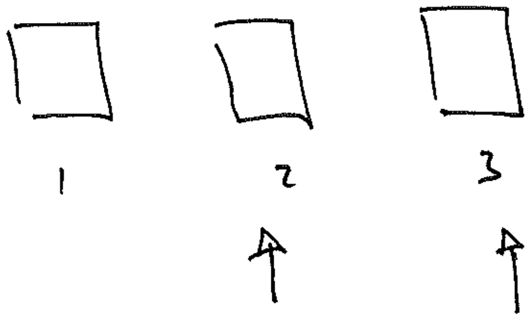
Monty Hall Problem

Simpson's Paradox



car + 2 goats

I chose door 2.

host opens a door, selected so that a goat is revealed.

Do you stick with your currently chosen door, or switch to the other closed door?

or does it make no difference?

solve this via 1/ LOTP
2/ Bayes' Thm
3/ Counting

UC Berkely Grad School Admissions.

\# applying   \# accepted.

men

women.

women accepted at a lower rate than men.

---

look at acceptance rate by department.

for most departments
women accepted at a higher rate than men.

---

why does the direction flip?

Gambler's Ruin.

2 gamblers, A and B, play a sequence of rounds.

They bet \$1 on each round.

$$p = P(\text{A wins a given round})$$

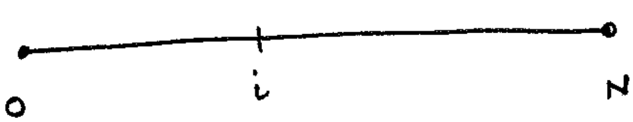
$$q = P(\text{B wins a given round}) = 1 - p$$

They continue playing until one player runs out of money.

What's the probability that A wins the game?

Assume: A starts with \$$i$

B \$$(N-i)$



~~Mov~~ Random walk on the line segment.

Move to the right with prob. p

left q.

0, N are "absorbing states"

$$\frac{d^2x}{dt^2} + 4\frac{dx}{dt} + 7 = 2\sin t$$

Guess + test.

$x^2 + 4x + 7$ ← find the roots
+ use them to
inspire our guesses.

e^{-at} ← solutions based
on the roots

Recursive structure.

1^{st} round. A wins or A loses.

A wins : same problem, but starting from $i+1$ instead of i

A loses : $i-1$ i

$P($A wins the game | A starts with \$$i) = P_i$

$$P_i = p * P_{i+1} + q * P_{i-1}$$

chance of winning with \$$i$

prob. of winning this round.

prob. of winning with \$$(i+1)$

prob of losing this round

prob of ~~losing~~ winning ~~starting~~ with \$$(i-1)$

Boundary conditions $P_0 = 0$

$P_N = 1$

this is a <u>difference equation.</u>

z - transform

$$P_i = p \times P_{i+1} + q \times P_{i-1}$$

$$p \times P_{i+1} - P_i + q \times P_{i-1} = 0$$

$$z^2 p - z + q = 0$$

$$z = \frac{1 \pm \sqrt{1 - 4pq}}{2p}$$

$$z = 1, \cancel{\frac{p}{q}} \; \frac{q}{p}$$

Try solutions of the form $P_i = z^i$

$$P_i = A \times 1^i + B \times \left(\frac{q}{p}\right)^i \qquad q \neq p$$

Boundary conditions

$$P_0 = 0 \qquad 0 = A + B$$

$$P_N = 1 \qquad 1 = A + B\left(\frac{q}{p}\right)^N$$

$$1 = A - A\left(\frac{q}{p}\right)^N$$

$$1 = A\left(1 - \left(\frac{q}{p}\right)^N\right)$$

$$A = \cancel{\frac{1}{\left(1 - \left(\frac{q}{p}\right)\right)^N}}$$

$$A = \frac{1}{1 - \left(\frac{q}{p}\right)^N}$$

$$p_i = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N} \quad \text{for } p \neq q.$$

limiting cases.

$p \to 1 \qquad p_i \to 1 \qquad$ (A wins each round)

$p \to 0 \qquad p_i \to 0$

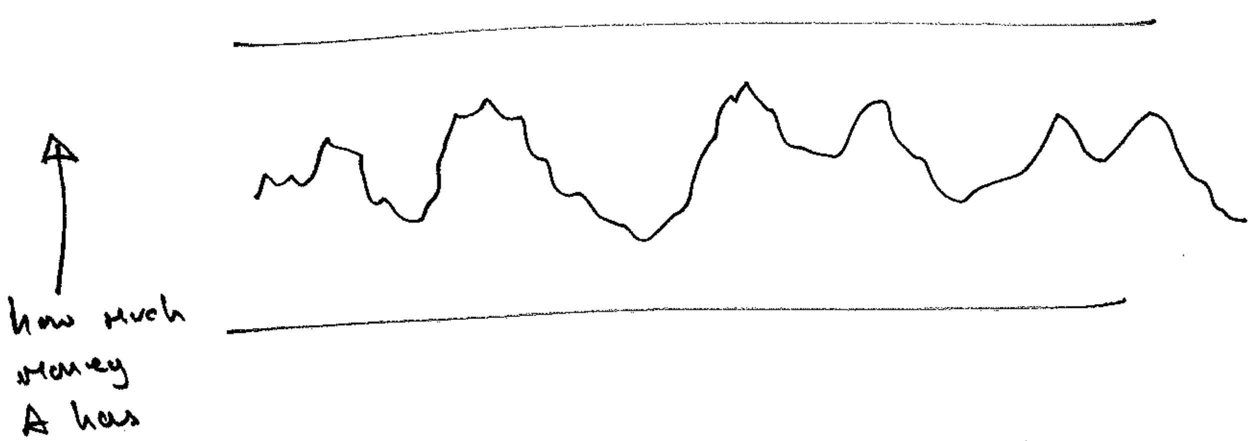
"almost fair" $\qquad p = 0.49.$

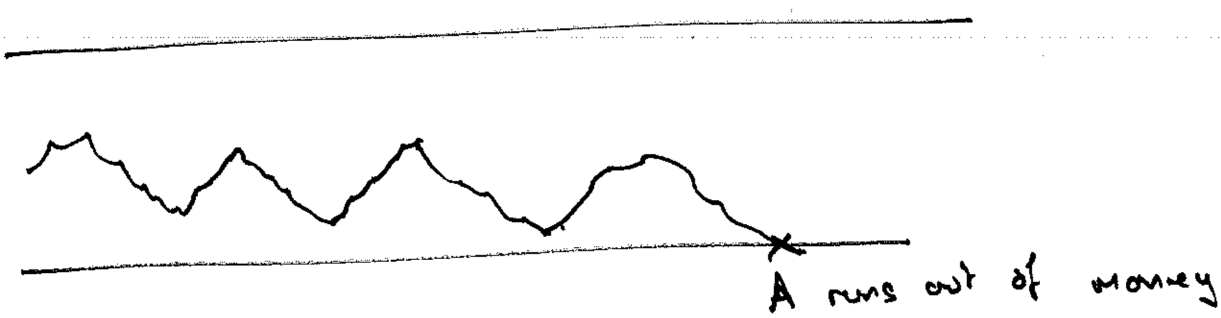
$$i = N - i$$

$N = 20 \qquad p_i = 0.4 \qquad$ (A wins with prob 0.4)

$= 100 \qquad p_i = 0.12$

$= 200 \qquad p_i = 0.02$





what happens when $i \ll N$.

– A loses.

$p = q$. → the game is <u>fair</u>

when the roots of the characteristic equation are not distinct, the solution is of the form

$$P_i = A + B \times i \times z^i$$

$$P_i = A \times 1^i + B \times i \times 1^i$$

when $p = q$ the roots of the eq^n are 1, 1

Put in the boundary conditions

$$P_i = \frac{i}{N}$$

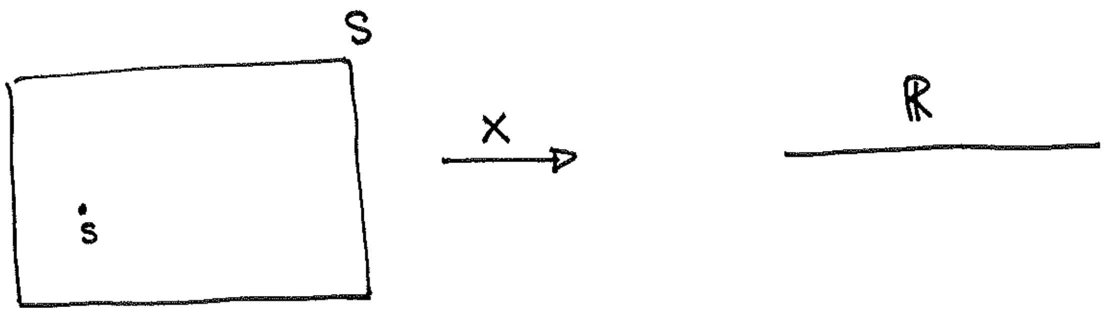
$$P(A \text{ wins}) = \frac{i}{N}$$

$$P(B \text{ wins}) = \frac{N-i}{N}$$

$$P(A \text{ wins}) + P(B \text{ wins}) = 1.$$

Random Variables.

What is a Random Variable.



RV is a function from the sample space to the real line.

~ numerical summary of an aspect of the experiment.

Randomness is in the experiment
(which subset of S is chosen)

After the experiment has happened,
we have a specific outcome, s, and
the RV maps <u>that</u> to a real number.

X $\qquad$ x

Random variable $\qquad$ particular realization

Bernoulli Random Variable

A R.V. X is said to have the Bernoulli Distribution if X has two possible values, 0 and 1

and

$$P(\underbrace{X = 1}_{\text{event}}) = p \quad \text{and} \quad P(X = 0) = 1 - p$$

$$\{ s \in S : X(s) = 1 \}$$

$$X \sim \text{Bernoulli}(p)$$

" RV X has Bernoulli distribution with parameter p "

Binomial Distribution

Binomial (n, p)

The distribution of the number of successes, X, in n independent Bernoulli (p) trials is called a Binomial (n, p), and its distribution is given by

$$\underline{P(X = k)} = \binom{n}{k} p^k (1-p)^{n-k}$$

k · integers $0 \dots n$

specification of the

probabilities associated with the RV

Probability Mass Function (PMF)

$Bin(n, p)$ $\quad X \sim Bin(n, p)$

n independent Bernoulli (p) trials

$Bin(n,p)$ # successes.

$$X = X_1 + X_2 + X_3 + \dots X_n$$

where $X_i = 1$ if i^{th} trial is success

0 failure.

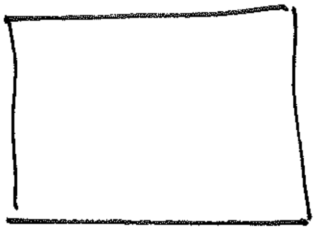
indicator RV

$X_i \sim$ Bernoulli (p)

$X_1 \dots X_n$ iid Bernoulli (p).

independent and identically distributed

S



RV assigns a number to each element of S.

RV $\neq$ Distribution

$n > 5$

$$\left.\begin{array}{ccccc} 0&0&0&0&0\\ 0&0&0&0&1\\ 0&0&0&1&0\\ 0&0&0&1&1\\ &&\vdots&&\\ 1&1&1&1&1 \end{array}\right\} \quad X \longrightarrow \quad \begin{array}{c} 0\\ 1\\ 1\\ 2\\ \vdots\\ 5 \end{array}$$

RV maps elements of the state space to numbers

Distribution assigns probabilities to events.

$X =$ some value. is an event.

Cumulative Distribution Function

$X \leq x$ ← event.

$F(x) = P(X \leq x)$ — this also assigns probabilities to all possible events.

another way to write the distribution that can be easier to work with in some situations.