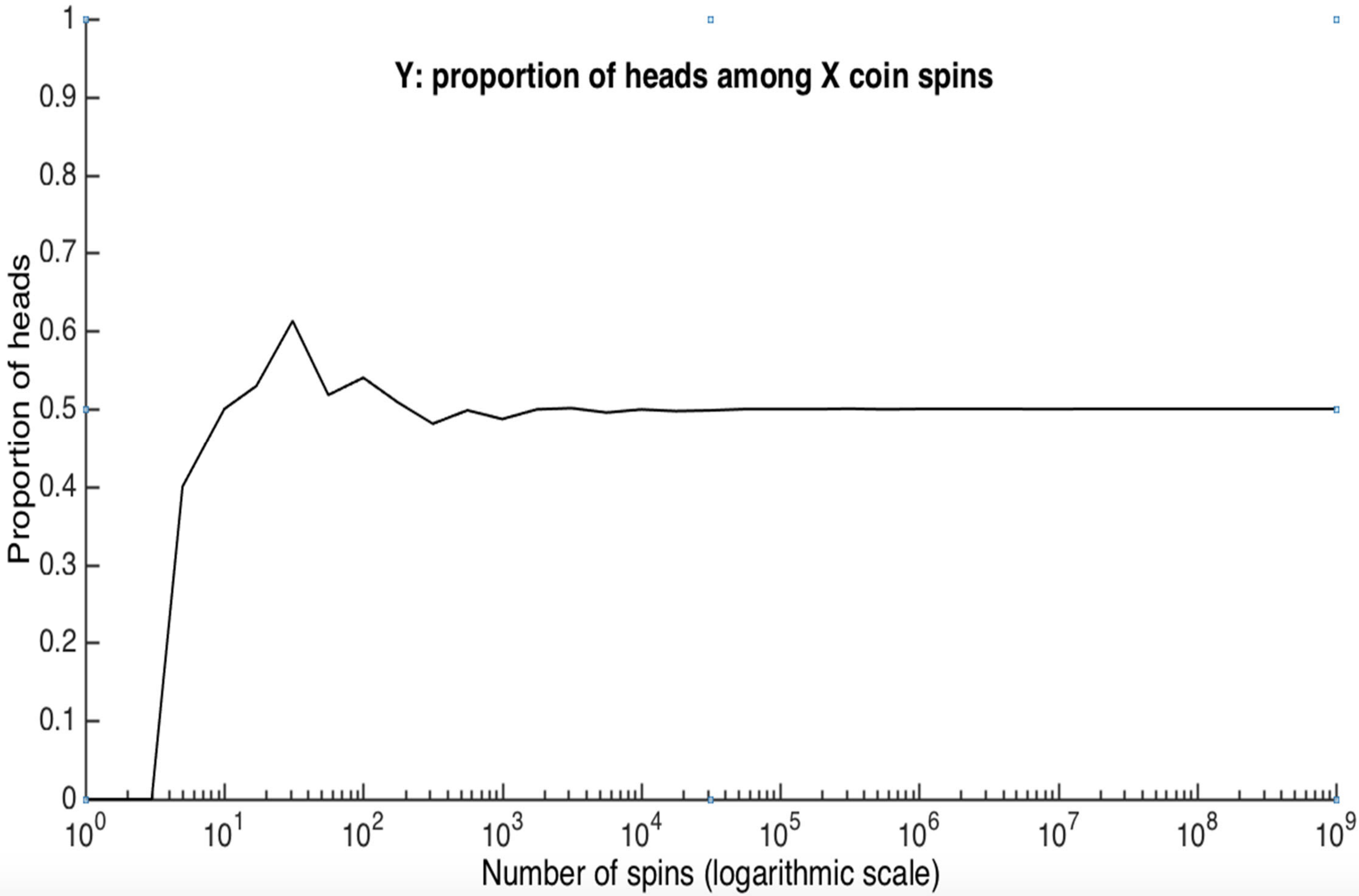
Matlab group exercise

Each table to edit the file coin_toss_template.m (replace all ?? with commands/variables/operations) or writes a new Matlab (Python, R, or anything else) script to:

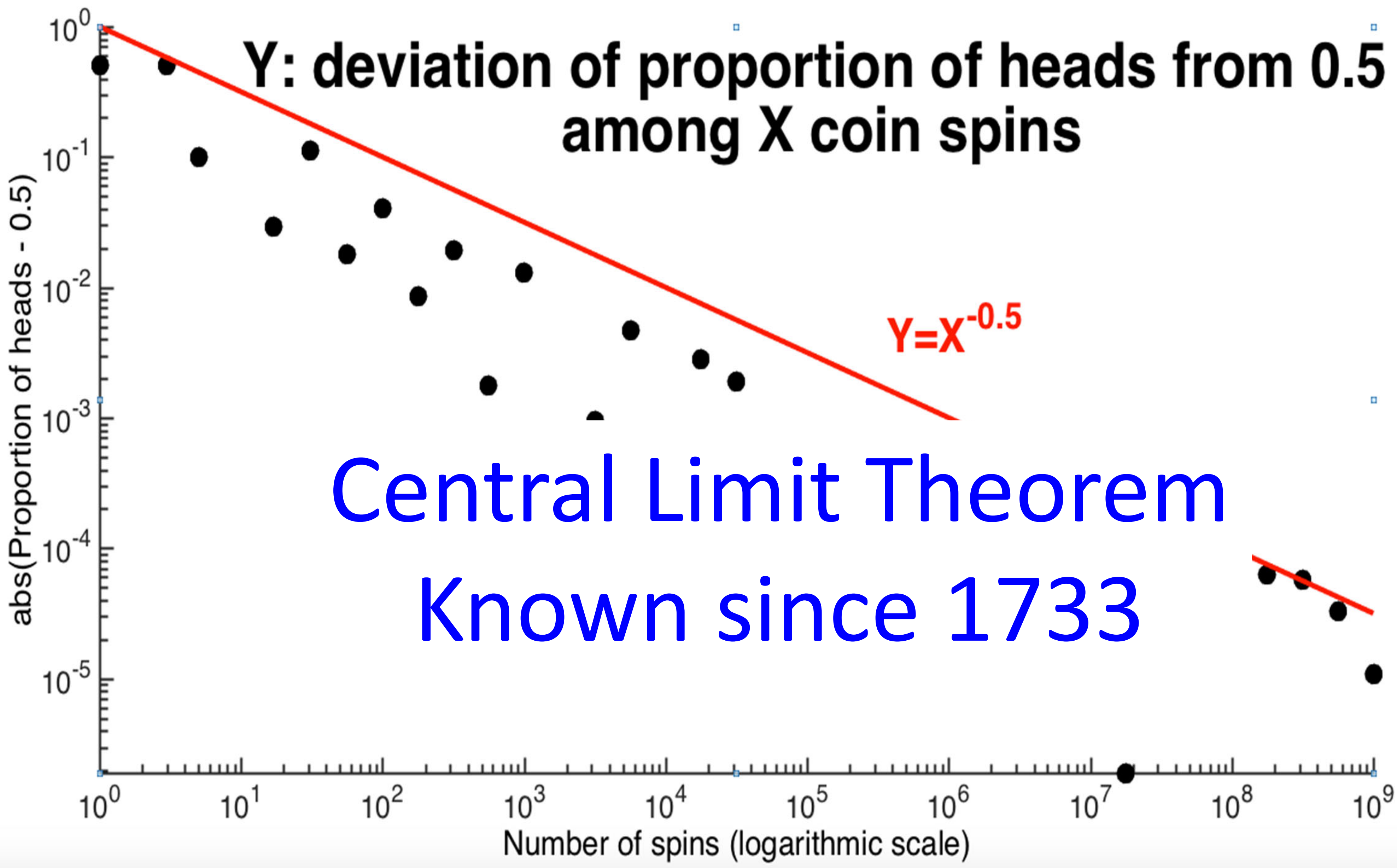
- Simulate a fair coin toss experiment
- Generate multiple tosses of a fair coin:
  1 – heads, 0 - tails
- Calculate the fraction of heads (f_heads(t)) at timepoints:
  t=10; 100; 1000; 10,000; 100,000; 1,000,000;10,000,000 coin tosses
- Plot fraction of heads f_heads(t) vs t with a logarithmic t-axis
- Plot abs(f_heads(t)-0.5) vs t on a log-log plot
  (both axes are logarithmic)

How I did it

- Stats=1e7;
- r0=rand(Stats,1); r1=floor(2.*r0);
- n_heads(1)=r1(1);
- for t=2:Stats; n_heads(t)=n_heads(t-1)+r1(t); end;
- tp=[1, 10,100,1000, 10000, 100000, 1000000, 10000000]
- np=n_heads(tp); fp=np./tp
- figure; semilogx(tp,fp,'ko-');
- hold on; semilogx([1,10000000],[0.5,0.5],'r--');
- figure; loglog(tp,abs(fp-0.5),'ko-');
- hold on; loglog(tp,0.5./sqrt(tp),'r--');



Proportion of heads among 1,000,000,000 coin tosses
(10^5 more than Kerrich) took me 33 seconds on my Surface Book



ABS(Proportion of heads-0.5)
among 100,000,000 coin tosses

Definitions of Probability

Two definitions of probability

- (1) STATISTICAL PROBABILITY: the relative frequency with which an event occurs in the long run

- (2) INDUCTIVE PROBABILITY: the degree of belief which it is reasonable to place in a proposition on given evidence

Bulmer, M. G.. Principles of Statistics (Dover Books on Mathematics)

Inductive Probability

An inductive probability of an event the degree of belief which it is rational to place in a hypothesis or proposition on given evidence.

Logical

Principle of indifference

- Principle of Indifference states that two events are equally probable if we have no reason to suppose that one of them will happen rather than the other. (Laplace, 1814)

Pierre-Simon, marquis de Laplace
(1749 –1827)
French mathematician, physicist, astronomer

- Unbiased coin:

probability Heads = probability Tails = ½

- Symmetric die:

probability of each side = 1/6

Inductive = Naïve probability

- If space S is finite and all outcomes are equally likely, then

$$\text{Prob(Event E)} = \frac{\text{\# of outcomes in E}}{\text{\# of all outcomes in S}}$$

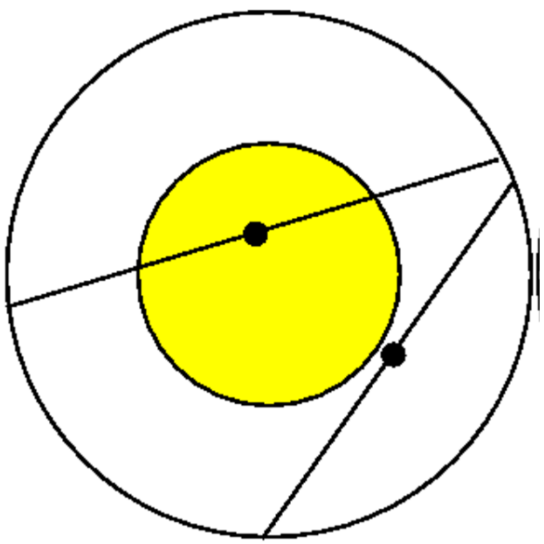
- Can also work with continuous is # is replaced with Area or Volume
- Unbiased coin: Prob(Heads) = Prob(Tails) = 1/2
- Symmetric die: probability of each side = 1/6
- Lottery outcomes are not symmetric: It is not a 50%-50% chance to win or loose in a lottery

Inductive probability can lead to trouble

- Glass contains a mixture of wine and water and proportion of water to wine can be anywhere between 1:1 and 2:1
- (i) We can argue that the proportion of water to wine is equally likely to lie between 1 and 1.5 as between 1.5 and 2.
- (ii) Consider now ratio of wine to water. It is between 0.5 and 1. Based on the same argument it is equally likely in [1/2, 3/4] as it is in [3/4, 1]. But then water to wine ratio is equally likely to lie between 1 and 4/3=1.333... as it is to lie between 1.333.. and 2. This is clearly inconsistent with the previous calculation...
- Paradox solved by clearly defining the experimental design:
  - For (i) use fixed amount of wine (1 liter) and select a uniformly-distributed random number between 1 and 2 for water.
  - For (ii) use 1 liter of water and select uniformly-distributed a random number between 0.5 and 1 for wine.
  - Different experiments – different answers
- Paradox is old. It is attributed to (among others) Joseph Bertrand

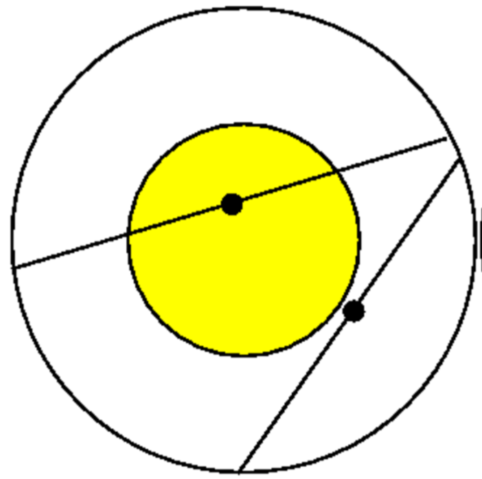
Better known Bertrand's paradox

- Take a circle of radius 2 and randomly draw a line segment through the circle. What is the probability P that the line intersects a concentric circle of radius 1?



Joseph Bertrand
(1822 –1900)
French mathematician

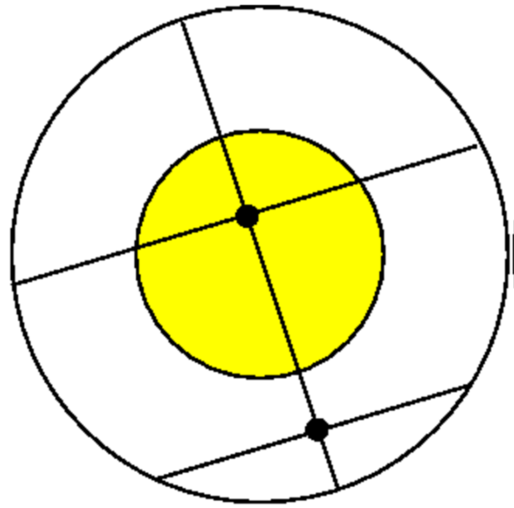
Solution #1



1. **Random point in 2D**: Each line has a unique midpoint, and a line will intersect the inner circle if its midpoint lies inside inner circle. Thus, P = probability that a randomly chosen midpoint lies in the inner circle:

$$P = \frac{\text{Area of the inner circle}}{\text{Area of the outer circle}} = \frac{\pi}{\pi 2^2} = \frac{1}{4}.$$

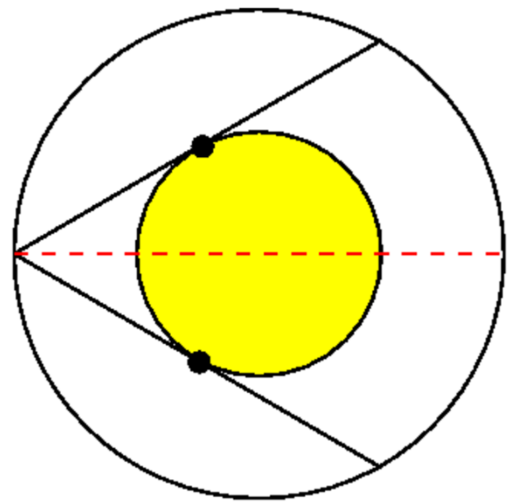
Solution #2



2. **Random point along the diameter**: Each line has a unique perpendicular bisector of length 4. So, P = probability that the midpoint lies on the inner part of the diameter:

$$P = \frac{\text{Length of the inner part of the diameter}}{\text{Length of the diameter}} = \frac{2}{4} = \boxed{\frac{1}{2}}.$$

Solution #3

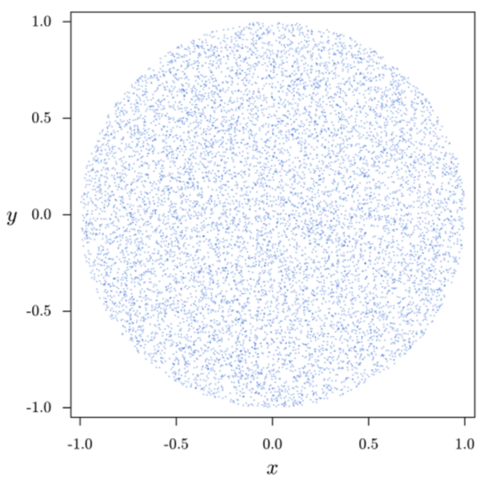


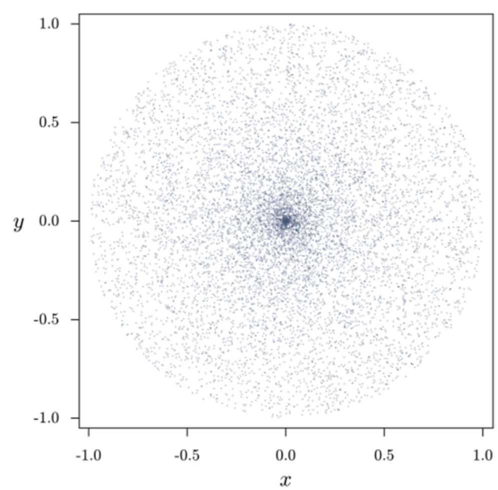
3. **Random angle**: Whether a line intersects the inner circle is determined by the angle it makes with the diameter intersecting the line on the outer circle:

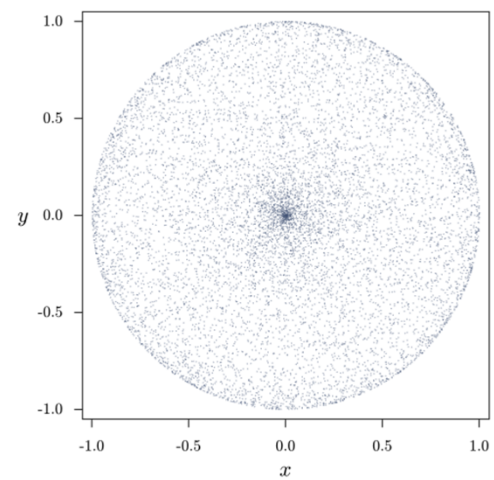
$$P = \frac{\pi/6}{\pi/2} = \frac{1}{3}.$$

So, is probability 1/4, 1/2 , or 1/3 ?

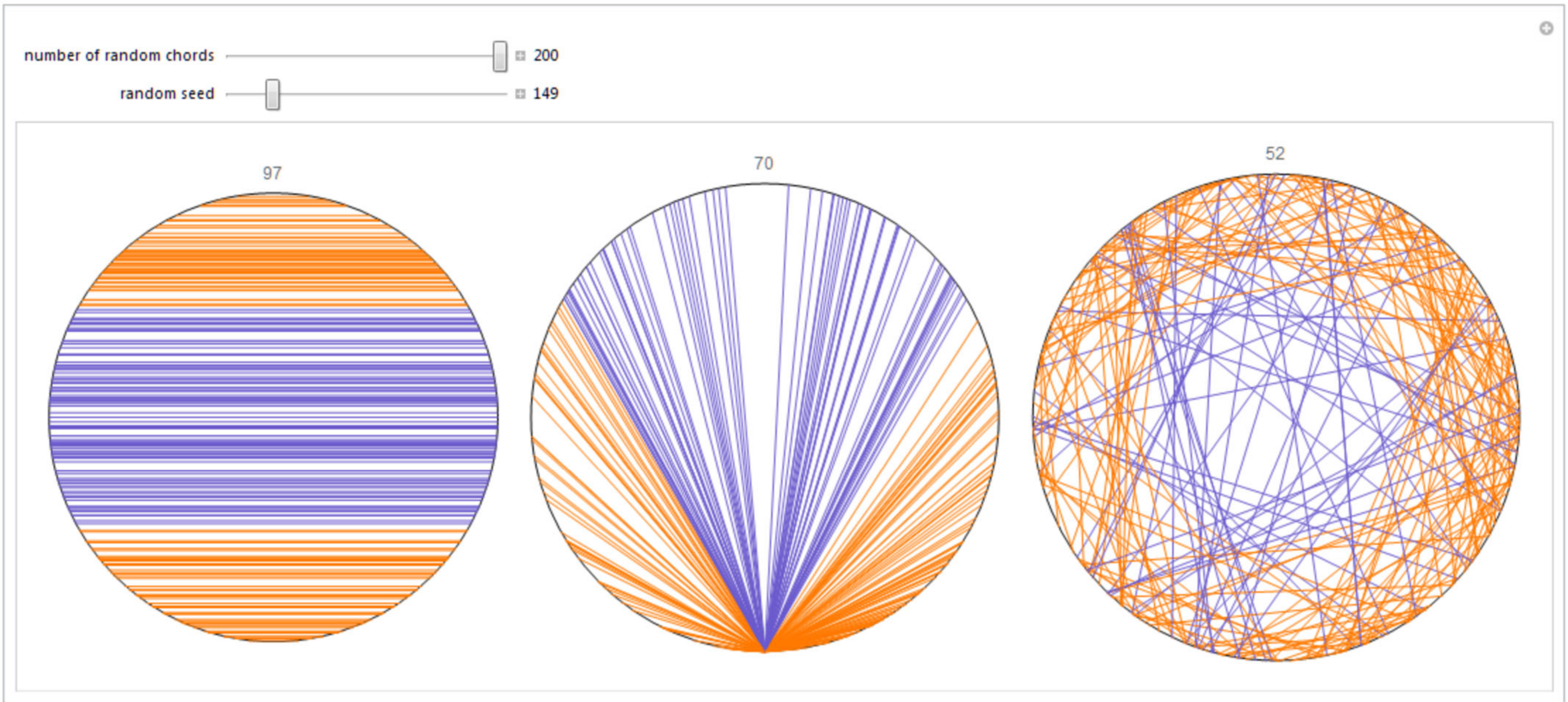
- Depends on how a “random” arc is selected:
  - For #1: select a point inside big circle and then draw an arc with this point as the center. Prob=1/4
  - For #2: select a diameter and a point on this diameter, then draw an arc. Prob=1/2
  - For #3: select a point on the circle and random angle. Prob=1/3







Mathematica visualization



BertrandRandomChordParadox.cdf

I have two children. One of them is a boy.

What is the probability I have two boys?

A. 1/2
B. 1/3
C. 2/3
D. 13/27
E. I don't know

Get your i-clickers

I have two children. One of them is a boy.

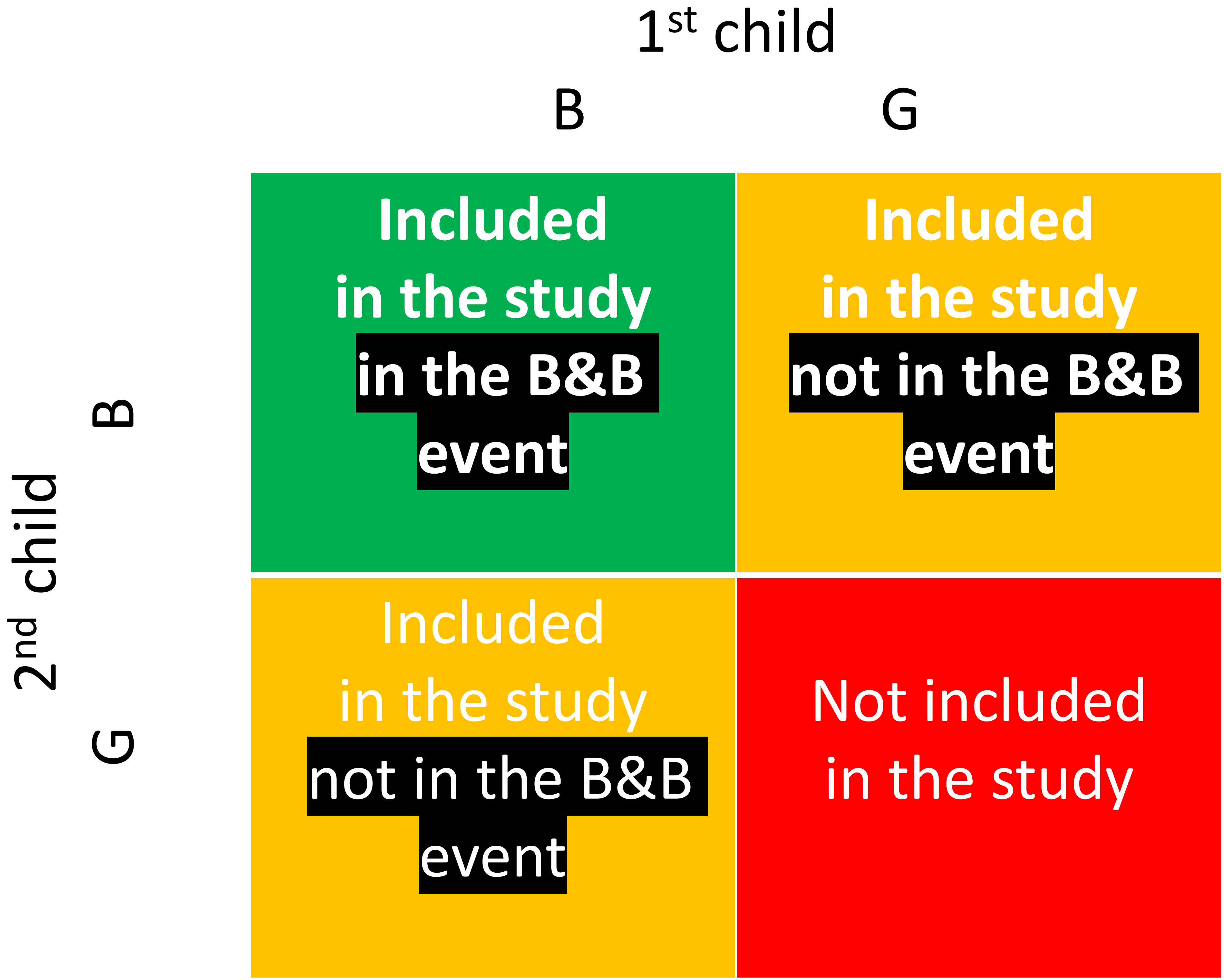
What is the probability I have two boys?

A. 1/2
B. 1/3
C. 2/3
D. 13/27
E. I don't know

Get your i-clickers

Solution

- Naïve answer: probability is 1/2
  - It would be correct if I told you that my first child was a boy, and I was asking for a probability that my second child would also be a boy
- Correct answer: probability is 1/2
  - Two children can come in four configurations: 1) boy/girl, 2) girl/boy, 3) boy/boy, 4) girl/girl. Since he has one boy, we are looking at the options 1, 2, or 3. Only the boy/boy combination includes two boys, so the probability is 1/3
- Consider doing an NIH-funded study:
  - recruit 1000 parents with two children
  - send ~250 parents with two girls straight home
  - Out of remaining ~750 parents ~250 (or 1/3 of the total) have two boys. The probability is 1/3


1st child
B
G
2nd child
B
G
Included in the study in the B&B event
Included in the study not in the B&B event
Included in the study not in the B&B event
Not included in the study

I have two children.

One of them is a boy born on Tuesday.

What is the probability I have two boys?

A. 1/2
B. 1/3
C. 2/3
D. 13/27
E. I don't know

Get your i-clickers

I have two children.

One of them is a boy born on Tuesday.

What is the probability I have two boys?

A. 1/2
B. 1/3
C. 2/3
D. 13/27
E. I don't know

Get your i-clickers

- 4x7x7=196 outcomes, 13+7+7=27 of which satisfy “Boy born on Tuesday”
- The probability of having two boys =13/(13+7+7)=13/27
- Close but not equal to 14/28=1/2

		Child 1: B						
		M	T	W	R	F	S	S
	M		1					
	T	2	3	4	5	6	7	8
Child 2: B	W		9					
	R		10					
13	F		11					
	S		12					
	S		13					

		Child 1: G						
		M	T	W	R	F	S	S
	M							
	T	1	2	3	4	5	6	7
Child 2: B	W							
	R							
7	F							
	S							
	S							

		Child 1: B						
		M	T	W	R	F	S	S
	M		8					
	T		9					
Child 2: G	W		10					
	R		11					
7	F		12					
	S		13					
	S		14					

		Child 1: G						
		M	T	W	R	F	S	S
	M							
	T							
Child 2: G	W							
	R							
	F							
	S							
	S							

Inductive probability relies on combinatorics or the art of counting combinations

Counting – Multiplication Rule

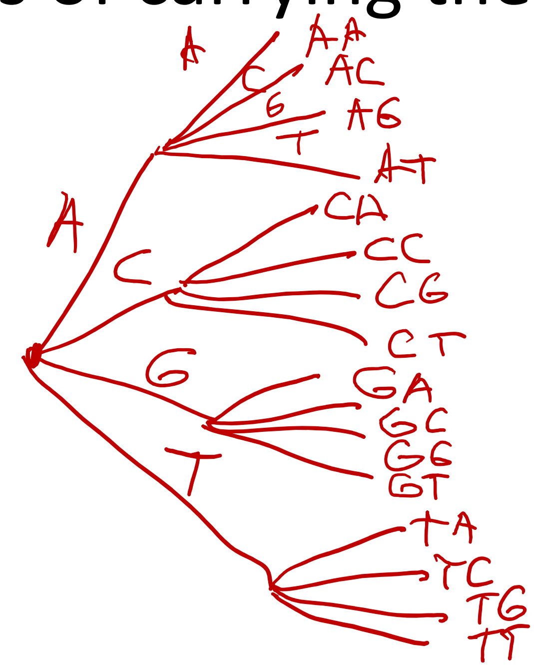
- Multiplication rule:
  - Let an operation consist of k steps and
    - n_1 ways of completing the step 1,
    - n_2 ways of completing the step 2, … and

    …….

    - n_k ways of completing the step k.
  - Then, the total number of ways of carrying the entire operation is:
    - $n_1 * n_2 * \ldots * n_k$

$n_1 = n_2 = 4$

Example: DNA 2-mer



- $S = \{A, C, G, T\}$ the set of 4 DNA bases
  - Number of k-mers is $4^k=4*4*4...*4$ (k –times)
  - There are $4^3=64$ triplets in the genetic code
  - There are only 20 amino acids (AA)+1 stop codon
  - There is redundancy: same AA coded by 1-3 codons
  - Evidence of natural selection: "silent" changes of bases are more common than AA changing ones
- A protein-coding part of the gene is typically 1000 bases long
  - There are $4^{1000}= 2^{2000} \sim 10^{600}$ possible sequences of just one gene
  - Or $(10^{600})^{25,000}=10^{15,000,000}$ of 25,000 human genes.
  - For comparison, the Universe has between 10^{78} and 10^{80} atoms and is $4*10^{17}$ seconds old.

Counting – Permutation Rule

- A permutation is a unique sequence of distinct items.
- If $S = \{a, b, c\}$, then there are 6 permutations
  - Namely: abc, acb, bac, bca, cab, cba (order matters)
- # of permutations for a set of n items is n!
- n! (factorial function) = n*(n-1)*(n-2)*…*2*1
- 7! = 7*6*5*4*3*2*1 = 5,040
- By definition: 0! = 1

A class has n students.
What is the smallest n so that there is 100% probability that there is a pair people with the same birthday e.g. May 1 (in any year)

A. 366
B. 367
C. 730
D. 32

Get your i-clickers

A class has n students.
What is the smallest n so that there is 50% probability that there is a pair people with the same birthday e.g. May 1 (in any year)

A. 734
B. 184
C. 5
D. 23

Get your i-clickers

Probability n people have different birthdays is:

$$\frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \cdots \frac{(365-n+1)}{365}$$

Let's find n when this is $\approx \frac{1}{2}$

$$\frac{1}{2} = \exp\left(\sum_{k=1}^{n} \log_{6}\left(1 - \frac{k-1}{365}\right)\right)$$

$$\log\left(1-\frac{k-1}{365}\right) \approx -\frac{k-1}{365}$$

$$\sum_{k=1}^{n} \log\left(1-\frac{k-1}{365}\right) \approx -\frac{(n-1)n}{2\cdot 365}$$

We need

$$\frac{1}{2} \approx \exp\left(-\frac{n(n-1)}{2\cdot 365}\right)$$

or

$$-\log 2 \approx -\frac{n^2}{2\cdot 365}$$

$$n \approx \sqrt{2\cdot 365\cdot \log 2} \approx 22.5$$

SOLUTION Because each person can celebrate his or her birthday on any one of 365 days, there are a total of $(365)^n$ possible outcomes. (We are ignoring the possibility of someone having been born on February 29.) Furthermore, there are $(365)(364)(363)\cdot(365-n+1)$ possible outcomes that result in no two of the people having the same birthday. This is so because the first person could have any one of 365 birthdays, the next person any of the remaining 364 days, the next any of the remaining 363, and so on. Hence, assuming that each outcome is equally likely, we see that the desired probability is

$$\frac{(365)(364)(363)\cdots(365-n+1)}{(365)^n}$$

It is a rather surprising fact that when $n \geq 23$, this probability is less than $\frac{1}{2}$. That is, if there are 23 or more people in a room, then the probability that at least two of them have the same birthday exceeds $\frac{1}{2}$. Many people are initially surprised by this result, since 23 seems so small in relation to 365, the number of days of the year. However, every pair of individuals has probability $\frac{365}{(365)^2} = \frac{1}{365}$ of having the same birthday, and in a group of 23 people there are $\binom{23}{2} = 253$ different pairs of individuals. Looked at this way, the result no longer seems so surprising. ■

Let's check it on our class

When I point to you,
say your month and date of birth

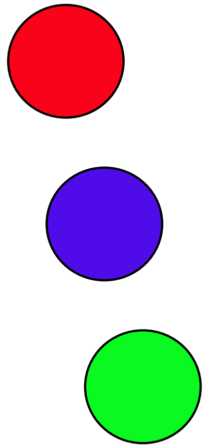
Multiplication and permutation rules are two examples of a general problem, where a sample of size k is drawn from a population of n distinct objects

Balls drawn from an urn (or bowl)

1 ball is red

1 ball is blue

1 ball is green

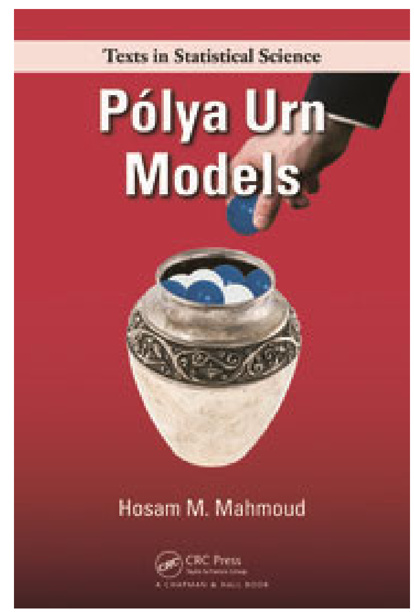


n=3 balls of different colors in an urn from which I draw k=2 balls one at a time

- Do I put each ball back to the bag after drawing it?
  - Yes: problem with replacement
  - No: problem without replacement
- Do I keep track of the order in which balls are drawn?
  - Yes: the order matters
  - No: the order does not matter

George Pólya

- George Pólya (December 13, 1887 – September 7, 1985) was a Hungarian mathematician. He was a professor of mathematics from 1914 to 1940 at ETH Zürich and from 1940 to 1953 at Stanford University. He made fundamental contributions to combinatorics, number theory, numerical analysis and probability theory.

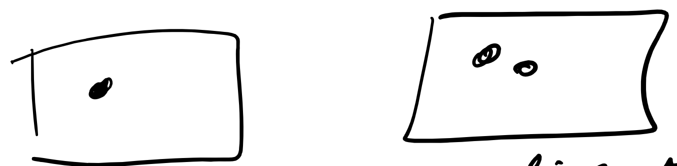


How many ways to choose a sample of K objects out of a population of n objects

	Order matters	Order does not matter
Replace	$n \times n \times n \times \ldots \times n$ $= n^K$	~~$\frac{n^K}{K!}$~~ not all objects are different
Do not replace	$n \times (n-1) \times (n-2) \times \ldots \times (n-K+1) =$ $= \frac{n!}{(n-K)!}$.	All objects are different $\rightarrow$ $\frac{n!}{(n-K)!} \times \frac{1}{K!} = \binom{n}{K}$

How to solve the problem of k out of n with replacement but where order does not matter?

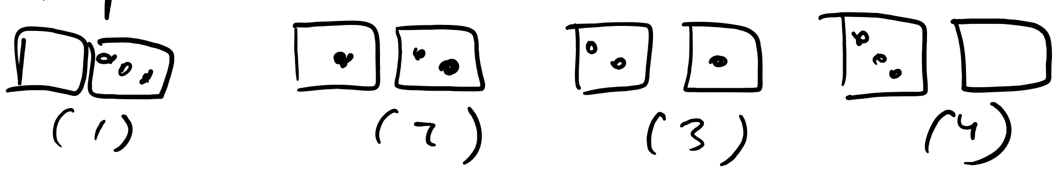
Let's solve $n=2$ problem first:

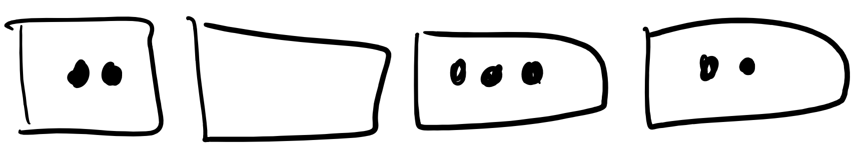


object 1 object 2

$k=3$

4 possibilities





$n=4, k=7$

$\bullet\bullet|||\bullet\bullet\bullet|\bullet\bullet$ k dots, $n-1$ box boundaries

$$\binom{k+n-1}{k} = \frac{(k+n-1)!}{k!\,(n-1)!}$$ ways to distribute

Sampling table

How many ways to choose a sample of k objects out of population of n objects?

	Order matters	Order does not matter
Replacement	$(n)^k$	Difficult: $\binom{n+k-1}{k} = \frac{(n+k-1)!}{(n-1)!k!}$
No replacement	$n(n-1)(n-2)..(n-k+1)= \\ = \frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{(n-k)!k!}$