \frac{p(F|E)\,p(E)}{p(F)}$$

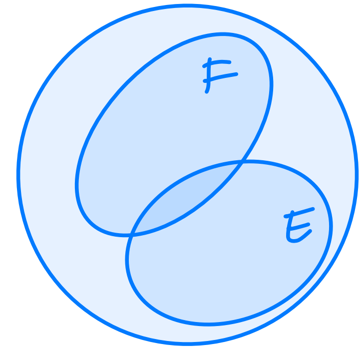
interpretations (philosophical):

Bayesian – we are computing a "degree of belief" in proposition E given evidence F ← P(E|F) tells me how likely it is you have COVID

Frequentist – we are measuring the relative # of outcomes in which events E & F occur ← you either have COVID or not, but this describes the population at large

e.g., E = has COVID
F = positive test

Bayes' Rule (extended form):



Bayes' $\quad p(E|F) = \dfrac{p(F|E)\,p(E)}{p(F)}$

know that $p(F) = p(F \cap E) + p(F \cap \bar{E})$

by conditional probability: $p(F \cap E) = p(F|E)\,p(E)$

$p(F \cap \bar{E}) = p(F|\bar{E})\,p(\bar{E})$

$$\therefore\ p(E|F) = \frac{p(F|E)\,p(E)}{p(F|E)\,p(E) + p(F|\bar{E})\,p(\bar{E})}$$

e.g., Consider a covid-19 test with a 3% false negative rate, and a 30% false positive rate. I.e.,

- if you have covid-19, there is a 3% chance the test says you don't
- if you don't have covid-19, there is a 30% chance the test says you do

assuming an infection rate of 5%, how accurate is the test?

E = someone has covid-19

F = test is positive , using

$$P(E|F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|\bar{E})P(\bar{E})} = \frac{(0.97)(0.05)}{(0.97)(0.05) + (0.3)(0.95)}$$

$$\approx 14.5\%$$

Independence:

if E and F are events w/ $p(F) > 0$, the events are independent iff:

$$p(E|F) = p(E), \text{ and}$$

$$p(E \cap F) = p(E) \cdot p(F)$$

e.g., probability of rolling "snake-eyes" (two 1's) using two six-sided dice

E = rolling a 1 on first die

F = rolling a 1 on second die

$P(E) = \frac{1}{6}$ $\quad P(F) = \frac{1}{6}$

$P(E|F) = \frac{1}{6}$

$P(E \cap F) = P(E) \cdot P(F) = \frac{1}{36}$

Mutual independence:

a set of events $E_1, E_2, \ldots, E_n$ are mutually independent iff for any subset of the events $E_i, \ldots, E_j$,

$$p(E_i \cap \ldots \cap E_j) = p(E_i) \cdots p(E_j)$$

e.g., the three events E_1, E_2, E_3 are mutually independent if

$$p(E_1 \cap E_2) = p(E_1)p(E_2)$$

$$p(E_1 \cap E_3) = p(E_1)p(E_3)$$

$$p(E_2 \cap E_3) = p(E_2)p(E_3)$$

$$p(E_1 \cap E_2 \cap E_3) = p(E_1)p(E_2)p(E_3)$$

e.g., suppose we flip three coins and consider these events:

E_1 = coin 1 matches coin 2

E_2 = coin 2 matches coin 3

E_3 = coin 3 matches coin 1

are E_1, E_2, E_3 mutually independent?

$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

$p(E_1) = p(\{HHH, HHT, TTH, TTT\}) = \frac{1}{2}$

$p(E_2) = p(E_3) = \frac{1}{2}$ as well, by symmetry.

$p(E_1 \cap E_2) = p(\{HHH, TTT\}) = \frac{1}{4} = p(E_1)p(E_2)$

$p(E_1 \cap E_3) = p(E_1)p(E_3)$ and $p(E_2 \cap E_3) = p(E_2)p(E_3)$ by symmetry.

no!

$p(E_1 \cap E_2 \cap E_3) = p(\{HHH, TTT\}) = \frac{1}{4} \neq p(E_1)p(E_2)p(E_3)$

k-way independence:

a set of events $E_1, E_2, \ldots, E_n$ are k-way independent iff every k-sized subset of these events is mutually independent

(2-way indepence is aka "pairwise" independence)

e.g., the events on the previous page are pairwise independent, but <u>not</u> mutually independent!

when we want to perform mathematical analysis of probabilities, especially across many different events, focusing on individual events is unwieldy.

e.g., p(flipping a coin heads up 10 times in a row)

p(flipping a coin heads up between 0–10 times in a row)

\# of times to flip a coin before we expect to see heads

prefer to write:

$$P(C=10)\ ,\ P(C\leq 10)$$

Random Variables

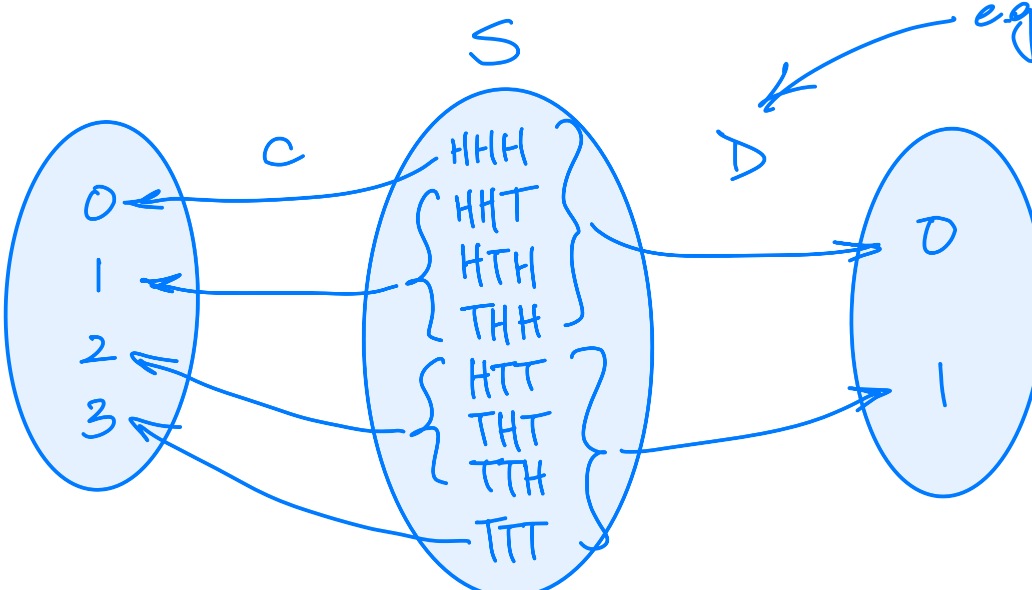
a R.V. is a complete function from the sample space of an experiment onto $\mathbb{R}$ (the set of real numbers)

e.g., given an experiment w/ sample space S, we can describe a R.V. $X : S \to \mathbb{R}$

e.g. based on the experiment of tossing 3 coins, define the random variables

C: which maps each outcome to its # of tails

D: which maps each outcome to 1 if it has at least 2 tails, and 0 otherwise



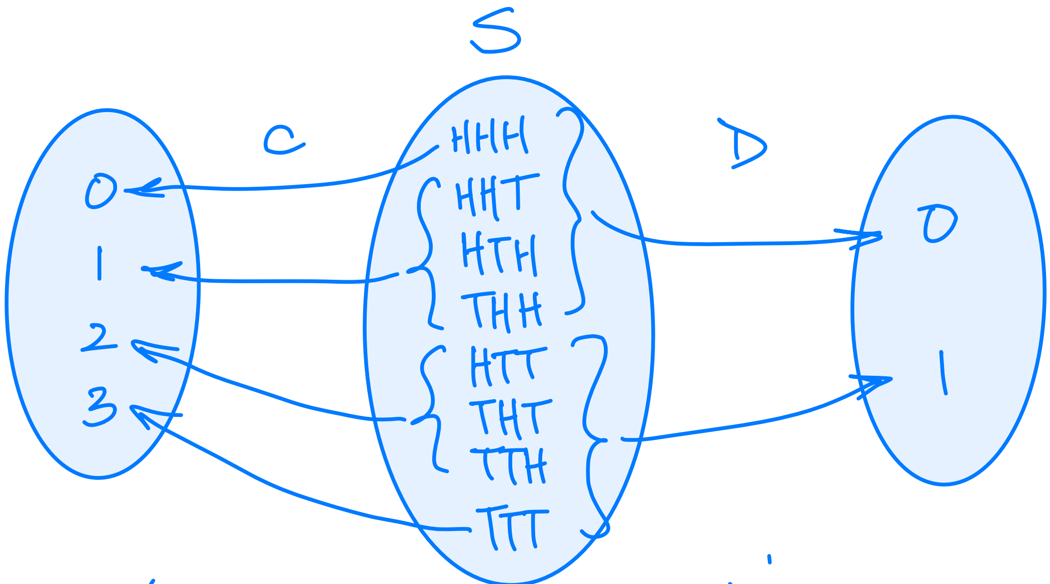
e.g. of Bernoulli R.V. (0/1 valued) aka. "characteristic" or "indicator" variable.

given sample space S and R.V. X, the event where $X=y$ is :

$$\{w \in S \mid X(w)=y\}$$

and the probability of this event is :

$$p(X=y) = \sum_{w \in S \mid X(w)=y} p(w)$$



e.g.,

$$P(C=1) = 3/8$$

$$P(C<2) = 4/8 = 1/2$$

$$P(C\geq 2) = 4/8 = 1/2 = P(D=1)$$

$$P(C\geq 0) = 1$$

observations:

- a R.V. partitions the sample space
- for any R.V. X w/ range R,

$$\sum_{y\in R} P(X=y) = 1$$

R.V.s X and Y are independent iff

$$\forall x\, y \in \mathbb{R} \left(p(X=x \wedge Y=y) = p(X=x) \cdot p(Y=y) \right)$$

alternatively, using conditional probability:

$$\forall x\, y \in \mathbb{R} \left(p(X=x \mid Y=y) = p(X=x) \text{ or } p(X=x) = 0 \right)$$

given a R.V. X, the probability mass function (PMF) is:

$$f(x) = p(X=x)$$

and the cumulative distribution function (CDF) is:

$$F(x) = p(X \leq x) = \sum_{y \leq x} p(X=y)$$

together, the PMF and CDF describe the distribution of probabilities over the range of a R.V.

— many R.V.s have the same distributions, and frequently arising distributions are well studied. The most common distributions used in computer science are:

1. the Bernoulli distribution
2. the uniform distribution
3. the Binomial distribution

The Bernoulli distribution describes a R.V. w/ range $\{0, 1\}$, where

$f(0) = p$, $f(1) = 1-p$, and

$F(0) = p$, $F(1) = 1$

a Bernoulli R.V. describes the probability of success/failure of a "Bernoulli trial" — an experiment w/ two possible outcomes.

e.g., flipping a coin; success = H ($p = 0.5$)

rolling two six-sided dice; success ≥ 11 ($p = 3/36 = 1/13$)

The uniform distribution describes a R.V. w/ range $R = \{a, a+1, \ldots, b-1, b\}$ where all values are assigned the same probability, i.e.,

$$\forall k \in R \quad f(k) = \frac{1}{|R|}$$

$$\forall k \in R \quad F(k) = \frac{k-a+1}{|R|}$$

uniform R.V.s are found in many "fair" experiments w/ multiple outcomes

e.g., rolling a six-sided die ; $f(\text{any outcome}) = \frac{1}{6}$

drawing a card from a shuffled deck ; $f(\text{any card}) = \frac{1}{52}$

an element being at position k in an unsorted array of size N ; $f(k) = \frac{1}{N}$

The ==Binomial distribution== describes a R.V. which counts the # of successes in n independent Bernoulli trials

e.g., consider flipping a coin four times, where "success" = H

R.V. X maps each outcome in S to the # of H's

$|S| = 2^4 = 16$

$f(1) = \frac{1}{16}$

$f(3) = \frac{\binom{4}{3}}{16} = \frac{4}{16} = \frac{1}{4}$

For a Binomial R.V. that models $n \geq 1$ independant Bernoulli trials, each w/ probabilty $0 < p < 1$ of sucess:

$$f_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

If $p = \frac{1}{2}$ ("fair"/"unbiased" trials), we have:

$$f_n(k) = \binom{n}{k} \frac{1}{2^n}$$

e.g., what is the probability that exactly 50 of 100 coin tosses result in a heads?

$$\binom{100}{50}\frac{1}{2^{100}} \approx 7.9\%$$

e.g. what is the probability that between 1-25 of 100 coin tosses result in a heads?

$$\sum_{k=1}^{25}\binom{100}{k}\frac{1}{2^{100}} \approx 0.000028\%$$

the **Expected Value** (aka average/mean) of a R.V. X over the sample space S is:

$$E(X) = \sum_{w \in S} P(w)\, X(w)$$

e.g. what is the expected value of rolling a 6-sided die?

outcomes = $\{1, 2, 3, 4, 5, 6\}$

$P(X) = \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6}$

$$E(X) = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

E(X) can also be computed as the weighted average of values in the range R of X:

$$E(X) = \sum_{y \in R} y \, p(X=y)$$

proof: $E(X) = \sum_{w \in S} p(w) X(w) = \sum_{y \in R} \sum_{w \in S \mid X(w)=y} p(w) \cancel{X(w)}\, y$

$$= \sum_{y \in R} y \cancel{\sum_{w \in S \mid X(w)=y} p(w)}\, p(X=y) \qquad = \sum_{y \in R} y \, p(X=y)$$

e.g., consider a game where you roll two 6-sided die and win \$1000 if you get a 2, and lose \$100 otherwise. Would you play?

What would you expect to win/lose per game?

$$\text{expected winnings} = (+\$1000)\frac{1}{36} + (-\$100)\frac{35}{36} = -\$69.44$$

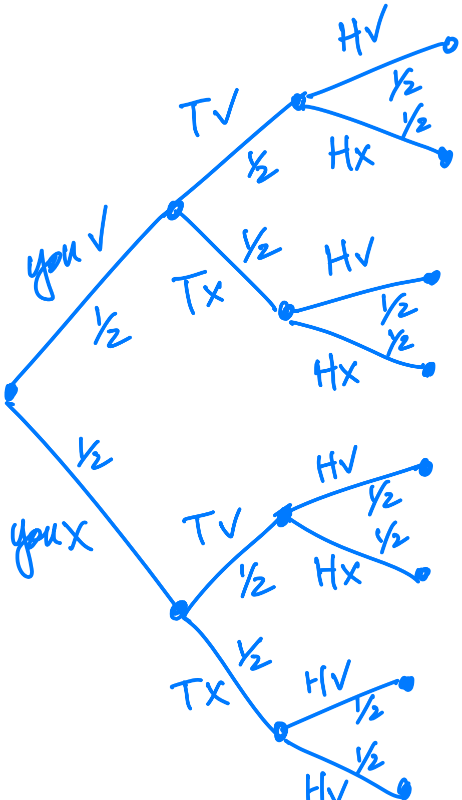
<u>Don't play</u>!

e.g., consider a game where 3 players wager \$10 each on the outcome of a coin toss. If all players guess correctly/wrongly, nobody wins, otherwise the players who guess wrongly lose their wagers, and the players who guess correctly split the pot.

Would you play?

demo:

You	Tom	Harry.	Result
T(+5)	T(+5)	H (-10)	T
H (0)	H (0)	H (0)	T
H (+20)	T (-10)	T (-10)	H



probability	winnings
$\frac{1}{8}$	0
$\frac{1}{8}$	+5
$\frac{1}{8}$	+5
$\frac{1}{8}$	+20
$\frac{1}{8}$	−10
$\frac{1}{8}$	−10
$\frac{1}{8}$	−10
$\frac{1}{8}$	0

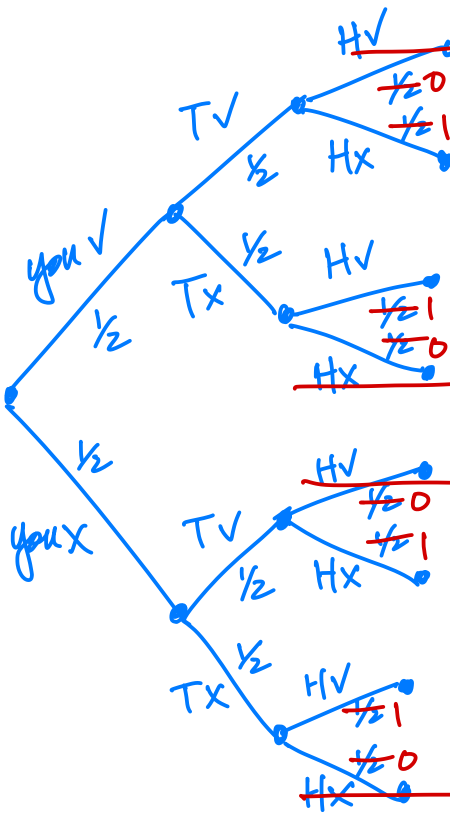
expected winnings per game = 0

OK to play!

e.g., consider a game where 3 players wager $10 each on the outcome of a coin toss. If all players guess correctly/wrongly, nobody wins, otherwise the players who guess wrongly lose their wagers, and the players who guess correctly split the pot.

You play 100 games, and have lost over $250. Is this just bad luck, or is there some other explanation?

the other players must be cheating!



probability	winnings
~~$\frac{1}{8}$~~	~~0~~
$\cancel{\frac{1}{8}}$ $\frac{1}{4}$	+5
$\cancel{\frac{1}{8}}$ $\frac{1}{4}$	+5
~~$\frac{1}{8}$~~	~~+20~~
~~$\frac{1}{8}$~~	~~−10~~
$\cancel{\frac{1}{8}}$ $\frac{1}{4}$	−10
$\cancel{\frac{1}{8}}$ $\frac{1}{4}$	−10
~~$\frac{1}{8}$~~	~~0~~

expected winnings per game $= \cancel{0}$

$= \frac{10}{4} = \$2.50$

Tom and Harry are colluding!
(and likely splitting their winnings)

When R.V. $X : S \to \mathbb{N}$, then we can also compute:

$$E(X) = \sum_{i=0}^{\infty} p(X > i) = \sum_{i=1}^{\infty} p(X \geq i)$$

proof: $\sum_{i=0}^{\infty} p(X > i) =$

$$
\begin{aligned}
& p(X>0) &&= p(X=1) + p(X=2) + p(X=3) + \dots \\
+ & p(X>1) &&= + \qquad\qquad\quad p(X=2) + p(X=3) + \dots \\
+ & p(X>2) &&= + \qquad\qquad\qquad\qquad\quad\; p(X=3) + \dots \\
& \vdots \\
& &&\overline{\quad 1 \cdot p(X=1) + 2 \cdot p(X=2) + 3 \cdot p(X=3) + \dots \quad} \\
& &&= E(X)
\end{aligned}
$$

e.g. consider a network router that drops each incoming packet w/ probability q (mutually independently)

on average, how long until the first dropped packet?

let X = # of first dropped packet ; find $E(X)$

$$E(X) = \sum_{i=0}^{\infty} \underbrace{p(X > i)}$$

p(no packets dropped up to ith packet)

$= p(1 \text{ not dropped}) \cdot p(2 \text{ not dropped}) \cdots p(i \text{ not dropped})$

$= (1-q) \cdot (1-q) \cdots (1-q)$

$= (1-q)^i$

$$E(X) = \sum_{i=0}^{\infty} \underbrace{(1-q)}_{r}{}^{i} = \underbrace{\sum_{i=0}^{\infty} r^{i}}_{\text{geometric series}}, \quad 0<r<1 \quad = \frac{1}{1-r} = \frac{1}{q}$$

first n terms: $\quad s = r^0 + r^1 + r^2 + \dots + r^{n-1}$

$$rs = r^1 + r^2 + \dots + r^n$$

$$s - rs = r^0 - r^n$$

$$s(1-r) = 1 - r^n$$

$$s = \frac{1-r^n}{1-r}$$

$$\lim_{n \to \infty} \frac{1-r^n}{1-r} = \frac{1}{1-r}$$

e.g. consider a network router that drops each incoming packet w/ 0.1% probability.

on average, how long until the first dropped packet?

$$\frac{1}{0.001} = 1000 \text{ packets}.$$ (1000th is dropped)

"mean time to failure" — MTTF

e.g. suppose you flip a coin repeatedly until you see a heads. How many tails are you likely to see before the first heads?

MTTF, where "failure" is heads

$$\frac{1}{\frac{1}{2}} = 2$$ — i.e., second flip, on average, is head,

so $2 - 1 = 1$ tail is expected.

e.g. suppose you flip a coin repeatedly until you see at least one tail and one head. On average, how many coin flips are required?

1st flip is H or T

then, on average 2 flips to see the other face

i.e., $1+2=3$ flips

Expected values obey a rule called "Linearity of Expectations",
which says that for R.V.s $X_1, X_2, \ldots, X_n$ on sample space S,

$$E(\underbrace{X_1 + X_2 + \ldots + X_n}_{\text{note: independence is not necessary!}}) = E(X_1) + E(X_2) + \ldots + E(X_n)$$

and that for any R.V. X and $a, b \in \mathbb{R}$,

$$E(aX + b) = aE(X) + b$$

e.g. "hat-check problem": a hat-check clerk loses track of which of n hats belong to whom, and returns them at random.

What is the average # of hats returned correctly?

let X = # of customers that get their hat back

$$E(X) = \sum_{i=1}^{n} i\, p(X=i) \quad ?$$

$$p(X=i) = \begin{cases} \dfrac{1}{i!\,(n-i)}, & 1 \le i \le n-2 \\ \dfrac{1}{n!}, & n-2 < i \le n \end{cases}$$

e.g. "hat-check problem": a hat-check clerk loses track of which of n hats belong to whom, and returns them at random.

What is the average # of hats returned correctly?

let X_i be the Bernoulli R.V. that indicates if customer i receives the correct hat.

$$E(X_i) = 0 \cdot p(X_i = 0) + 1 \cdot p(X_i = 1) = \frac{1}{n}$$

the R.V. that describes the # of hats returned correctly is

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

$$E(X) = E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

$$= \frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n} = \frac{n}{n} = 1$$

i.e., the expectation of a Binomial R.V. X that models $n \geq 1$ trials w/ probability p success each has expectation:

$$E(X) = np$$

recall: $p(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

$$\therefore \quad E(X) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = np$$

e.g., if we manufacture 10,000 widgets, each w/ a 0.01% chance of introducing a manufacturing defect, how many defective widgets will we have on average?

$$10{,}000 \cdot 0.01\% = 1$$

e.g., if we roll 3000 6-sided dice, what is the number of 3's we expect to see, on average, if we cannot assume the rolls are mutually independent?

$p = \frac{1}{6}$

$np = \frac{3000}{6} = 500$ — linearity of expectation doesn't assume independence!

Average case computational complexity of an algorithm can be found by computing the expectation of the R.V. X, where:

- the sample space of X are its possible inputs $i_0, i_1, \ldots i_n$, and
- X assigns to each input the # of operations carried out by the algorithm for that input.

we just need to assign a probability to each input, and

$$E(X) = \sum_{j=0}^{n} p(i_j) X(i_j)$$

e.g. insertion sort:

unsorted

sorted

4 | 2 3 7 5 1 6

insert

2 4 | 3 7 5 1 6

2 3 4 | 7 5 1 6

2 3 4 7 | 5 1 6

2 3 4 5 7 | 1 6

1 2 3 4 5 7 | 6

1 2 3 1 5 6 7 |

let X = # of comparisons needed to sort a list of $a_1, a_2, \dots, a_n$ elements.

let X_i = # of comparisons needed to insert a_i into sorted list of $a_1, a_2, \dots, a_{i-1}$

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

$$E(X) = E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$ (by lin. of exp)

$$E(X_i) = \,?$$ — assuming random list $= \sum_{k=1}^{i} k \cdot p(X_i = k) = \sum_{k=1}^{i} k \cdot \frac{1}{i} = \frac{\cancel{i}(i+1)}{2} \cdot \frac{1}{\cancel{i}}$

e.g. $a_1 \quad a_2 \quad a_3 \quad a_4 \mid a_5$ # comparisons = 1, 2, 3, 4, 4

can end up anywhere

$$E(X) = \sum_{i=2}^{n} E(X_i) = \sum_{i=2}^{n} \frac{i+1}{2} = \frac{1}{2} \sum_{j=3}^{n+1} j = \frac{n^2 + 3n - 4}{4} \in \Theta(n^2)$$