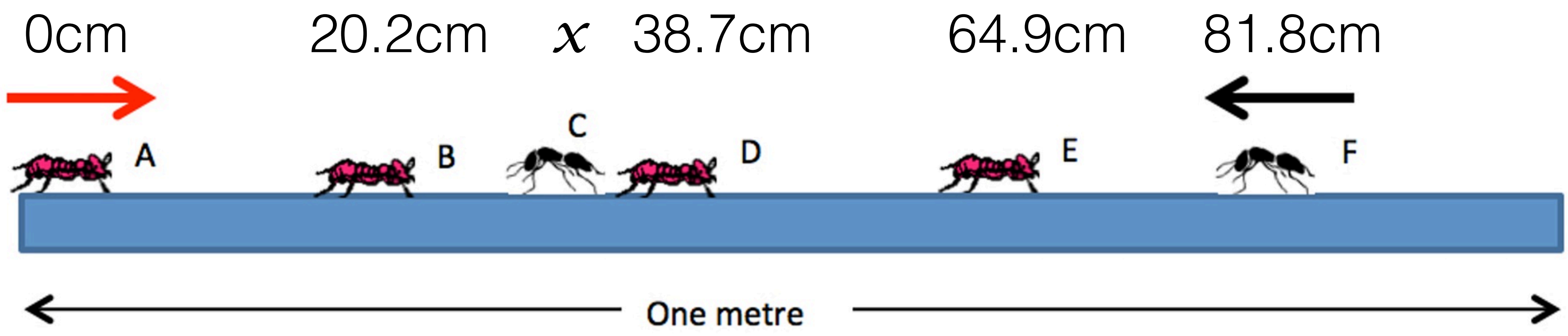
Ants on a stick



- Ants walk at 1cm/second. When they meet, each ant turns around and walks in the other direction. When they reach the end of the stick they fall off.
  - How many seconds until the last ant falls off?
  - Which ant is the last to fall off the stick?

Last week’s problems

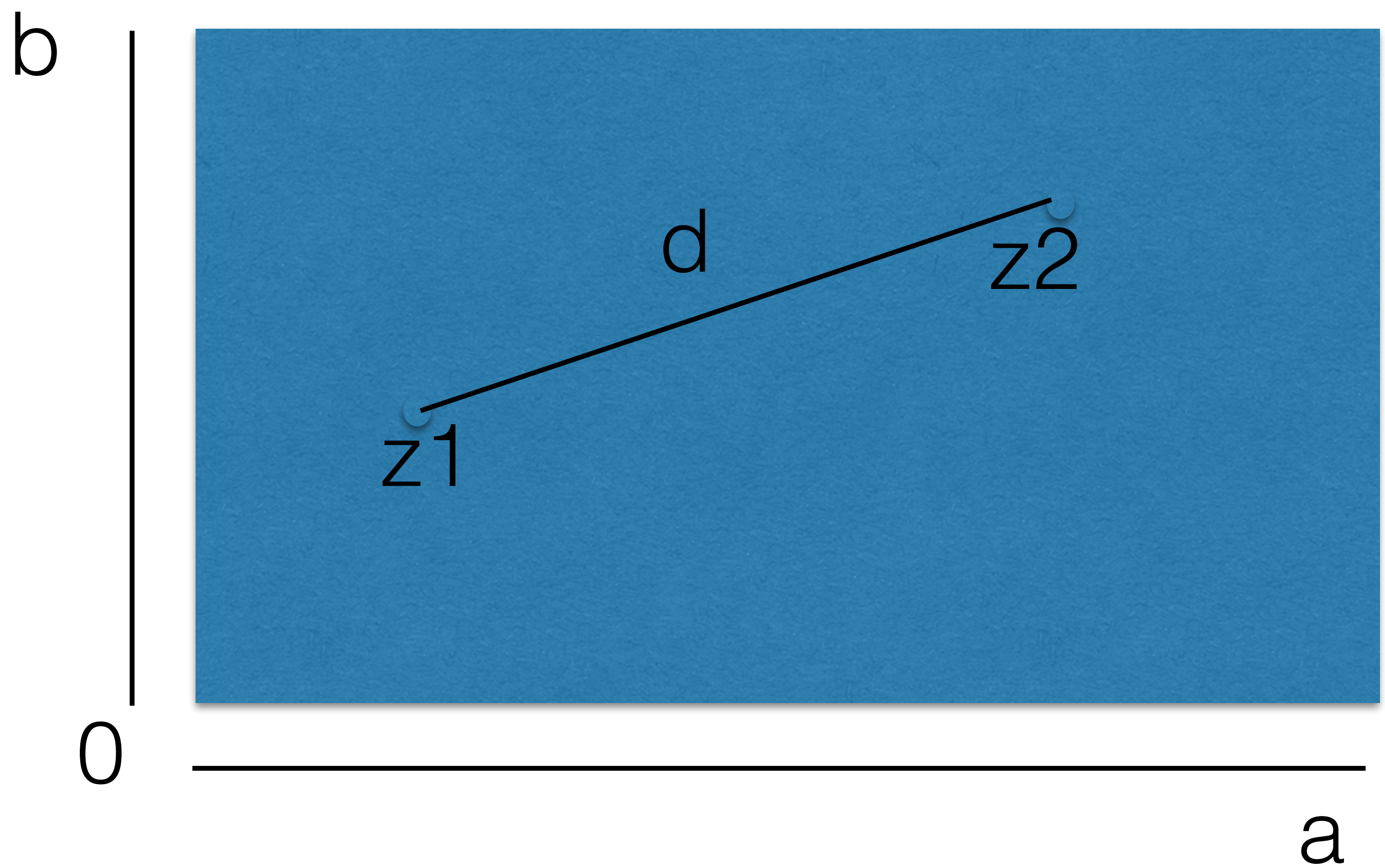
For feedback, etc., commit your changes to GitHub and then create an issue to let me know that it is ready for review.

Github organization

- The repos for Week n will all have names that start with “Weekn-“.
- We have a repository called “Useful-Tips” for sharing useful tricks that you learn.
- Github does not naturally render .html files. So if you upload those I won’t be able to see your output properly. It’s better to either knit to a pdf, which works fine, or knit to a .md (markdown) file. If you want the .md to render properly you also need to upload a folder that is created when you knit it and that contains the figures. This is explained in the “Useful-Tips” repo on our github page (or see Slack). I have uploaded an example (“week1-md-example" repo)

Problems from last week: Distance between points

- Suppose we have a rectangle [0,a]x[0,b]
- If we generate two points, z1 and z2, randomly in the rectangle, what is the expected distance, d, between them?



Questions

- How many simulations should you run? (The program will return an answer for any value of NumberOfSims that you give it.)
-

$$\mathrm{E(d)} = \frac{1}{15}\left[\frac{a^3}{b^2} + \frac{b^3}{a^2} + \sqrt{a^2+b^2}\left(3 - \frac{a^2}{b^2} - \frac{b^2}{a^2}\right)\right] + \frac{1}{6}\left[\frac{b^2}{a}\operatorname{arccosh}\left(\frac{\sqrt{a^2+b^2}}{b}\right) + \frac{a^2}{b}\operatorname{arccosh}\left(\frac{\sqrt{a^2+b^2}}{a}\right)\right]$$

where

$$\operatorname{arccosh}(t) = \log\left(t + \sqrt{t^2-1}\right)$$

Further questions

- For fixed area A, is the expected distance between two randomly chosen points, E(d), bigger for a square or a rectangle?

- Show that, for any fixed area A, and distance D, there exists a rectangle such that E(d)>D.

Monte Carlo Application 3: coin tossing

- Suppose we simulate a sequence of 500 coin tosses. Define a 'run' as a sequence of all Heads (and of all Tails). Answer the following:
  - What is the distribution of the number of heads?
  - What is the distribution of the length of the run starting at the first toss?
  - What is the expected number and distribution of the number of runs in total?
  - What is the expected value and distribution of the length of the longest run?

  - There was a bug in the provided code that caused the final histogram to look wrong!

MC Application 3: Answers

- Suppose we simulate a sequence of n coin tosses. Define a ‘run’ as a sequence of all Heads (or all Tails). Answer the following:
  - What is the distribution of the number of heads?
    - Binomial(n,1/2)
  - What is the distribution of the length of the run starting at the first toss?
    - Geometric(0.5)
  - What is the expected number and distribution of the number of runs in total?
    - 1+Binomial(n-1,1/2)
  - What is the expected value and distribution of the length of the longest run?
    - ?

Examinable project1: Randomization tests - golf balls

- Allan Rossman used to live along a golf course and collected the golf balls that landed in his yard. Most of these golf balls had a number on them.

- Question: What is the distribution of these numbers?
- In particular, are the numbers 1, 2, 3, and 4 equally likely?

[Originally due to Allan Rossman - via Randall Pruim]

Examinable project 1: golf balls

- Population: Golf balls at that driving range
- Allan tallied the numbers on the first 500 golf balls that landed in his yard one summer.
- Sample: Those golf balls driven ~150 yards and sliced.

1	2	3	4
137	138	107	104

486 balls in total

- There were 14 “others”, which we will ignore
- Question: What is the distribution of these numbers? In particular, are the numbers 1, 2, 3, and 4 equally likely?
- Meta-Question: How do we answer this question using the data?

Randomization test set-up

- Null hypothesis: Our default belief about the data.
  - Here, it is that the numbers on the population golf balls are uniformly distributed between 1 and 4.
- Test statistic: A single number that can be calculated from the data and used to test whether our null hypothesis is true.

1	2	3	4
137	138	107	104

486 balls in total

- What test-statistic should we use?
- How should we conduct the test **using simulations**?

Examinable Project - what to turn in

- A 3-5 page write-up, in the form of a knitted file (ideally a pdf or .md file) including the following:

- Explain the scenario you are asked to consider (golf balls)
- Explain the logic of a hypothesis test
  - 1. State Hypotheses
    - Null hypothesis that you must provide a model for (for simulations)
  - 2. Calculate a Test Statistic
  - 3. Determine the “p-value” - how do you do this via simulation?
  - 4. Interpret the p-value you get
- What makes a good test statistic? **Compare several**.
- Comment on things such as the following in a Conclusion section:
  - What would you do if you knew the sampling distribution of the test statistic? How does simulation help if you don’t?
  - Power against particular types of alternatives:
  - Does the best test statistic depend on what hypoth. you are testing?

Examinable Projects - what to turn in

- What to turn in (due in one week at midnight on Monday evening):
  - 1/2 page introduction - describe the idea of Monte Carlo simulation
    - 1-2 page of methods - describe the specific method you are using
    - 1-2 pages of results - show your results
    - 1/2 page of conclusions - summarize the results; what did you learn

- Submit both an Rmarkdown file and a ‘knitted’ file (i.e. a file that also shows the output).

Monte Carlo application 2: Hypercubes

- Consider n-dimensional unit ‘cubes’
  - n=3 -> a cube. Coordinates= (x_1,x_2,x_3) [or (x,y,z)].
  - n=2 -> a square. Coordinates= (x_1,x_2) [or (x,y)].
  - n=1 -> line. Coordinates= (x_1) [or (x)].

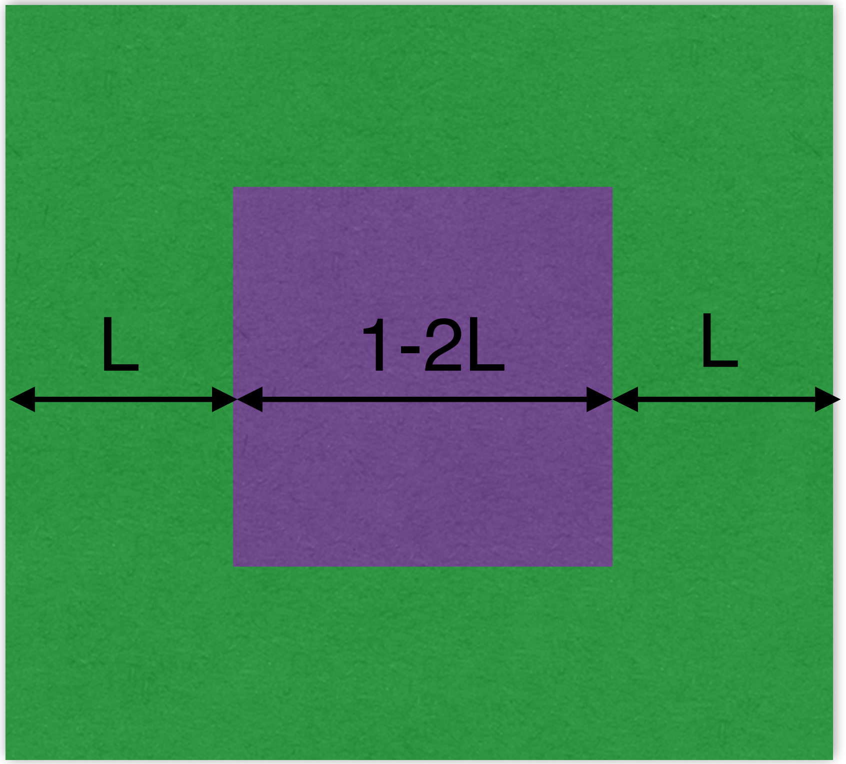
  - n=4 -> 4-dimensional hypercube. Coordinates= (x_1,x_2,x_3,x_4).
  - n=5 -> 5-dimensional hypercube. Coordinates= (x_1,x_2,x_3,x_4,x_5).
  - n=N -> N-dimensional hypercube. Coordinates= $(x_1,x_2,x_3,x_4,\ldots,x_N)$.

- Let’s suppose that all are unit cubes, so $0 \leq x_i \leq 1$, for all i.
- **Question: what proportion of the volume of an N-dimensional hypercube is within a distance of 0.1 of the surface?**

Question: If we sample a point ***uniformly at random*** from inside a hypercube, what is the probability that the point is within a distance of 0.1 of the surface?

Hypercubes

- Answer:
  - The proportion of the volume that is within a distance of L of the surface of an N-dimensional hypercube with sides of length 1 is …. $1-(1-2L)^N \rightarrow 1$ as $N \rightarrow \infty$ (for $0 \leq L \leq 0.5$).

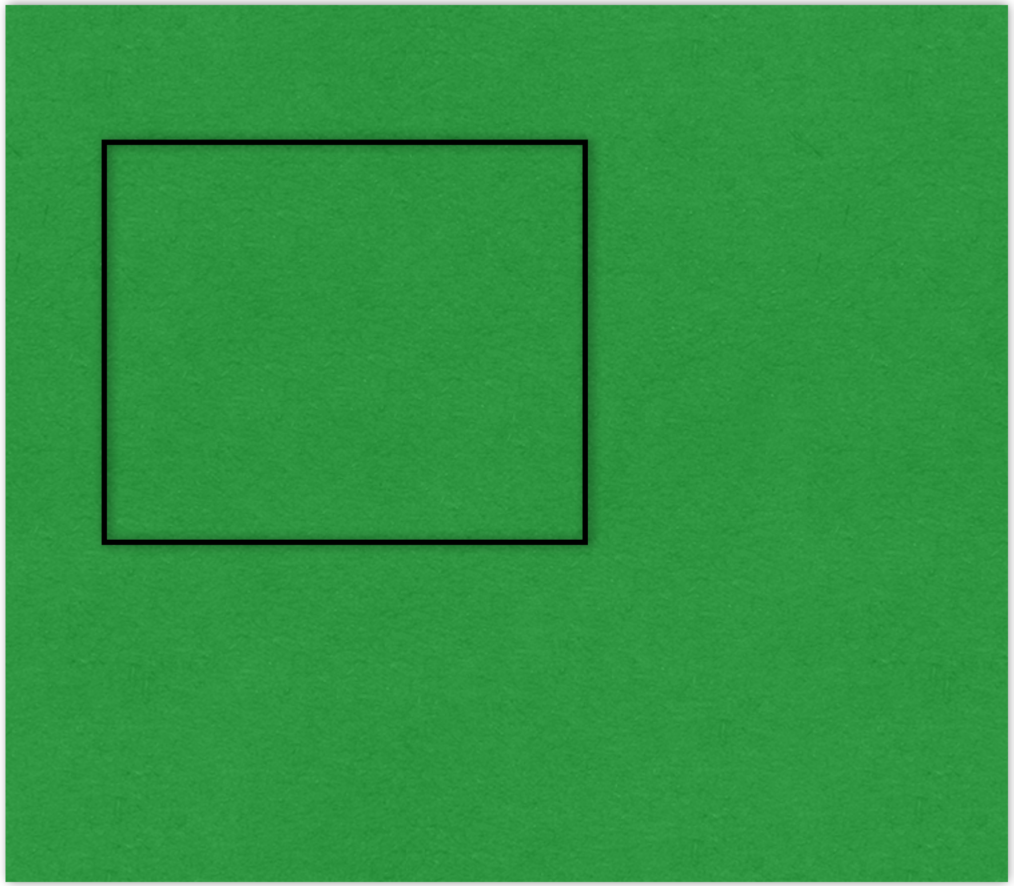


In-class exercise:

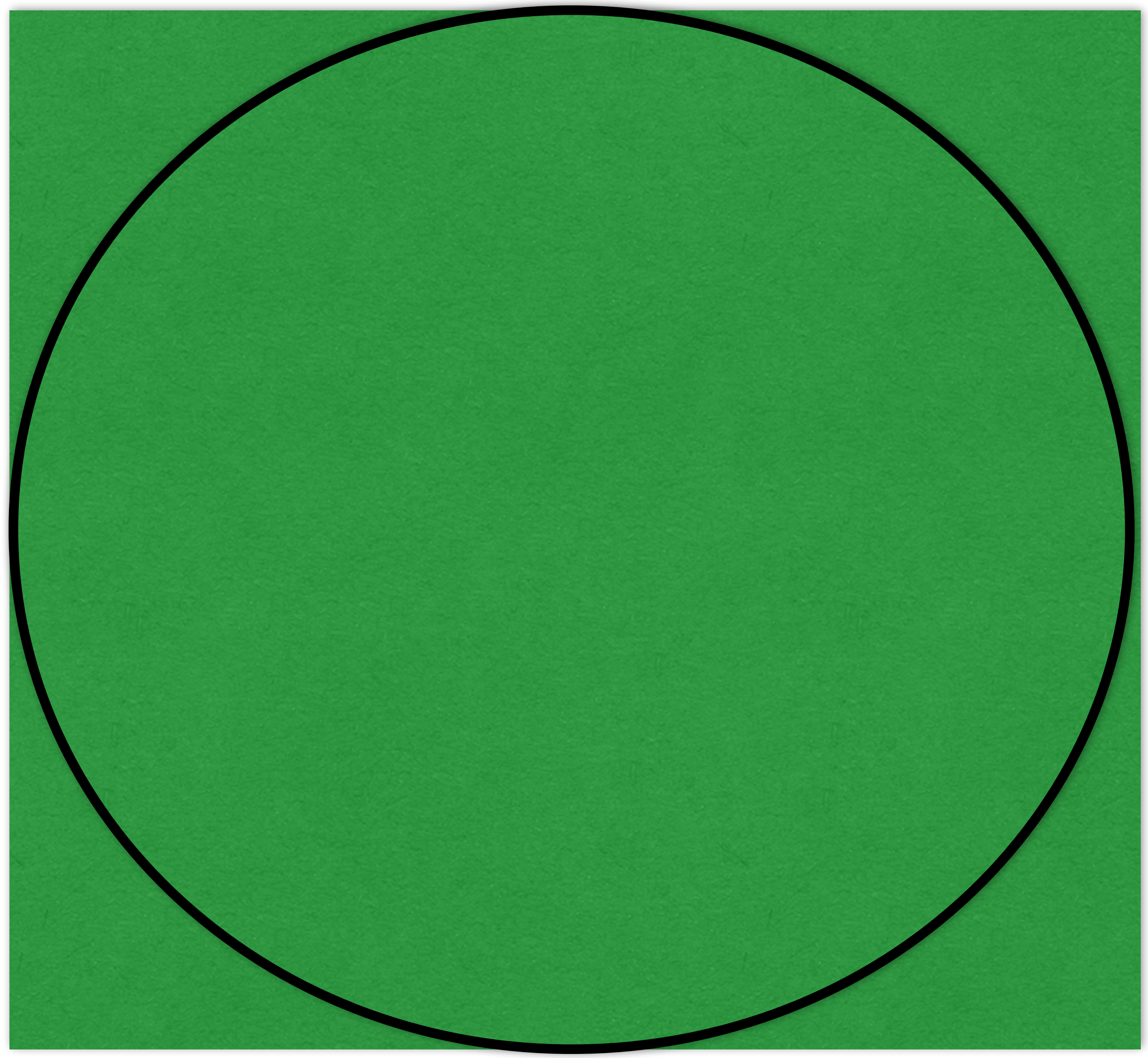
- Estimate (not calculate) π using Monte Carlo methods.
- Recall: The Monte Carlo estimate of the probability of an event is the proportion of times that the event happens when you repeat the experiment over and over (usually via computer simulation).
- So, you need to think of something you can simulate in which the probability of success depends upon π.

Hint

- Suppose we randomly pick a point in the green square
- What is the probability that the point lands inside the box area?



- Answer:
  - (area of box)/(total area of green square)



Jocky Wilson - Athlete, legend.

Jocky Wilson is the Citizen Kane of darts." - Guardian newspaper

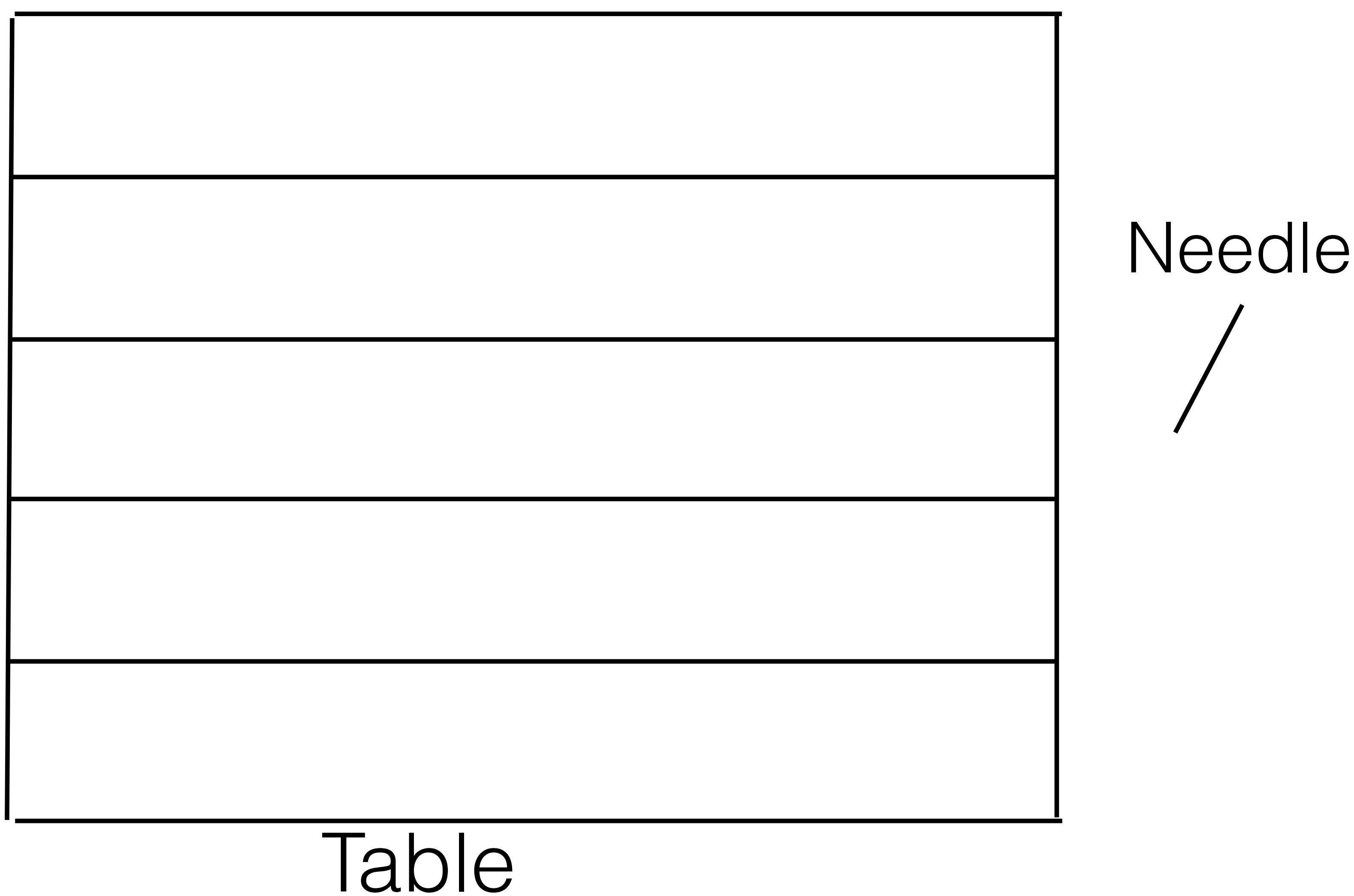
In 1989 he released a record "Jocky on the oche" but it failed to spark the public imagination and is reputed to have sold just 850 copies. (Wikipedia).

- As he mourned him yesterday, old rival and pal Bobby George called him “the king of darts”.

“In any case, a lot of Jocky’s diet was liquid. He liked to prepare for a game with “seven or eight vodkas, to keep my nerves”. Jocky also knocked back pints of lager during matches, which didn’t always help his game.

http://www.dailyrecord.co.uk/news/editors-choice/2012/03/26/pals-and-stars-honour-scots-darts-legend-jocky-wilson-after-he-dies-at-home-aged-62-86908-23801861/

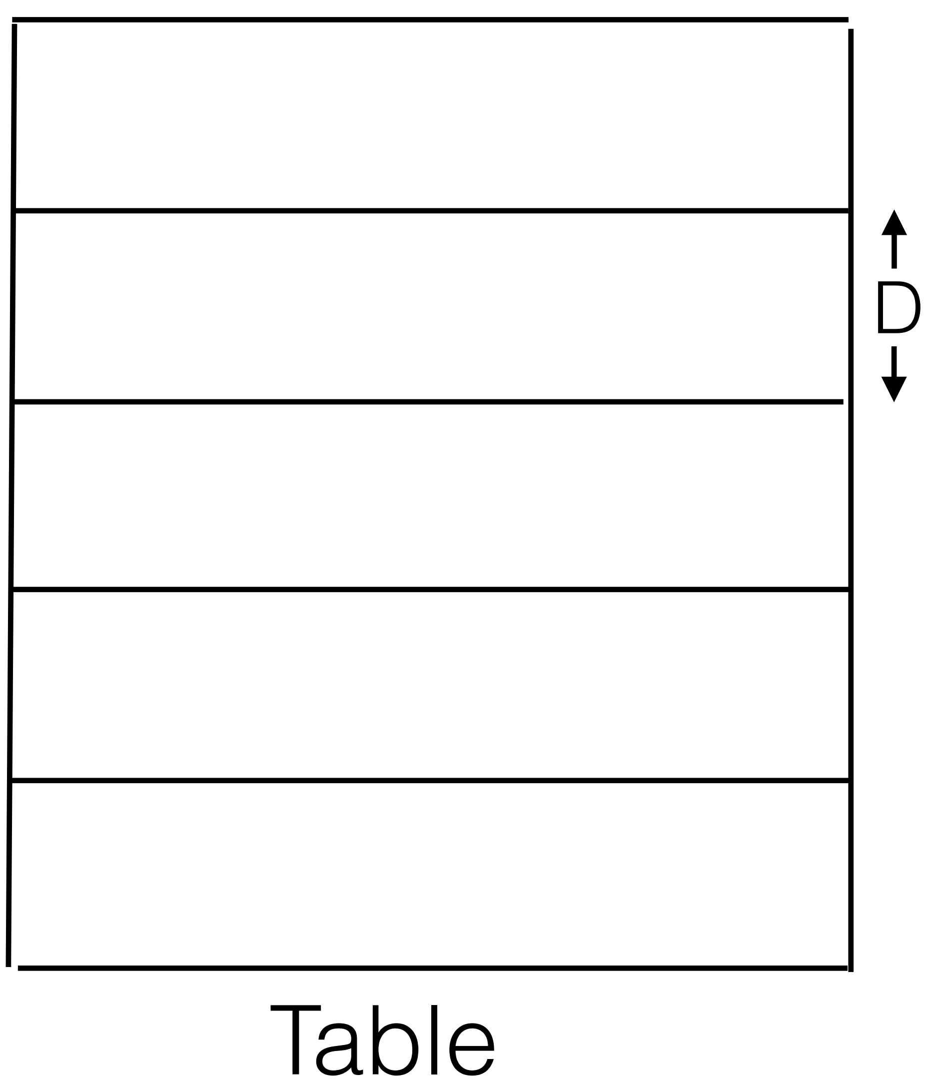
Monte Carlo Simulation: Calculating π - Buffon’s needle

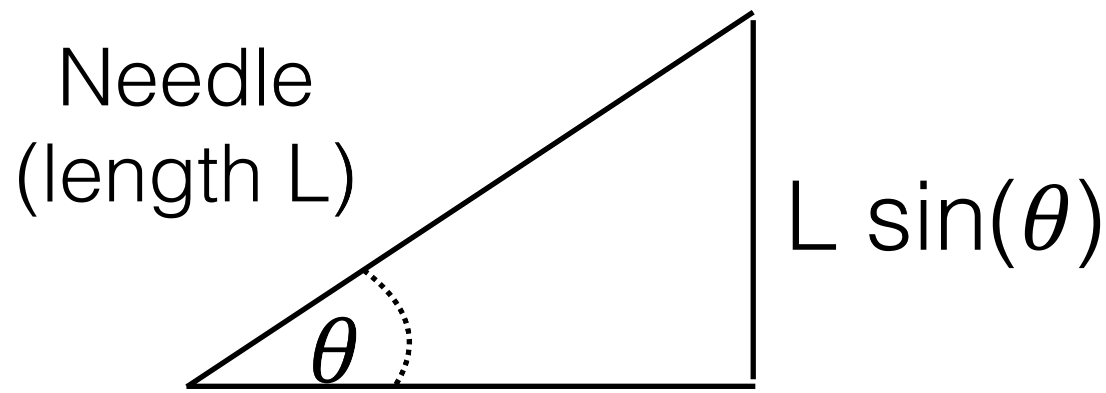


Georges-Louis Leclerc, Comte de Buffon
(French pronunciation: [ʒɔʁʒ lwi ləklɛʁ kɔ̃t də byfɔ̃]; 7 September 1707 – 16 April 1788)

- http://www.angelfire.com/wa/hurben/buff.html

Monte Carlo Simulation: Calculating π - Buffon's needle





So, P(cross line) = min(1,LSin(θ)/D)

If N= number of trials, lines are distance D=1 apart,
C = number of times needle crosses line,
and L≤D, then:
$C/N \sim \int L\sin(\theta)(1/\pi)\, d\pi$ [integral goes from 0 to π in radians]
So $C/N \sim L[-\cos(\pi)-(-\cos(0))]/\pi = 2L/\pi$
So $\pi \sim 2LN/C$

Results:

- Example:
  - Mario Lazzerini (1901):
    - Needle length: 2.5cm
    - Distance between lines: 3cm
    - 3408 throws
    - 1809 hits of the lines
    - Led to estimate of pi of 3.1415929

  - In order to guarantee accuracy to 6 decimal places would need 134 trillion needles.

- What is the optimal length of needle?

END