Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 1 Video 2-5-25: https://youtu.be/bSdLy3-BVvQ
Slides: https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Math 341: Probability
First Lecture

Steven J Miller
Williams College

sjm1@williams.edu

http://www.williams.edu/Mathematics/sjmiller/public_html/317

Introduction and Objectives

Introduction / Objectives

Probability theory: model the real world, predict likelihood of events.

One of the three most important quantitative classes (statistics, programming).

Introduction / Objectives

Probability theory: model the real world, predict likelihood of events.

One of the three most important quantitative classes (statistics, programming).

[redacted]

- Obviously learn probability.
- Emphasize techniques / asking the right questions.
- Model problems and analyze model.
- Elegant solutions vs brute force (parameters in closed form versus numerical solutions).
- Looking at equations and getting a sense: $\log-5$ Method: $\frac{p \pm pq}{p+q \pm 2pq}$.

Types of Problems

- Biology: will a species survive?
- Physics / Chemistry / Number Theory: Random Matrix Theory.
- Gambling: Double-plus-one.
- Economics: Stock market / economy.
- Finance: Monte Carlo integration.
- Marketing: Movie schedules. Cryptography: Markov Chain Monte Carlo. 8 ever 9 never (bridge).

$1, 2, \dots, 9$

$\frac{\square}{\square\square} + \frac{\square}{\square\square} + \frac{\square}{\square\square} = 1$

$9! \approx 330{,}000$

$6! \approx 1000$

$7 \cdot 8 \cdot 9 \approx 1000$

$1{,}000{,}000$

$500{,}000$

My (applied) experiences

- Marketing: parameters for linear programming (SilverScreener).
- Data integrity: detecting fraud with Benford's Law (IRS, Iranian elections).
- Sabermetrics: Pythagorean Won-Loss Theorem.

Course Mechanics

Grading / Administrative

- Fast Pace: Class Participation: 5%. HW: 15%. Writing Assigbment: 10% (Solutions to HW Problems from Chapters, Project in Probability / Statistics / Data Science). Midterm: 30% (if there are two exams only best counts). ‘Final’ exam: 40%. May do a project for 10% of your grade (reduces other categories proportionally).
- Pre-reqs: Calc III, basic combinatorics / set theory, linear algebra.

Office hours / feedback

- TBD and when I’m in my office (schedule online)

Other

- Webpage: numerous handouts, additional comments each day (mix of review and optional advanced material).
- Probability Lifesaver: opportunity to help write a solution key, lots of worked examples.
- Gather and analyze some data set of interest.
- PREPARE FOR CLASS! Must do readings before each class.

Being Prepared

Never know when an opportunity presents itself....

S. J. Miller at the Sarnak 61^{st} Dinner

(copyright C. J. Mozzochi, Princeton N.J)

Being Prepared

- Your Job:
  - Be prepared for class: do reading, think about material.
  - Come to me, the TAs and each other with questions.
- My/TAs Job:
  - Provide resources, guiding questions.
  - Be available.

Other: Advice from Jeff Miller

- Party less than the person next to you.

Other: Advice from Jeff Miller

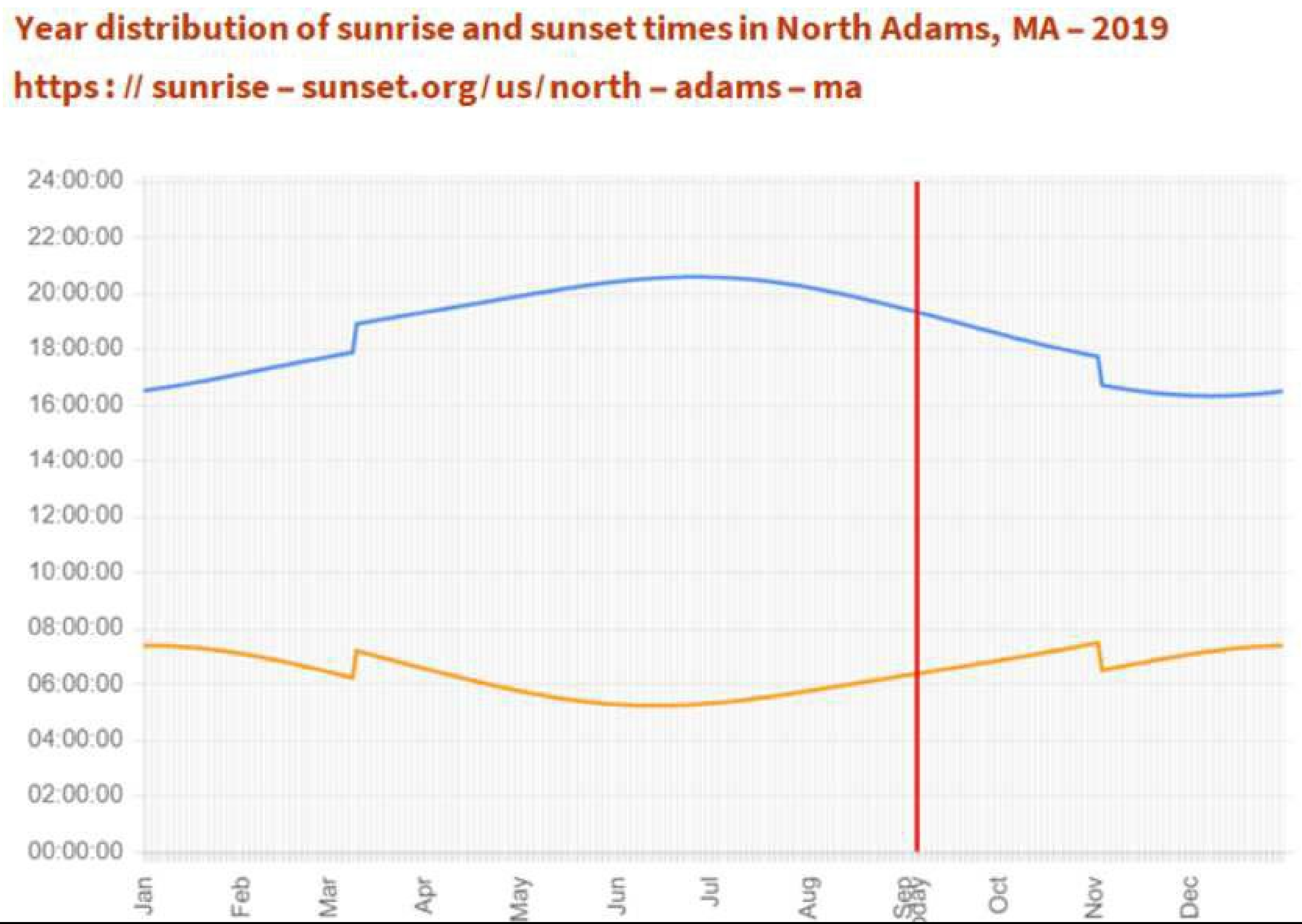
- Party less than the person next to you.
- Take advantage of office hours / mentoring.

Other: Advice from Jeff Miller

- Party less than the person next to you.
- Take advantage of office hours / mentoring. Learn
- to manage your time: no one else wants to.

Other: Advice from Jeff Miller

- Party less than the person next to you.
- Take advantage of office hours / mentoring. Learn to manage your time: no one else wants to.
- Happy to do practice interviews, adjust deadlines....


Year distribution of sunrise and sunset times in North Adams, MA – 2019
https : // sunrise – sunset.org / us / north – adams – ma
24:00:00
22:00:00
20:00:00
18:00:00
16:00:00
14:00:00
12:00:00
10:00:00
08:00:00
06:00:00
04:00:00
02:00:00
00:00:00
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


Who America is rooting for in the Super Bowl:
@NFL_MEMES

Gambling

Football Wager

2007: Friend of a favorite student bet $500 at 1000:1 odds on Patriots going undefeated and winning the Superbowl.

Football Wager

2007: Friend of a favorite student bet $500 at 1000:1 odds on Patriots going undefeated and winning the Superbowl.

Football Wager

2008: In third quarter, Pats leading, Vegas offers to buy back the bet at 300:1, told no....

WHAT WAS THE BETTOR’S MISTAKE?

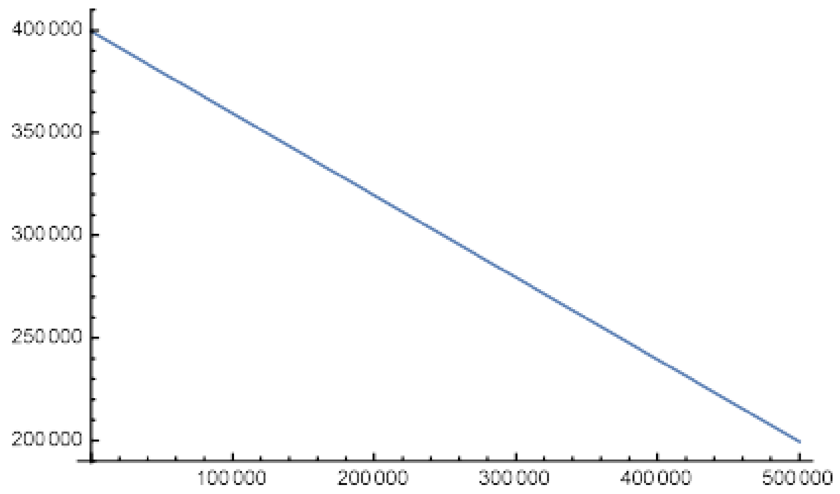
Hedging

Pats win with probability p, Giants $q = 1 - p$.

Bet \$1 bet on Giants, if they win get \$$x$.
Already bet \$500 on Patriots, now bet \$$B$ on the Giants.

Expected Winning:

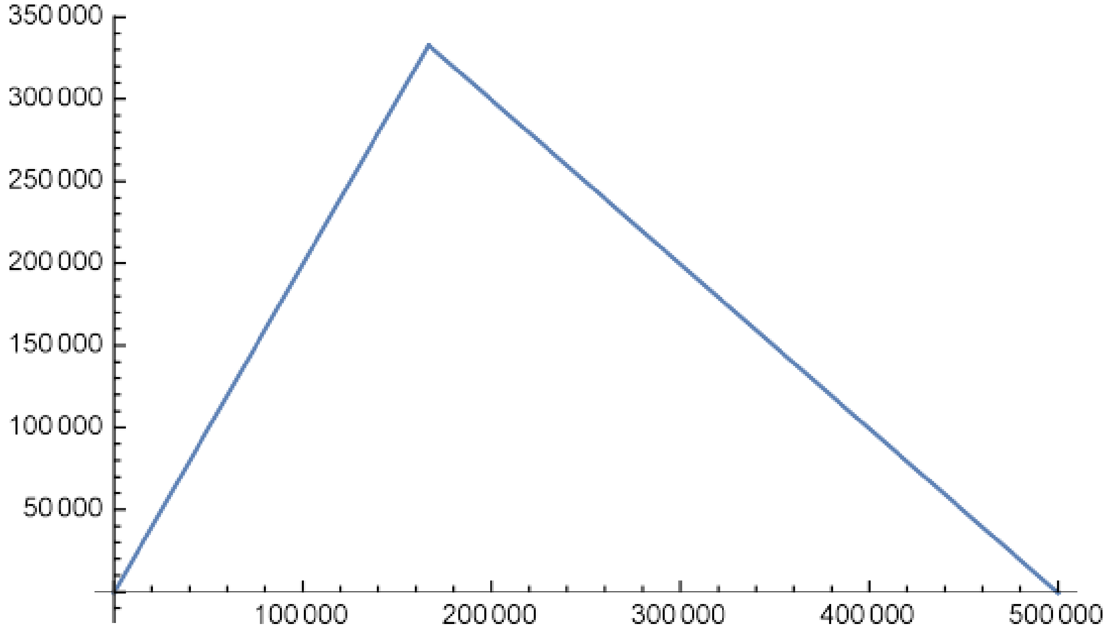
$$f(p, x, B) = p \cdot 500000 + (1 - p)Bx - 500 - B.$$



Guaranteed Winnings

By hedging can ensure some winnings:

$$g(p, x, B) = \min(500000, Bx) - 500 - B.$$

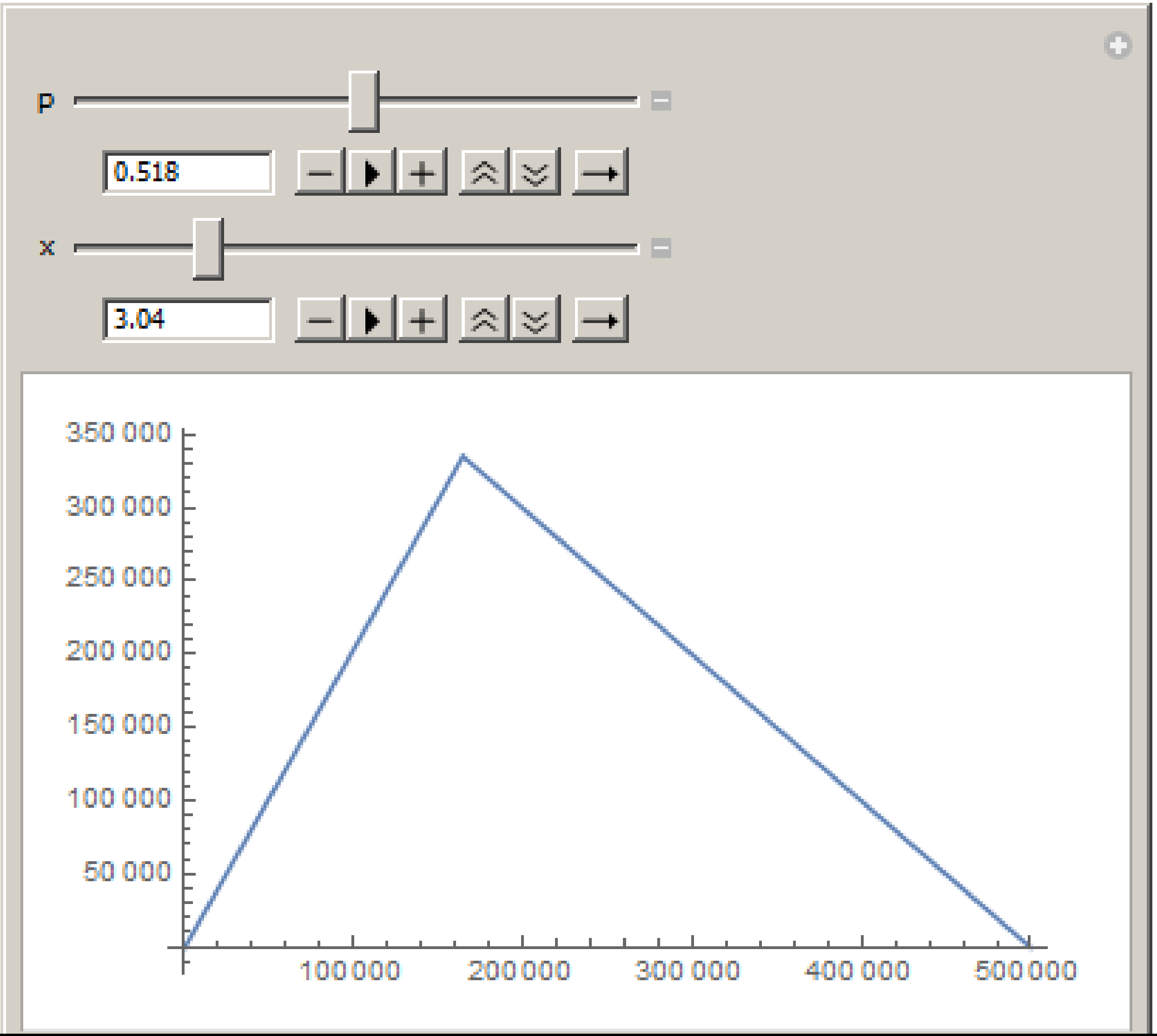


Here $p = .8, x = 3.$

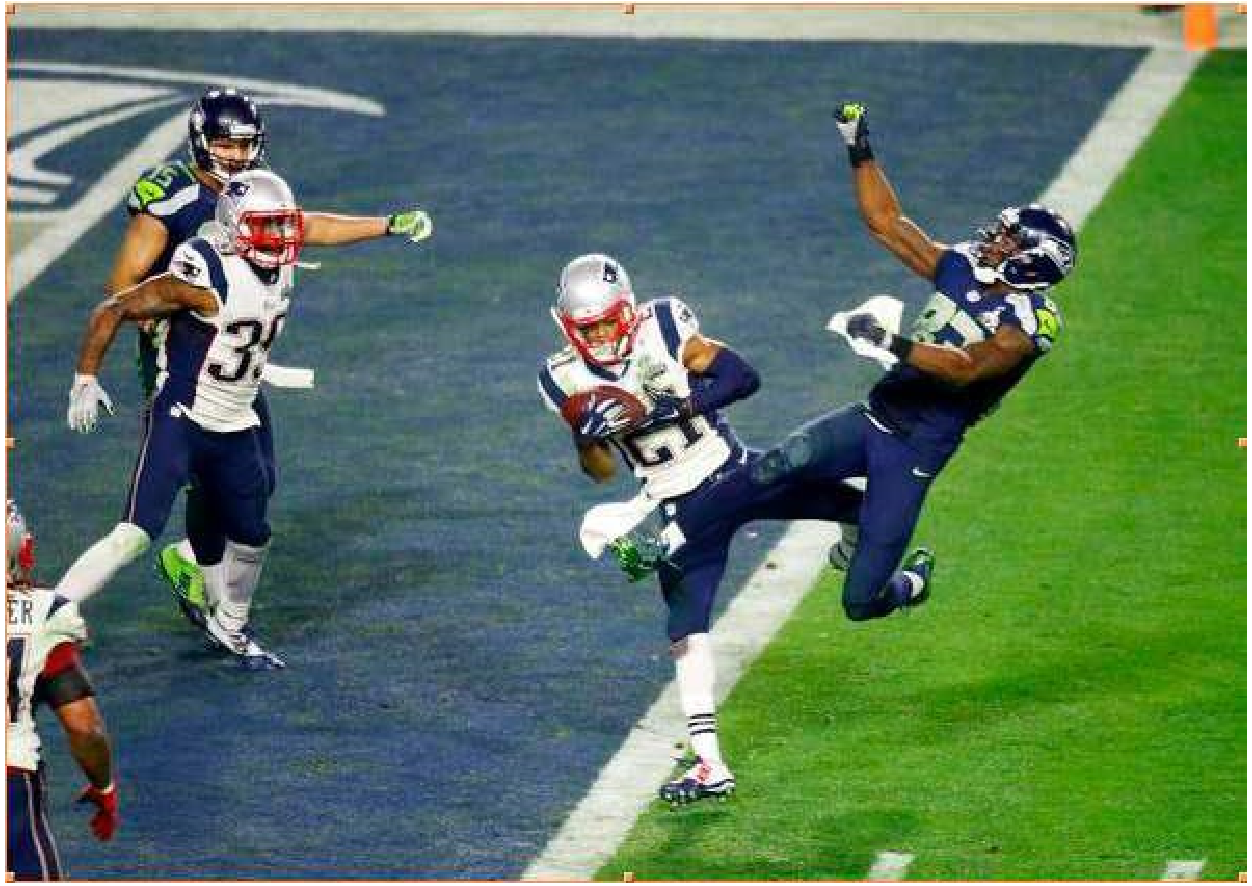
Mathematica Code

```
f[p_, x_, B_] := 500000 p + (1 - p) B x - 500 - B
g[p_, x_, B_] := Min[500000, B x] - 500 - B
Plot[f[.8, 3, B], {B, 0, 500000}]
Plot[g[.8, 3, B], {B, 0, 500000}]
Manipulate[Plot[g[p, x, B], {B, 0, 500000}], {p, 0, 1}, {x, 1, 10}]
```

Mathematica Code



Sabermetrics Club at Williams....



http://fivethirtyeight.com/features/

a-head-coach-botched-the-end-of-the-super-bowl-and-it-wasnt-pete-carroll/

Clicker Problems

Birthday Problem I

[redacted]

How large must N be for there to be at least a 50% probability that two of the N people share a birthday?

Birthday Problem I

[redacted]

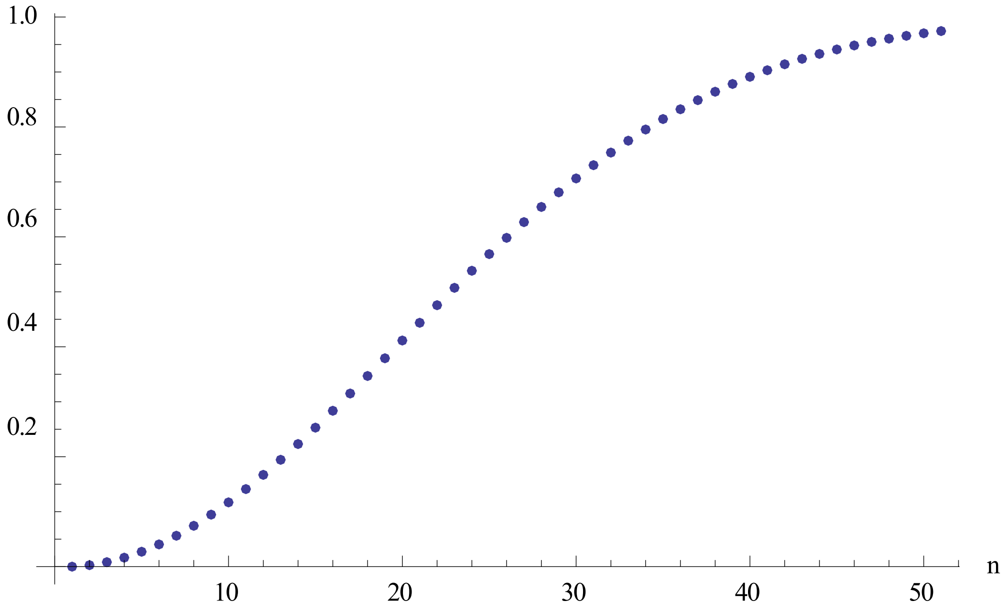
How large must N be for there to be at least a 50% probability that two of the N people share a birthday?

- (A) 11 people
- (B) 22 people
- (C) 33 people
- (D) 44 people
- (E) 90 people
- (F) 180 people
- (G) 365 people
- (H) 500 people.

Birthday Problem I

[redacted]

How large must N be for there to be at least a 50% probability that two of the N people share a birthday?



Birthday Problem II

How large must N be for there to be at least a 50% probability that two of N Plutonians share a birthday?

Birthday Problem II

How large must N be for there to be at least a 50% probability that two of N Plutonians share a birthday? 'Recall' one Plutonian year is about 248 Earth years (or 90,520 days).

Birthday Problem II

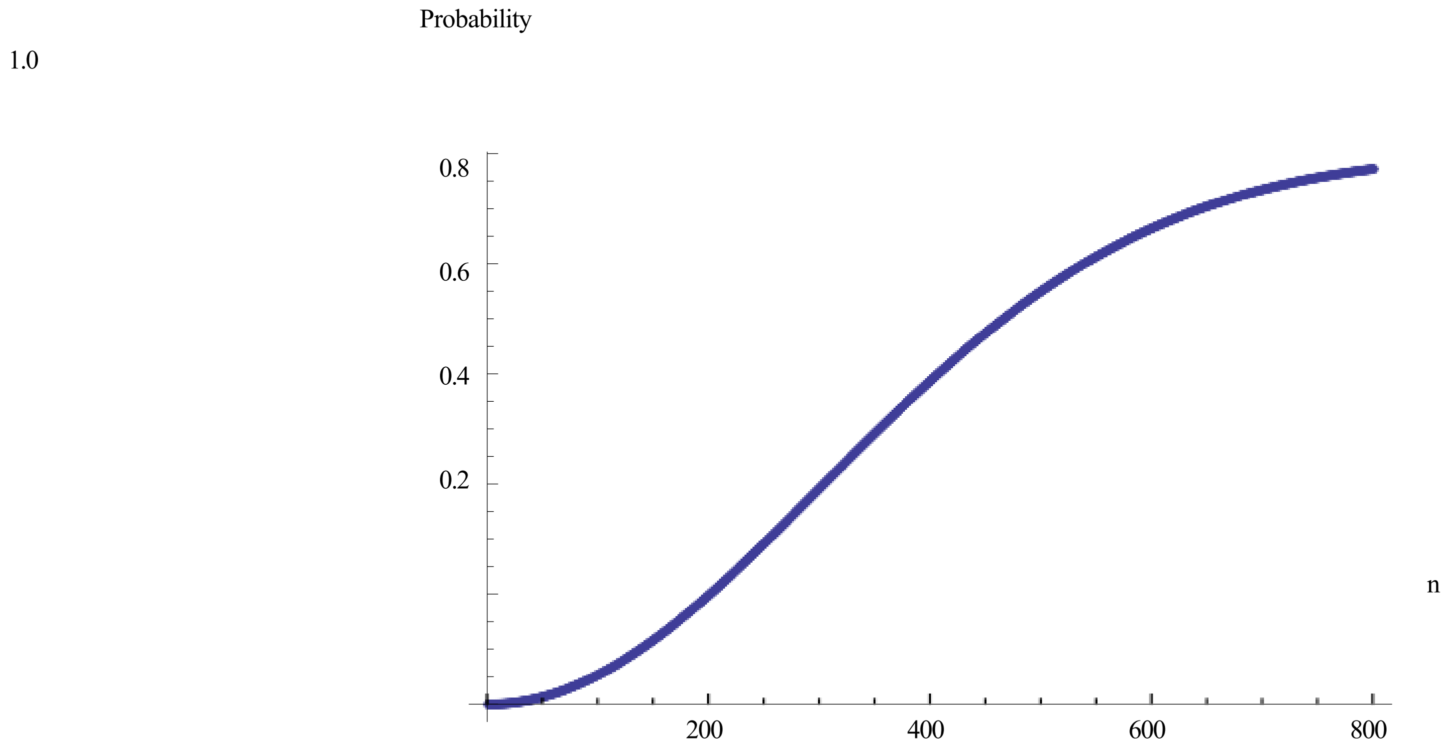
How large must N be for there to be at least a 50% probability that two of N Plutonians share a birthday? 'Recall' one Plutonian year is about 248 Earth years (or 90,520 days).

(A) 110 people

- (B) 220 people
- (C) 330 people
- (D) 440 people
- (E) 1,000 people
- (F) 5,000 people
- (G) 10,000 people
- (H) 20,000 people
- (I) more than 30,000 people.

Birthday Problem II

How large must N be for there to be at least a 50% probability that two of N Plutonians share a birthday? 'Recall' one Plutonian year is about 248 Earth years (or 90,520 days).



Voting: Democratic Primaries

During the Democratic primaries in 2008, Clinton and Obama received exactly the same number of votes in Syracuse, NY. How probable was this?

Voting: Democratic Primaries

During the Democratic primaries in 2008, Clinton and Obama received exactly the same number of votes in Syracuse, NY. How probable was this? (Note: they each received 6001 votes.)

- (A) 1 / 10
- (B) 1 / 100
- (C) 1 / 1,000
- (D) 1 / 10,000
- (E) 1 / 100,000
- (F) 1 / 1,000,000 (one in a million)
- (G) 1 / 1,000,000,000 (one in a billion).

Voting: Democratic Primaries (continued)

Syracuse University mathematics Professor Hyune-Ju Kim said the result was less than one in a million, according to the Syracuse Post-Standard, which quoted the professor as saying, "It's almost impossible." Her comments were reprinted widely, as the Associated Press picked up the story. (Carl Bialik, WSJ, 2/12/08)

Voting: Democratic Primaries (continued)

Syracuse University mathematics Professor Hyune-Ju Kim said the result was less than one in a million, according to the Syracuse Post-Standard, which quoted the professor as saying, "It's almost impossible." Her comments were reprinted widely, as the Associated Press picked up the story. (Carl Bialik, WSJ, 2/12/08)

Far greater than 1/137! What's going on?

Voting: Democratic Primaries (continued)

Syracuse University mathematics Professor Hyune-Ju Kim said the result was less than one in a million, according to the Syracuse Post-Standard, which quoted the professor as saying, "It's almost impossible." Her comments were reprinted widely, as the Associated Press picked up the story. (Carl Bialik, WSJ, 2/12/08)

Far greater than 1/137! What's going on?

Prof. Kim's calculation ... was based on the assumption that Syracuse voters were likely to vote in equal proportions to the state as a whole, which went for Ms. Clinton, its junior senator, 57%-40%. .... Prof. Kim said she had little time to make the calculation, so she made the questionable assumption ... for simplicity.

From Shooting Hoops
to the Geometric Series Formula

Simpler Game: Hoops

Game of hoops: first basket wins, alternate shooting.

Simpler Game: Hoops: Mathematical Formulation

Bird and **Magic** (I'm old!) alternate shooting; first basket wins.

- **Bird** always gets basket with probability p.
- **Magic** always gets basket with probability q.

Let x be the probability **Bird** wins – what is x?

Solving the Hoop Game

Classic solution involves the geometric series.

Break into cases:

Solving the Hoop Game

Classic solution involves the geometric series.

Break into cases:

- **Bird** wins on 1^{st} shot: p.

Solving the Hoop Game

Classic solution involves the geometric series.

Break into cases:

- **Bird** wins on 1st shot: p.
- **Bird** wins on 2nd shot: $(1 - p)(1 - q) \cdot p$.

Solving the Hoop Game

Classic solution involves the geometric series.

Break into cases:

- **Bird** wins on 1st shot: p.
- **Bird** wins on 2nd shot: $(1 - p)(1 - q) \cdot p$.
- **Bird** wins on 3rd shot: $(1 - p)(1 - q) \cdot (1 - p)(1 - q) \cdot p$.

Solving the Hoop Game

Classic solution involves the geometric series.

Break into cases:

- **Bird** wins on 1st shot: p.
- **Bird** wins on 2nd shot: $(1 - p)(1 - q) \cdot p$.
- **Bird** wins on 3rd shot: $(1 - p)(1 - q) \cdot (1 - p)(1 - q) \cdot p$.
- **Bird** wins on nth shot:
  $(1 - p)(1 - q) \cdot (1 - p)(1 - q) \cdots (1 - p)(1 - q) \cdot p$.

Solving the Hoop Game

Classic solution involves the geometric series.

Break into cases:

Bird wins on 1st shot: p.

- **Bird** wins on 2nd shot: $(1-p)(1-q)\cdot p$.
- **Bird** wins on 3rd shot: $(1-p)(1-q)\cdot(1-p)(1-q)\cdot p$.
- **Bird** wins on nth shot:
- $(1-p)(1-q)\cdot(1-p)(1-q)\cdots(1-p)(1-q)\cdot p$.

Let $r=(1-p)(1-q)$. Then

$$\begin{aligned} x &= \text{Prob}(\textbf{Bird}\text{ wins}) \\ &= p+rp+r^2p+r^3p+\cdots \\ &= p\left(1+r+r^2+r^3+\cdots\right), \end{aligned}$$

the geometric series.

Solving the Hoop Game: The Power of Perspective

Showed

$$x = \text{Prob}(\textbf{Bird}\text{ wins}) = p(1 + r + r^2 + r^3 + \cdots);$$

will solve without the geometric series formula.

Solving the Hoop Game: The Power of Perspective

Showed

$$x = \text{Prob}(\textbf{Bird}\text{ wins}) = p(1 + r + r^2 + r^3 + \cdots);$$

will solve without the geometric series formula.

Have

$$x = \text{Prob}(\textbf{Bird}\text{ wins}) = p +$$

Solving the Hoop Game: The Power of Perspective

Showed

$$x = \mathrm{Prob}(\textbf{Bird}\ \mathrm{wins}) = p(1 + r + r^2 + r^3 + \cdots);$$

will solve without the geometric series formula.

Have

$$x = \mathrm{Prob}(\textbf{Bird}\ \mathrm{wins}) = p + (1 - p)(1 - q)$$

Solving the Hoop Game: The Power of Perspective

Showed

$$x = \mathrm{Prob}(\textbf{Bird}\text{ wins}) = p(1 + r + r^2 + r^3 + \cdots);$$

will solve without the geometric series formula.

Have

$$x = \mathrm{Prob}(\textbf{Bird}\text{ wins}) = p + (1 - p)(1 - q)x$$

Solving the Hoop Game: The Power of Perspective

Showed

$$x = \text{Prob}(\textbf{Bird} \text{ wins}) = p(1 + r + r^2 + r^3 + \cdots);$$

will solve without the geometric series formula.

Have

$$x = \text{Prob}(\textbf{Bird} \text{ wins}) = p + (1 - p)(1 - q)x = p + rx.$$

Solving the Hoop Game: The Power of Perspective

Showed

$$x = \text{Prob}(\textbf{Bird} \text{ wins}) = p(1 + r + r^2 + r^3 + \cdots);$$

will solve without the geometric series formula.

Have

$$x = \text{Prob}(\textbf{Bird} \text{ wins}) = p + (1 - p)(1 - q)x = p + rx.$$

Thus

$$(1 - r)x = p \quad \text{or} \quad x = \frac{p}{1 - r}.$$

Solving the Hoop Game: The Power of Perspective

Showed

$$x = \text{Prob}(\textbf{Bird} \text{ wins}) = p(1 + r + r^2 + r^3 + \cdots);$$

will solve without the geometric series formula.

Have

$$x = \text{Prob}(\textbf{Bird} \text{ wins}) = p + (1 - p)(1 - q)x = p + rx.$$

Thus

$$(1 - r)x = p \quad \text{or} \quad x = \frac{p}{1 - r}.$$

As $x = p(1 + r + r^2 + r^3 + \cdots)$, find

$$1 + r + r^2 + r^3 + \cdots = \frac{1}{1 - r}.$$

$r = (1-p)(1-q)$

$0 \le p, q \le 1$

$0 \le r \le 1$

Lessons from Hoop Problem

- Power of Perspective: Memoryless process.
- Can circumvent algebra with deeper understanding! (Hard)
- Depth of a problem not always what expect.
- Importance of knowing more than the minimum: connections.
- Math is fun!

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 2 : Video 2-10-25: https://youtu.be/r2vIDHJOoSQ Birthday Problem, Coding, Integration, Sniffing out Equations
Slides: https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 2: February 11, 2025:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/handouts/341Notes_Chap1.pdf

- Review factorials, binomials, choosing....
- Birthday problems.
- QWERTY, Babylonian Mathematics
- General review: combinatorics, integration, ...?
- Power of coding

General items.

- Difference between a formula and a useful formula.
- Institutionalization: think hard on notation, actions.
- Sniffing out formulas?

HW: Due February 20:

#1: Section 1.2: Modify the basketball game so that there are 2015 players, numbered 1, 2, ..., 2015. Player i always gets a basket with probability 1/2i. What is the probability the first player wins?

#2: Section 1.2: Is the answer for Example 1.2.1 consistent with what you would expect in the limit as c tends to minus infinity?

#3: Section 1.2: Compute the first 42 terms of 1/998999 and comment on what you find; you may use a computer.

#4: Section 2.2.1: Find sets A and B such that |A| = |B|, A is a subset of the real line and B is a subset of the plane (i.e., R2) but is not a subset of any line.

#5: Section 2.2.1: Write at most a paragraph on the continuum hypothesis (you may use Wikipedia or any source to look it up).

#6: Section 2.2.2: Give an example of an open set, a closed set, and a set that is neither open nor closed (you may not use the examples in the book); say a few words justifying your answer.

#7: Section 2.3: Give another proof that the probability of the empty set is zero.

#8: Find the probability of rolling exactly k sixes when we roll five fair die for k = 0, 1, ..., 5. Compare the work needed here to the complement approach in the book. #9: If f and g are differentiable functions, prove the derivative of f(x)g(x) is f'(x)g(x) + f(x)g'(x). Emphasize where you add zero.

Birthday Problem: If all days equally likely to be a birthday, how many to have a 50% chance that two share?

Assume D days in a year

Calculate prob something doesn't happen, then prob happens is 1 − that

Prob(no two of n share a birthday with D days in a year)

$$= \frac{D-0}{D} \, \frac{D-1}{D} \cdots \frac{D-(n-1)}{D}$$

$$= \prod_{k=0}^{n-1} \frac{D-k}{D} = \prod_{k=0}^{n-1} \left(1 - \frac{k}{D}\right) =: q_{n,D}$$

$:$ means defn

PAVLOV: Prod: Think logarithm!

$$q_{n,D} = \prod_{k=0}^{n-1} \left(1 - \frac{k}{D}\right)$$

$$\log q_{n,D} = \log \prod_{k=0}^{n-1} \left(1 - \frac{k}{D}\right) = \sum_{k=0}^{n-1} \log\left(1 - \frac{k}{D}\right)$$

Taylor : $\log(1-x) = -x - x^2/2 - x^3/3 - x^4/4 \dots \quad |x|<1$

$$(Tf)(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)\, x^n}{n!}$$

First order: $\log(1-x) = -x$ (better is $-x - x^2/2$)

$$\underbrace{\log q_{n,D}}_{\substack{\text{prob}<1 \text{ so} \\ \log \text{ is neg}}} = \underbrace{\sum_{k=0}^{n-1} -\frac{k}{D}}_{\text{negative}} = -\frac{1}{D} \sum_{k=0}^{n-1} k = -\frac{1}{D} \frac{(n-1)\, n}{2}$$

$S_n = 0 + 1 + \dots + n$

$S_n = n + (n-1) + \dots + 0$

$2S_n = n + \dots + n = n(n+1) \rightarrow S_n = \frac{n(n+1)}{2}$

$$\log \varepsilon_{n,D} = -\frac{1}{D}(n-1)n$$

Find n st $\varepsilon_{n,D} = 1/2$

$$\log 1/2 = -\frac{1}{D}(n-1)n$$

but $\log 1/2 = -\log 2 \approx -.69$

$$(n-1)n = D\log 2$$

Solve by Quadratic Formula

$$(n-1)n \approx \left(n-\tfrac{1}{2}\right)^2 = n^2 - n + \tfrac{1}{4}$$

$$(n-1/2)^2 \approx D\log 2$$

or $n \approx \sqrt{D\log 2} + \frac{1}{2}$

So $n \approx (\log 2)^{\frac{1}{2}} D^{1/2} + 1/2$

Critical n grows like $D^{1/2}$

$D = 365 \quad n \approx 21$

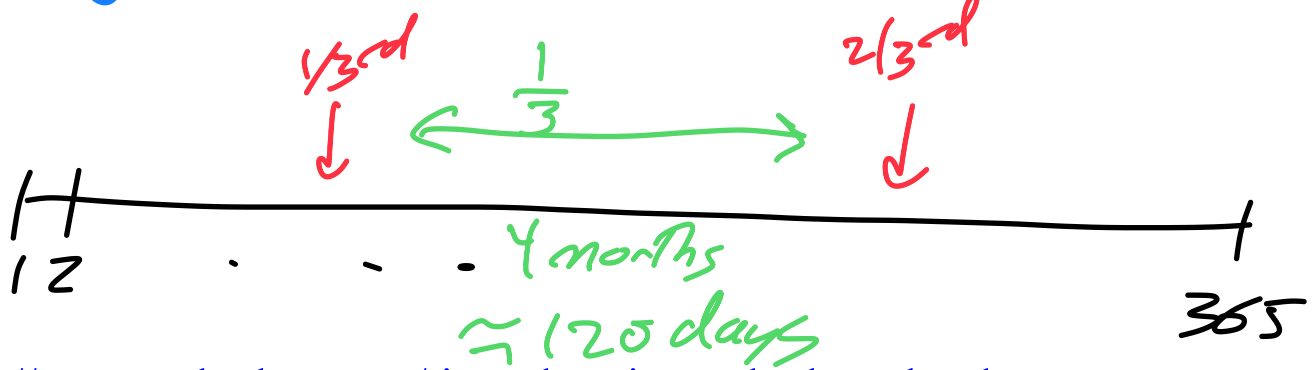
Choose two distinct days randomly in a year.

Conj: six months

Upper Bound: 1 year (minus a day)

Lower Bound: 1 day

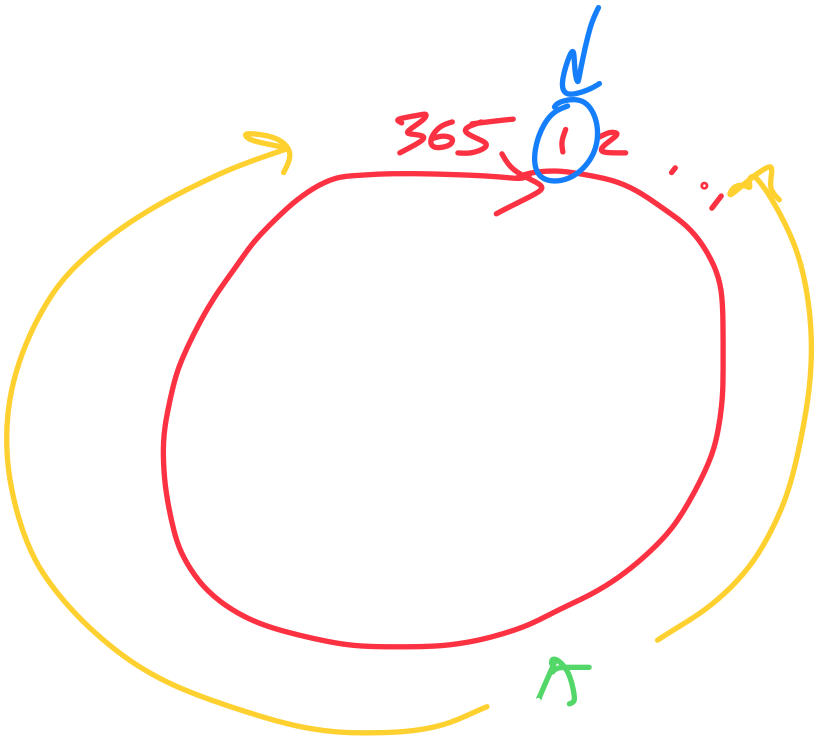
Conj: 4 or 5 months



https://www.calculator.net/time-duration-calculator.html

Class Mean: 144.846

13 data points



Max separation is 6 months, so 1/2 that or 3 months 90 days

```
twodayspread[d_, numiter_] := Module[{},
  spread = {}; (* record data here *)
  squarespread = {}; (* record square of data here *)
  year = Range[1, d]; (* list of days in the year *)For[n = 1, n ≤ numiter, n++,
   {
    dayschosen = RandomSample[year, 2];
    difference = Abs[dayschosen[[1]] - dayschosen[[2]]];
    spread = AppendTo[spread, difference];
    squarespread = AppendTo[squarespread, difference^2];
   }]; (* end of for loop on number of iterations *)
  Print["Observed average spread = ", 1.0 Mean[spread], ", theory = ",
   1.0 (d + 1) / 3, " days."];
  Print["Observed average spread^2 = ", 1.0 Mean[squarespread], ", theory = ",
   1.0 d (d + 1) / 6, " days."];
 ] (* end of program*)
```

```
10^m is 10
Observed average spread = 128.8, theory = 122. days.
Observed average spread^2 = 18 968.6, theory = 22 265. days.
{0., Null}

10^m is 100
Observed average spread = 126.84, theory = 122. days.
Observed average spread^2 = 23 812.2, theory = 22 265. days.
{0., Null}

10^m is 1000
Observed average spread = 123.927, theory = 122. days.
Observed average spread^2 = 22 684.9, theory = 22 265. days.
{0.015625, Null}

10^m is 10 000
Observed average spread = 121.704, theory = 122. days.
Observed average spread^2 = 22 094.9, theory = 22 265. days.
{0.21875, Null}
```

```
10^m is 1000
Observed average spread = 123.927, theory = 122. days.
Observed average spread^2 = 22 684.9, theory = 22 265. days.
{0.015625, Null}
```

>10

```
10^m is 10 000
Observed average spread = 121.704, theory = 122. days.
Observed average spread^2 = 22 094.9, theory = 22 265. days.
{0.21875, Null}
```

>100

```
10^m is 100 000
Observed average spread = 121.838, theory = 122. days.
Observed average spread^2 = 22 207.2, theory = 22 265. days.
{29.7344, Null}
```

>200

```
10^m is 1 000 000
Observed average spread = 121.988, theory = 122. days.
Observed average spread^2 = 22 258.1, theory = 22 265. days.
{7600.31, Null}
```

```
fasttwodayspread[d_, numiter_] := Module[{},
  spread = 0; (* record data here *)
  squarespread = 0; (* record square of data here *)
  year = Range[1, d]; (* list of days in the year *)For[n = 1, n ≤ numiter, n++,
   {
    dayschosen = RandomSample[year, 2];
    difference = Abs[dayschosen[[1]] - dayschosen[[2]]];
    spread = spread + difference;
    squarespread = squarespread + difference^2;
   }]; (* end of for loop on number of iterations *)
  Print["Observed average spread = ", 1.0 spread / numiter, ", theory = ",
   1.0 (d + 1) / 3, " days."];
  Print["Observed average spread^2 = ", 1.0 squarespread / numiter, ", theory = ",
   1.0 d (d + 1) / 6, " days."];
 ] (* end of program*)
```

numbers
not arrays

```
Doing 10^4 trials.

Observed average spread = 122.225, theory = 122. days.

Observed average spread^2 = 22 253., theory = 22 265. days.

{0.265625, Null}

Observed average spread = 122.121, theory = 122. days.

Observed average spread^2 = 22 457.1, theory = 22 265. days.

{0.03125, Null}


Doing 10^5 trials.

Observed average spread = 122.405, theory = 122. days.

Observed average spread^2 = 22 391.3, theory = 22 265. days.

{29.7656, Null}

Observed average spread = 122.106, theory = 122. days.

Observed average spread^2 = 22 292.9, theory = 22 265. days.

{0.375, Null}
```

≈ 10

```
Doing 10^7 trials.

Observed average spread = 122.042, theory = 122. days.

Observed average spread^2 = 22 279.7, theory = 22 265. days.

{48.125, Null}
```

≈ 10

```
Doing 10^8 trials.

Observed average spread = 122.009, theory = 122. days.

Observed average spread^2 = 22 267.9, theory = 22 265. days.

{405.875, Null}
```

≈ 10

```
Doing 10^9 trials.

Observed average spread = 122.001, theory = 122. days.

Observed average spread^2 = 22 264.7, theory = 22 265. days.

{3861.38, Null}
```

Die Problem: Do you take the bet?

I have a fair 6 sided die with[1 - 6] on the sides and you are paid the same amount of dollars as the number you roll.If you roll a 5 you get $5, if you roll a 1 you get $1, etc.The question was to find out "how much are you willing to pay to play the game?"

The game then evolved to how much are you willing to pay to play the game with two dice?In this case you get to keep your first roll or you get to roll again and are forced to take whatever the second roll is.Keep in mind, this is for a role that pays~$110, 000 a year out of college and I blurted the answers out within 15 seconds.Even if you don' t do well in probability at Williams, if you know basic concepts like expected value and can keep cool in an interview, then you can move your interview along quite nicely and put yourself in a good position to go from getting a D/D - in probability to not being broke😂.

```
twodiesim[iterations_] := Module[{},
   get = 0; (* keeps a running sum of what you get *)
   For[n = 1, n ≤ iterations, n++,
    {
     x = RandomInteger[{1, 6}]; (* roll a fair die *)
     (* if roll > 3 keep and add, else re-roll and keep that *)
     If[x > 3, get = get + x, get = get + RandomInteger[{1, 6}]];
    }];
   get = get / iterations; (* average earning per game *)
   Print["Earn on average ", get 1.0, "."]
  ];
```

```
Timing[twodiesim[100000]]
Timing[twodiesim[1000000]]
Timing[twodiesim[10000000]]
Timing[twodiesim[100000000]]
```

```
Earn on average 4.25225.

{0.1875, Null}

Earn on average 4.25142.

{1.875, Null}

Earn on average 4.25019.

{18.6406, Null}

Earn on average 4.25026.

{182.156, Null}
```

10

10

10

My logic : second die on average 3.5, take the first if it is a 4, 5, 6 (average is 5). So half the time get a 5, half the time a 3.5, average is 4.25

HANDOUTS:

•REVIEW MATERIAL: Numerous worked out calculus problems (differentiation, integration, statement of topics you should know)

•Notes on Induction, Calculus, Convergence, the Pigeon Hole Principle and Lengths of Sets (from the first appendix to *An Invitation to Modern Number Theory*, by myself and Ramin Takloo-Bighash, Princeton University Press 2006). We will not need all the material there for this course, but it is easier to just post the entire chapter.

•Handout on Types of Proofs (from a handout I wrote for math review sessions at Princeton, 1996-1997; this was written for students from calculus to linear algebra).

•Intermediate and Mean Value Theorems and Taylor Series (you should know this material already; the main results are stated and mostly proved, subject to some technical results from analysis which we need to rigorously prove the IVT).

•Sequences and series:

- From multivariable calculus (Cain and Herod)
- From calculus (Strang)

•Free textbooks:

- Numerous free textbooks from Georgia Tech
- Free textbook on multivariable calculus (Cain and Herod)
- Free probability textbook (Grinstead and Snell), Free real analysis textbook (from William Trench)

•Isaac Asimov: The Relativity of Wrong

Integration

Change of variable (u-substitution)
↳ Trig substitution
∫ by parts (product rule)
partial fractions
Converting to double integral
Differentiating under the integral sign
Bring it over method

See also Lecture 03: 9/15/21: Die Problem, Coding, Integration: https://youtu.be/ZhPypMaQWhc (slides)

Integration by parts

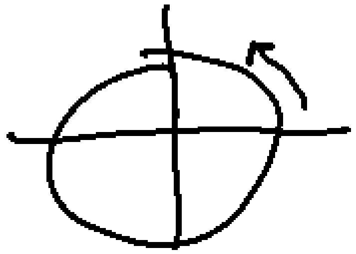
$$[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$$

$$[u(x)v(x)]' = u'(x)v(x) + u(x)v'(x)$$

$$u\,dv = u(x)v'(x)\,dx$$

$$\frac{du}{dx}dx$$

or $$\int_a^b u\,\overset{=v'(x)dx}{dv} = u(x)v(x)\Big|_a^b - \int_a^b v\,du$$



Ex: $$\int_0^\pi x\cos x\,dx$$

$$u(x) = x \qquad dv = \cos x\,dx$$

$$du = dx \qquad v = \sin x$$

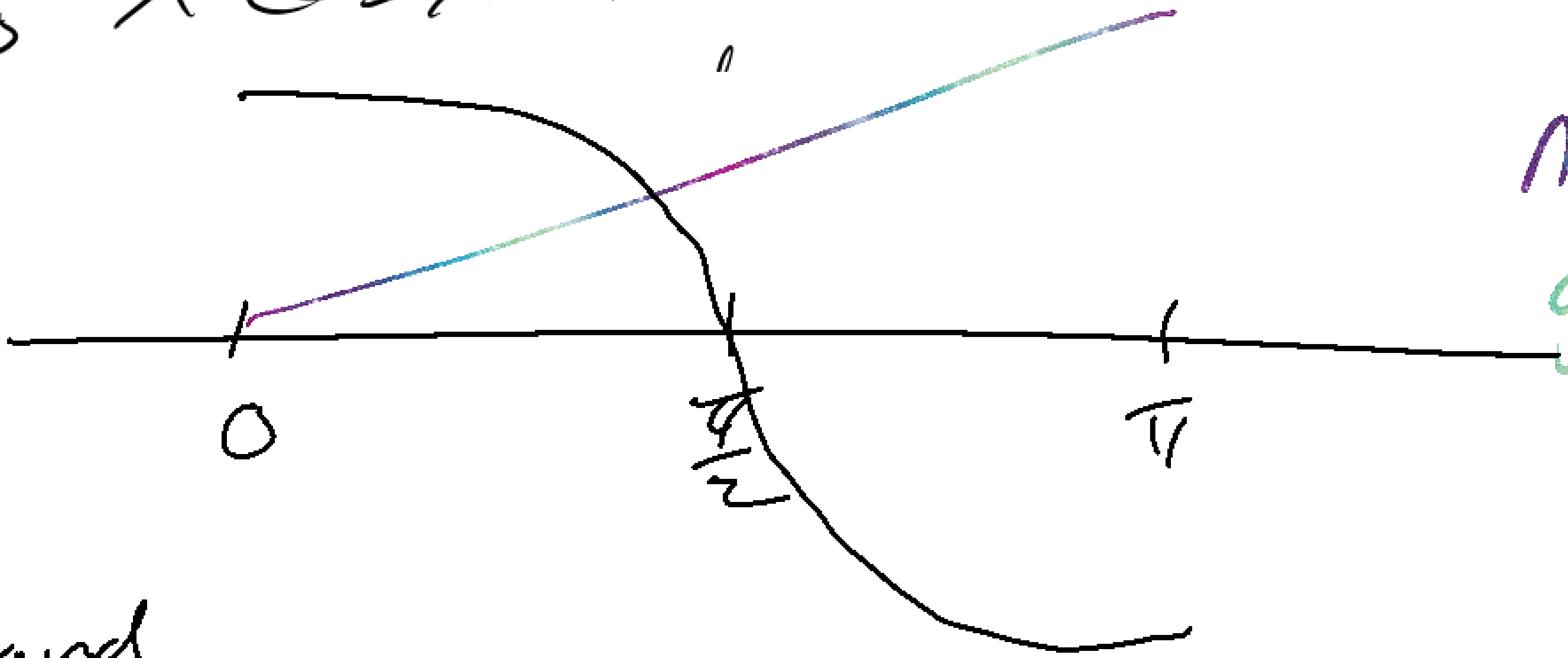
$$\cos'(x) = -\sin x$$

minus sign

$$= uv\Big|_0^\pi - \int_0^\pi v\,du = x\sin x\Big|_0^\pi - \int_0^\pi \sin x\,dx$$

$$= 0 + \cos x\Big|_0^\pi = -2$$

$\int_0^{\pi} x\cos x\,dx = -2$: is this reasonable?



neg region gets larger weight

Expect negative answer

Bound

$0 \geq \int_0^{\pi} x\cos x\,dx \geq -\pi^2$

improve to $\geq -\frac{1}{2}\pi^2$

Change of variable

$$\int_0^\infty x e^{-x^2/2} dx = \int_0^\infty e^{-x^2/2} \underline{\underline{x dx}}$$

$$u = x^2/2 \qquad du = x dx$$

$$\int_{u=0}^\infty e^{-u} du = -e^{-u}\Big|_0^\infty = e^{-u}\Big|_\infty^0 = 1$$

$$\int_{-\infty}^{\infty} x e^{-4x^2} dx = \int_{-\infty}^{\infty} e^{-4x^2} x\,dx$$

Odd function integral is 0

$u = 4x^2$ $\qquad$ $du = 8x\,dx$

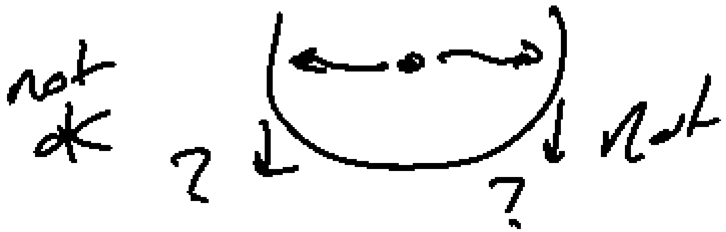
$$\frac{1}{8}\int_{x=-\infty}^{\infty} e^{-4x^2}\, 8x\,dx = \frac{1}{8}\int_{u=-\infty}^{\infty} e^{-u}\,du$$

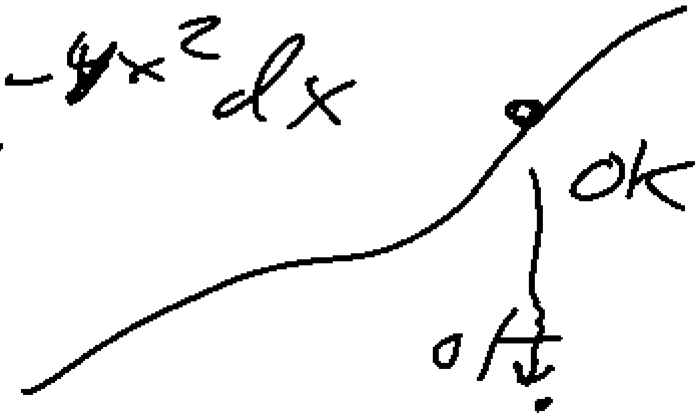
NO!

Be explicit: $x: -\infty$ to ∞, $u = 4x^2$

$u: \infty$ to ∞

not 1-1 (onto): need to do $2\int_0^{\infty} x e^{-4x^2} dx$





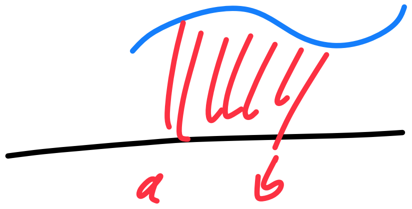
Gaussian Integral

$f(x) = e^{-x^2}$

$N(\mu, \sigma)$ normal random variable with mean μ, std dev σ

↳ density $f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$

Prob density: (1) $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (2) $f(x) \geq 0$



prob value btw a and b

Prove $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$\Longrightarrow \int_{-\infty}^{\infty} f_{\mu,\sigma}(x)\,dx = 1$

Comfortable!

changing variables

$$I := \int_{-\infty}^{\infty} e^{-x^2}\,dx$$

$$I^2 = \int_{-\infty}^{\infty} e^{-x^2}\,dx \int_{-\infty}^{\infty} e^{-y^2}\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dx\,dy$$

$$= \int_{\theta=0}^{2\pi}\int_{r=0}^{\infty} e^{-r^2/2}\, r\,dr\,d\theta$$

$$= \int_{\theta=0}^{2\pi} d\theta \int_{r=0}^{\infty} e^{-r^2/2}\, r\,dr$$

$u = r^2/2$

$du = \frac{2r\,dr}{2} = r\,dr$

$r: 0 \to \infty \Rightarrow u: 0 \to \infty$

$$= 1\Big|_0^{2\pi} \int_{u=0}^{\infty} e^{-u}\,du$$

$$= 2\pi\left[-e^{-u}\right]_0^{\infty} = 2\pi$$

so $I^2 = 2\pi \Rightarrow I = \sqrt{2\pi}$

pos square root as integrand in I is positive

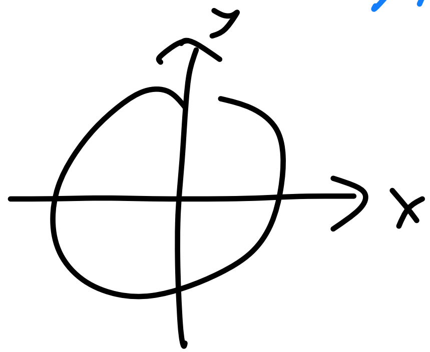
Polar coords:

x, y: $-\infty$ to ∞

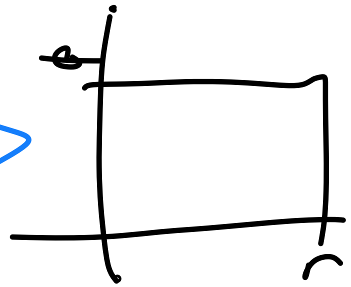
$r: 0 \to \infty$, $\theta: 0 \to 2\pi$

$dx\,dy = r\,dr\,d\theta$

$\underbrace{dx}_{m}\,\underbrace{dy}_{m} = \underbrace{(dr)}_{\text{meters}}\,\underbrace{(r\,d\theta)}_{\text{meters}}$

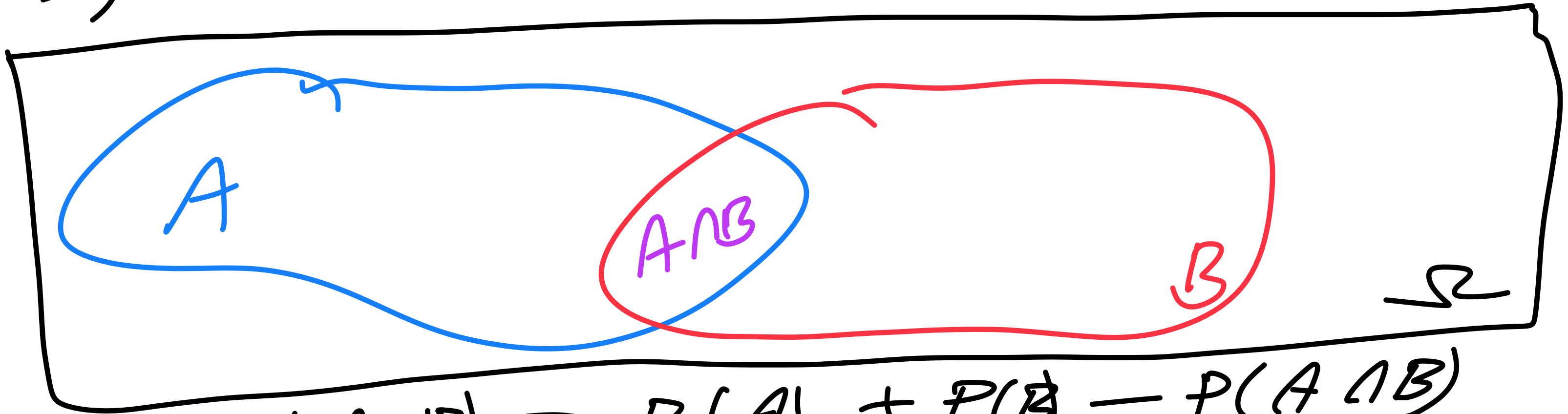


$\Rightarrow$



Sniffing at formulas

Prob$(A \cup B)$ as a function of $P(A)$, $P(B)$, $P(A \cap B)$



$\text{Prob}(\Omega) = 1$

$\text{Prob}(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\text{Prob}(A \cup B \cup C)$ or $\text{Prob}(A_1 \cup \cdots \cup A_n)$

Conj: $\text{Prob}(A \cup B) = \alpha P(A) + \beta \, \text{Prob}(B) + \gamma \, \text{Prob}(A \cap B)$

Conj: $\text{Prob}(A \cup B) = \alpha\, P(A) + \beta\, \text{Prob}(B) + \gamma\, \text{Prob}(A \cap B)$

Study special cases of (A, B), with $A, B \subset \Omega$, $\text{Prob}(\Omega) = 1$

A	B	$P(A)$	$P(B)$	$P(A \cap B)$		$P(A \cup B)$
		$\alpha = 1$	$\alpha = 1$	$\gamma = -1$	$=$	
A	A^c	$P(A)$	$1 - P(A)$	0		1
A	$\emptyset$	$P(A)$	0	0		$P(A)$
A	Ω	$P(A)$	1	$P(A)$		1
A	A	$P(A)$	$P(A)$	$P(A)$		$P(A)$

$\alpha = \beta$ as $\text{Prob}(A \cup B) = \text{Prob}(B \cup A)$ symmetry

consistent $\qquad \alpha = 1$

$p = \text{Prob}(A)$: $\alpha p + \alpha(1-p) = 1$, $\alpha p = p$, $\alpha p + \alpha 1 + \gamma p = 1$

$\alpha p + \alpha p + \gamma p = p$ $\qquad \gamma = -1 \qquad$ consistent

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 3 Video 2-12-25: Binomial Coefficients, Poker, Coding Efficiently, Generating Functions, Laws of Probability. https://youtu.be/aByMKJ8NPEE

Slides: https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan of the Day

- Review Binomial Coefficients
- Proof by Story
- Coding Efficiency in Poker
- Generating Functions
- Roulette: https://youtu.be/yyTOb7BtzbY
- Axioms of Probability (if time): Outcome space, probability functions, wish list, …..

Watch at home: Set Theory, Probability Wish List, Coding: https://youtu.be/2agUkFQJtnU (2019 lecture)

$n!$ (read n factorial) is $n(n-1)\cdots 3\cdot 2\cdot 1$
with $0! = 1$

↳ how many arrangements of n objects when order matters.

(extension: $(-\tfrac{1}{2})! = \sqrt{\pi}$)

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n!/(n-k)!}{k!} = \#\text{ ways to choose } k$$

from n when order does not matter.

$$\frac{n!}{(n-k)!} = (n-0)(n-1)\cdots(n-(k-1))$$

\# ways to choose k from n when order matters

$\boxed{n}\ \boxed{n-1}\ \cdots\ \boxed{n-(k-1)}$

Pascal's Triangle

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

↳ Binomial Thm

↳ use is in calculus: deriv of x^n: $\lim_{h \to 0} \frac{(x+h)^n - x^n}{h}$

↳ Proof: $(x+y)^n = (x+y)(x+y) \cdots (x+y)$

How get term $x^k y^{n-k}$?

↳ take x a total of k times, then have y total of $n-k$ times

$$\binom{n}{k}\binom{n-k}{n-k} = \binom{n}{k} \text{ as } \binom{n-k}{n-k} = 1$$

Challenge: $(x+y+z)^n = ((x+y)+z)^n$

Pascal's Triangle

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
$$

$(x+y)^n$: binomial coeff

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$$

Proof 1: Expand!

Proof 2: Story

n people from Williams

1 person from Amherst

Choose $k+1$ people from these $n+1$

$$\binom{n+1}{k+1} = \underbrace{\overbrace{\binom{1}{1}}^{=1}}_{\text{choose Amherst}}\binom{n}{k} + \underbrace{\overbrace{\binom{1}{0}}^{=1}}_{\text{no Amherst}}\binom{n}{k+1} \qquad \blacksquare$$

Choose k from n. order doesn't matter: $\binom{n}{k}$

Get stars all else get nothing

$||$

Choose $n-k$ from n, order doesn't matter, to be excluded $\binom{n}{n-k}$

algebra proof: $\binom{n}{n-k} = \frac{n!}{(n-k)!\,(n-(n-k))!} = \frac{n!}{(n-k)!\,k!} = \frac{n!}{k!\,(n-k)!} = \binom{n}{k}$

or do story: counting the same thing!

What is the prob have at least 2 kings in a hand of 5 cards?

K K ☆ ☆ ☆ K K K ☆ ☆ K K K K ☆

☆ = non king

Prob is

$$\frac{\binom{4}{2}\binom{48}{3} + \binom{4}{3}\binom{48}{2} + \binom{4}{4}\binom{48}{1}}{\binom{52}{5}}$$

$$1 - \frac{\binom{4}{0}\binom{48}{5} + \binom{4}{1}\binom{48}{4}}{\binom{52}{5}}$$

Hand of 5 cards, at least 2 Kings, at least 2 Queens

KKQQ☆ KKQQQ KKKQQ ☆= non-K, non-Q

Pseudo code:

- Create deck of 52 cards: 1, 2, ..., 52 and 1,2,3,4 King; 5,6,7,8 Queen
- randomly and uniformly choose 5 cards
- Check one at a time to see if 1,2,3,4 is in, if yes ↑ king counter by 1. Then do for queen counter. Both must be ≥ 2 for success.

Deck is {10, 10, 10, 10, 1, 1, 1, 1, 0, —— 0}

See if sum is in {22, 23, 32}

Generating Functions

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$ a_n sequence we care about

Hope: get a closed form expression for f, use that to deduce things about the a_n's.

Fibonacci Numbers: $F_0 = 0$, $F_1 = 1$, $F_{n+1} = F_n + F_{n-1}$

$$g(x) = \sum_{n=0}^{\infty} F_n x^n = \frac{x}{1-x-x^2}$$ (I think)

$\hookrightarrow$ yields Binet's Formula: $F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n$

$$f(x) = 1 + x + x^2 + \cdots = \frac{1}{1-x} \quad \text{if } |x|<1$$

Proof: $f(x) = 1 + x(1 + x + x^2 + \cdots)$

$$f(x) = 1 + x f(x) \Rightarrow (1-x) f(x) = 1 \Rightarrow f(x) = \frac{1}{1-x}$$

$$f(x) = 1 + x + x^2 + x^3 + \cdots = \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

$$x \frac{d}{dx} f(x) = 0 + 1x + 2x^2 + 3x^3 + \cdots = \sum_{n=0}^{\infty} n x^n = \frac{x}{(1-x)^2}$$

Ex: $x = \frac{1}{2} \Rightarrow 0 + \frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \frac{4}{2^4} + \cdots + \cdots = \frac{1/2}{(1-1/2)^2} = 2$

Fun Definitions

$$\binom{3}{5} = ?$$ Should be zero!

$$\binom{3}{5} = \frac{3!}{5!\,(3-5)!} = \frac{6}{120 \cdot (-2)!}$$

suggests $(-2)!$ is ∞

Gamma Function: $\Gamma(s) = \int_0^\infty e^{-x} x^{s-1}\,dx$

$\Gamma(n+1) = n!$ if n is a non-neg integer

and $\Gamma(1/2) =$ "$(-1/2)!$" $= \sqrt{\pi}$

Wish list for Prob

$$\text{Prob}(A \text{ and } B) = \text{Prob}(A) + \text{Prob}(B)$$

↳ if $B = A \Rightarrow \text{Prob}(A) = 2\,\text{Prob}(A)$ so all events have 0 probability.

↳ it is true if A and B are disjoint

More wishes

$$\mathrm{Prob}(A \cup B \cup C \cup \cdots) = \mathrm{Prob}(A) + \mathrm{Prob}(B) + \cdots$$

if pairwise disjoint

Danger is uncountable infinities

Integers / Rationals (countable) Irrationals / Reals (uncountable)

$$\mathrm{Prob}\left(\bigcup_{x \in [0,1]} \{x\}\right) \underset{\text{if true}}{=} \sum_{x \in [0,1]} \mathrm{Prob}(\{x\})$$

Need more than pairwise disjoint! Need countable

Wish list:

1. For any event A, we have $0 \leq \Pr(A) \leq 1$, and if Ω is the outcome space, then $\Pr(\Omega) = 1$.

2. If $\{A_i\}$ is a pairwise disjoint collection of sets (which means $A_j \cap A_k$ is empty if $j \neq k$), then $\Pr(\cup_i A_i) = \sum_i \Pr(A_i)$.

finite sum

or

Sum $i: 0 \to \infty$

Sets $A_1, A_2, A_3, \ldots$

(Kolmogorov's) Axioms of Probability $(\Omega, \Sigma, \mathrm{Prob})$, where Ω is the outcome space and Σ a σ-algebra, is a probability space if the probability function satisfies the following.

- If $A \in \Sigma$ then $\Pr(A)$ is defined and $0 \leq \Pr(A) \leq 1$.
- $\Pr(\emptyset) = 0$ and $\Pr(\Omega) = 1$.
- Let $\{A_i\}$ be a finite or countable collection of disjoint elements of Σ. Then $\Pr(\cup_i A_i) = \sum_i \Pr(A_i)$.

Useful Rules for Probability Spaces. Let $(\Omega, \Sigma, \mathrm{Prob})$ be a probability space. Then

1. **"Law of Total Probability"**: If $A \in \Sigma$, then $\Pr(A) + \Pr(A^c) = 1$. Equivalently, $\Pr(A) = 1 - \Pr(A^c)$.

2. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$. This can be generalized. For example, if we have three events, then

   $\mathrm{Prob}((A_1 \cup A_2) \cup A_3)$

   $$\begin{aligned}\Pr(A_1 \cup A_2 \cup A_3) \quad = \quad & \Pr(A_1) + \Pr(A_2) + \Pr(A_3) \\ & - \Pr(A_1 \cap A_2) - \Pr(A_1 \cap A_3) \\ & - \Pr(A_2 \cap A_3) \ + \ \Pr(A_1 \cap A_2 \cap A_3)\,.\end{aligned}$$

   This is also known as **inclusion-exclusion**.

3. If $A \subset B$, then $\Pr(A) \leq \Pr(B)$. If, however, A is a proper subset of B, we don't necessarily have $\Pr(A) < \Pr(B)$, but we do know for certain that $\Pr(B) = \Pr(A) + \Pr(B \cap A^c)$, where $B \cap A^c$ refers to all elements of B that aren't in A.

4. Let $A_i \subset B$ for all i. Then $\Pr(\cup_i A_i) \leq \Pr(B)$.

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 4 and 5 Slides: Replacement Videos:

- Roulette: Double Plus Ungood: https://youtu.be/Esa2TYwDmwA
- German Tank Problem: pdf (video: https://youtu.be/1N2IhpifwAk).

•Probability and Mathematical Modeling: I: (slides online here): https://youtu.be/T-35mhZioeo)

•Probability and Mathematical Modeling: II: (slides online here): https://youtu.be/ZGM6be90XqA)

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 6: Trump Splits, Conditional Probability, Bayes' Theorem, Method of Exhaustion, Inclusion/Exclusion, Independence: https://youtu.be/xMAzWda4yZU

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 06:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf (some notes on sets, probabilities, …)

- Trump Splits
- Conditional Probability (sniffing out formula)
- Inclusion/Exclusion
- Bayes' Theorem
- Independence
- Derangements

General items.
- Run simulations!
- Importance of phrasing.
- Explore extreme cases.

Probability of having a 5–0 trump split in bridge

4 hands of 13, partner and I have 8 trumps missing 5

$$\binom{2}{1}\binom{5}{5}\binom{21}{8}\binom{13}{13} \Big/ \binom{26}{13} * \binom{13}{13}$$

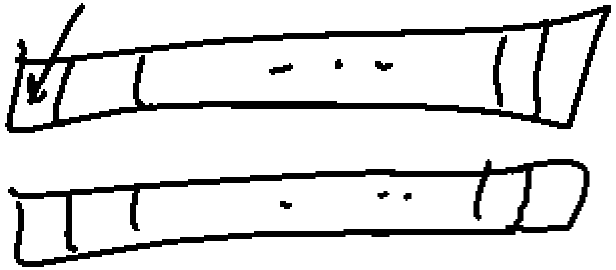
which opponent — Get all missing — fill up hand — other hand — first hand — second hand

$\frac{21!}{8!\,\cancel{13!}}$ $\qquad$ $\frac{26!}{13!\,\cancel{13!}}$

$$\binom{2}{1} \quad \frac{13}{26} * \frac{12}{25} * \frac{11}{24} * \frac{10}{23} * \frac{9}{22}$$

choose who gets

equal to $\frac{9}{230} \approx 3.91\%$



OR $\binom{2}{1} * \left(\frac{1}{2}\right)^5 = \frac{1}{16} \approx 6.25\%$

choose person

Probabilty of a 5-0 split: 9/230, or about 3.91% or is it 2/32 or about 6.25%?

-3 no comments!

```
badtrumpsplit[numbad_, numiter_] := Module[{},
   deck = {};
   For[c = 1, c ≤ numbad, c++, deck = AppendTo[deck, 1]];
   For[c = numbad + 1, c ≤ 26, c++, deck = AppendTo[deck, 0]];
   badsplits = 0;
   For[n = 1, n ≤ numiter, n++,
    {
     hand = RandomSample[deck, 13];
     If[Mod[Sum[hand[[i]], {i, 1, 13}], numbad] == 0, badsplits = badsplits + 1];
    }];
   Print["Observed badsplits is ", SetAccuracy[100.0 badsplits / numiter, 4], "%."];
  ];
```

```
Timing[badtrumpsplit[5, 1000000]]
```

```
Observed badsplits is 3.942%.
{15.5625, Null}
```

```
Timing[badtrumpsplit[5, 10000000]]
```

```
Observed badsplits is 3.913%.
{188.891, Null}
```

Are these two items equivalent:

Each person is equally likely to be chosen, form a group of two people from four.

Chose any group of two people, all groups equally likely to be chosen.

$A_1 \; A_2 \quad B_1 \; B_2$

$A_1 - B_1$ do not want to go together

$A_2 - B_2$ " " " " " "

$\binom{4}{2} = 6$ groups of 2

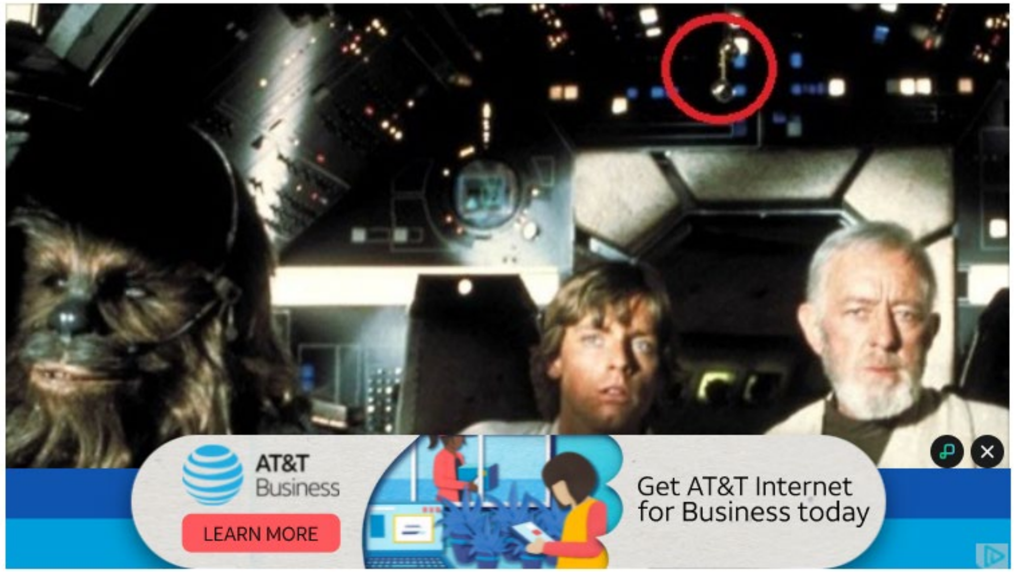
$A_1 A_2$	$A_2 B_1$
$\cancel{A_1 B_1}$	$\cancel{A_2 B_2}$
$A_1 B_2$	$B_1 B_2$

Rolling two fair independent die….

What is the probability that

1. The sum is an 11: $\frac{2}{36} = \frac{1}{18}$
2. The sum is a 7: $\frac{6}{36} = \frac{1}{6}$
3. Given first die is a 3, the sum is an 11: 0
4. Given first die is a 3, the sum is a 7: $\frac{1}{6}$

20. Han's Dice



When Luke boards the Millennium Falcon in The Last Jedi, he grabs a pair of gold dice which belonged to Han Solo, and though you may not have ever noticed them before, they were hanging up in the Falcon in A New Hope and also reappeared in The Force Awakens.

Conditional probability can be the same or different as the probability

Is it possible that there's some nice function F such that

$$\Pr(A|B) = F(\Pr(A), \Pr(B))?$$

$\Pr(A|B)$ is the prob of A given B happens

Choose A and B from: $A = B$, ϕ, Ω, $A = B^c$ or $B = A^c$

- If $A = B$ then $\Pr(A|A) = 1 = F(\Pr(A), \Pr(A))$
- If $B = \Omega$ then $\Pr(A|\Omega) = \Pr(A) = F(\Pr(A), 1)$
- If $B = A^c$ then $\Pr(A|A^c) = 0 = F(\Pr(A), 1 - \Pr(A))$

$F(x,y) \Rightarrow$ $F(x,x) = 1$

$F(x,1) = x$ — try $x = 1/2$

$F(x, 1-x) = 0$

If $\Pr(A) = 1/2$ if $B = A$ then $P(A|B) = 1$

if $B = A^c$ then $P(A|B) = 0$

but both are $F(1/2, 1/2)$: no such formula

$$\Pr(A|B) = G(\Pr(A), \Pr(B), \Pr(A \cap B))$$

$\Pr(A|A) = 1$, $\Pr(A|\Omega) = \Pr(A)$, $\Pr(A|A^c) = 0$

$$\frac{\Pr(A \cap B)}{\Pr(B)}$$

if $B = A$ get 1

if $B = A^c$ get 0

if $B = \Omega$ get $\Pr(A)$

$\tan(\arctan x) = x$

If $\Pr(B) = 0$ cannot talk about $\Pr(A|B)$ as B does not happen.

$$\Pr(A|B) = G(\Pr(A), \Pr(B), \underbrace{\Pr(A \cap B)}_{\text{A and B happen}})$$

- $\Pr(B|B) = 1$,
- $\Pr(B^c|B) = 0$, and
- $0 \leq \Pr(A|B) \leq 1$.

There's a simple expression using our three building blocks that has these three properties:

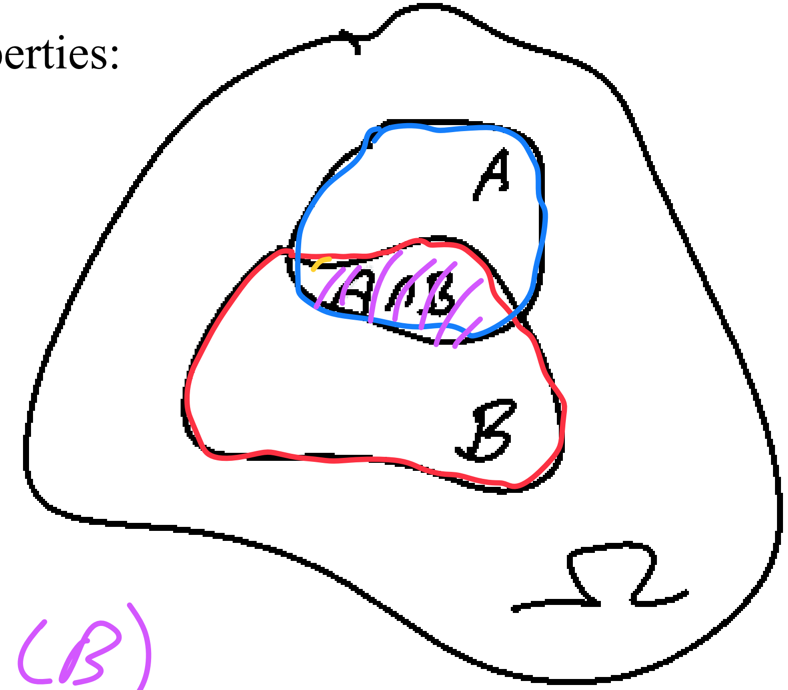
$$\Pr(A|B) = G(\Pr(A), \Pr(B), \underbrace{\Pr(A \cap B)}_{\text{A and B happen}})$$

- $\Pr(B|B) = 1$,
- $\Pr(B^c|B) = 0$, and
- $0 \leq \Pr(A|B) \leq 1$.

There's a simple expression using our three building blocks that has these three properties:

Try $P(A|B) = \dfrac{\Pr(A \cap B)}{P(B)}$

satisfies the three points above



Could write $\Pr(A \cap B) = \Pr(A|B)\Pr(B)$

$\Pr(B \cap A) = \Pr(B|A)\Pr(A)$

Expected Counts Approach

Suppose that you go out fishing one day, and you have the following set of rules: you stop fishing once you catch a fish, or after you've been on the water for four hours (whichever comes first). Let's also imagine that there's a 40% chance that you catch a trout, a 25% chance you catch a bass, and a 35% chance you don't catch anything. Notice that the percentages sum to 100%, and that you never catch more than one fish in a day. Now, if we know that you caught a fish one day, what are the odds that fish was a trout? Suppose that you went fishing 1000. Then….

Expected Counts Approach

Suppose that you go out fishing one day, and you have the following set of rules: you stop fishing once you catch a fish, or after you've been on the water for four hours (whichever comes first). Let's also imagine that there's a 40% chance that you catch a trout, a 25% chance you catch a bass, and a 35% chance you don't catch anything. Notice that the percentages sum to 100%, and that you never catch more than one fish in a day. Now, if we know that you caught a fish one day, what are the odds that fish was a trout? Suppose that you went fishing 1000. Then….

40% chance catch trout
25% chance catch a bass
35% chance catch nothing

$P(\text{trout} \mid \text{caught fish})$

$$= \frac{.40}{.65}$$

A: catch trout
B: catch a fish

1000 days

400 trout | 250 bass | 350 nothing

650 times caught something
400 of those are trout

So $\frac{400}{650}$

	A	A^c
B	$A \cap B$	$A^c \cap B$
B^c	$A \cap B^c$	$A^c \cap B^c$

Table 4.1: These are the possible outcomes for events A and B. If we know that event B has happened, we need only worry about the events in B's row.

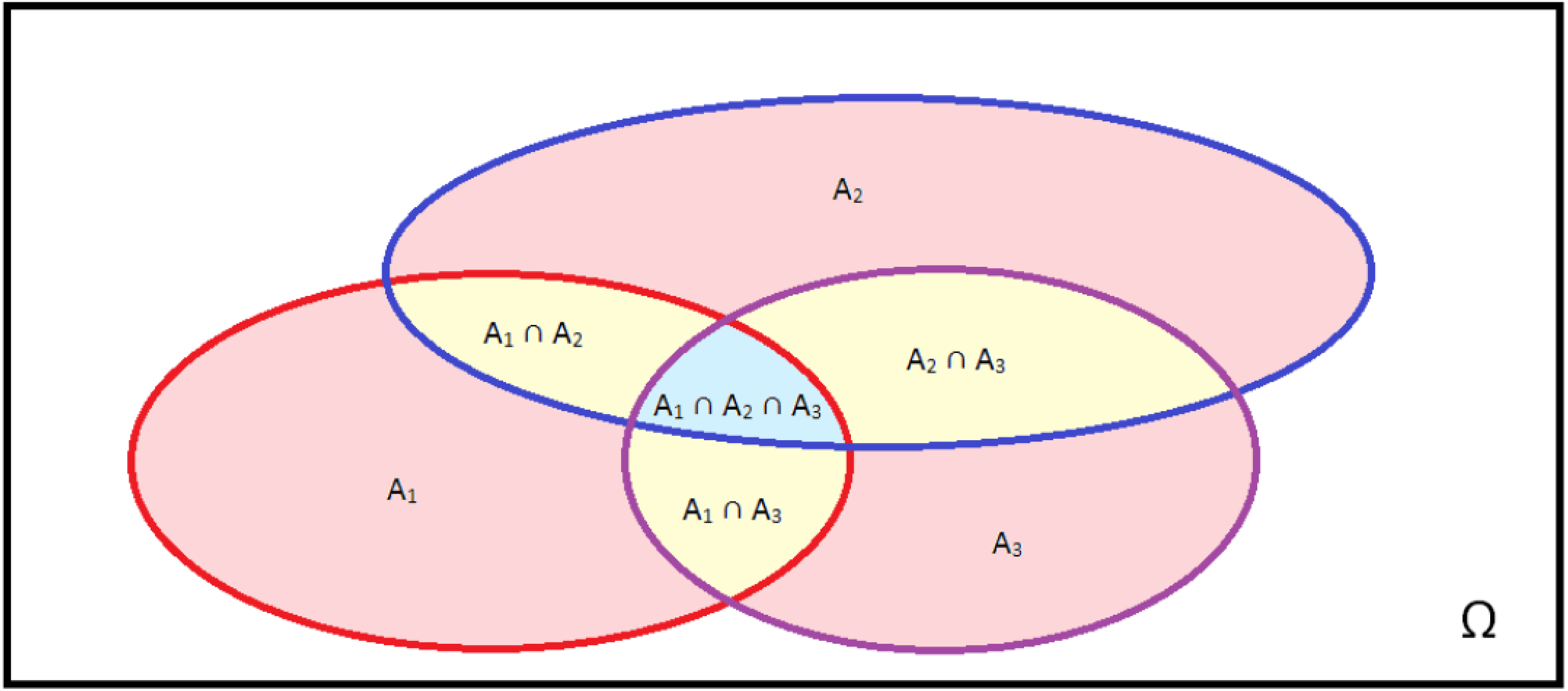
Conditional Probability: Let B be an event such that $\Pr(B) > 0$. Then the conditional probability of A given B is

$$\Pr(A|B) = \Pr(A \cap B) / \Pr(B).$$

Rewrite: $\Pr(A \cap B) = \Pr(A|B) * \Pr(B)$

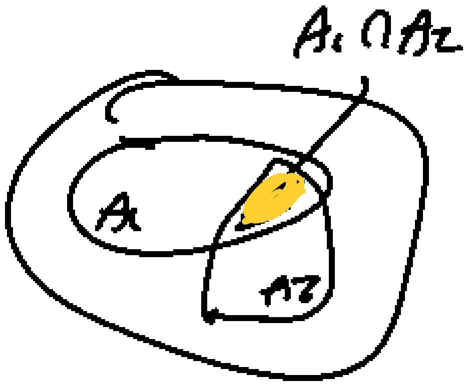
If you were really reading carefully, you might've noticed a new condition snuck into the box above: $\Pr(B) > 0$. If $\Pr(B) = 0$, then B cannot happen. If B cannot happen, it doesn't make sense to talk about the probability A happens given B happens! Fortunately if $\Pr(B) = 0$ then $\Pr(A \cap B)$ is also 0, and we have the indeterminate ratio 0/0, which warns us that we are in dangerous waters.

Illustration of inclusion-exclusion with three sets.



$$\Pr(A_1 \cup A_2 \cup A_3) = \Pr(A_1) + \Pr(A_2) + \Pr(A_3)$$
$$- \Pr(A_1 \cap A_2) - \Pr(A_1 \cap A_3) - \Pr(A_2 \cap A_3)$$
$$+ \Pr(A_1 \cap A_2 \cap A_3)$$

$$\Pr(A_1 \cup A_2)$$
$$= \Pr(A_1) + \Pr(A_2) - \Pr(A_1 \cap A_2)$$



Inclusion-Exclusion Principle: Consider sets $A_1, A_2, \ldots, A_n$. Denote the number of elements of a set S by $|S|$ and the probability of a set S by $\Pr(S)$. Then

$$\begin{aligned}\left|\bigcup_{i=1}^{n} A_i\right| &= \sum_{i=1}^{n}|A_i| - \sum_{1\le i<j\le n}|A_i\cap A_j| + \sum_{1\le i<j<k\le n}|A_i\cap A_j\cap A_k| - \\ &\quad \cdots + (-1)^{n-2}\sum_{1<\ell_1<\ell_2<\cdots<\ell_{n-1}\le n}|A_{\ell_1}\cap A_{\ell_2}\cap\cdots\cap A_{\ell_{n-1}}| \\ &\quad +(-1)^{n-1}|A_1\cap A_2\cap\cdots\cap A_n|;\end{aligned}$$

this also holds if we replace the size of all the sets above with their probabilities. We may write this more concisely. Let $A_{\ell_1\ell_2\ldots\ell_k} = A_{\ell_1}\cap A_{\ell_2}\cap\cdots\cap A_{\ell_k}$ (so $A_{12} = A_1\cap A_2$ and $A_{489} = A_4\cap A_8\cap A_9$). Then

$$\begin{aligned}\left|\bigcup_{i=1}^{n} A_i\right| &= \sum_{i=1}^{n}|A_i| - \sum_{1\le i<j\le n}|A_{ij}| + \sum_{1\le i<j<k\le n}|A_{ijk}| - \cdots \\ &\quad +(-1)^{n-2}\sum_{1<\ell_1<\ell_2<\cdots<\ell_{n-1}\le n}|A_{\ell_1\ell_2\cdots\ell_{n-1}}| + (-1)^{n-1}|A_{12\cdots n}|.\end{aligned}$$

If the A_i's live in a finite set and we use the counting measure where each element of our outcome space is equally likely, we may replace all $|S|$ above with $\Pr(S)$.

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n}|A_i| - \sum_{1\le i<j\le n}|A_i\cap A_j| + \sum_{1\le i<j<k\le n}|A_i\cap A_j\cap A_k| - \cdots + (-1)^{n-2}\sum_{1<\ell_1<\ell_2<\cdots<\ell_{n-1}\le n}|A_{\ell_1}\cap A_{\ell_2}\cap\cdots\cap A_{\ell_{n-1}}| + (-1)^{n-1}|A_1\cap A_2\cap\cdots\cap A_n|;$$

Imagine $x \in A_1 \cap A_2 \cap \cdots \cap A_n$

from $\sum_{i=1}^{n} |A_i|$ count it $\binom{n}{1}$

$\sum_{1\le i<j\le n} |A_i \cap A_j|$ count it $\binom{n}{2}$

Get $\binom{n}{1} - \binom{n}{2} + \binom{n}{3} - \binom{n}{4} \cdots + (-1)^{n-1}\binom{n}{n}$ $-\binom{n}{0}$ $+\binom{n}{0} = 1$

$= 1 - \left[\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + (-1)^n\binom{n}{n}\right]$

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

$$\Pr(A_1 \cup \cdots \cup A_n)$$

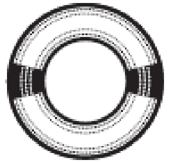
$$= n \Pr(A_1)$$

$$- \binom{n}{2} \Pr(A_1 \cap A_2)$$

$$+ \binom{n}{3} \Pr(A_1 \cap A_2 \cap A_3)$$

$$- \ldots$$

$$+ (-1)^{n-1} \binom{n}{n} \Pr(A_1 \cap \cdots \cap A_n)$$



A big caveat for independence of three or more events is that any combination of two of those events may be independent of each other, but three or more might be dependent. For example, roll a die twice. Let

- A denote the event that the first time the die shows an even number,
- B the second time the die shows an even number, and
- C the sum of the first two numbers is even.

We can see that

$$\begin{aligned}
\Pr(A \cap B) &= \Pr(A) \cdot \Pr(B) \\
\Pr(A \cap C) &= \Pr(A) \cdot \Pr(C) \\
\Pr(B \cap C) &= \Pr(B) \cdot \Pr(C).
\end{aligned}$$

However, in this case

$$\Pr(A \cap B \cap C) \neq \Pr(A) \cdot \Pr(B) \cdot \Pr(C),$$

as $\Pr(A \cap B \cap C)$ is the probability of getting an even number the first time and an even number again the second time (if the first two rolls are even, then the sum *must* be even – this is what causes the problems). We thus have $\Pr(A \cap B \cap C) = \frac{1}{4}$, but if the three events were independent then, according to the formula,

$$\Pr(A \cap B \cap C) = \Pr(A) \cdot \Pr(B) \cdot \Pr(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

Bayes' Theorem: The General Multiplication Rule implies

$$\Pr(B|A) \cdot \Pr(A) \;=\; \Pr(A|B) \cdot \Pr(B)$$

for events A and B. Therefore, so long as $\Pr(B) \neq 0$, we have

$$\Pr(A|B) \;=\; \Pr(B|A) \cdot \frac{\Pr(A)}{\Pr(B)}.$$

$P(A|B)$ and $P(A \cap B)$: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

<u>OB</u> : $P(A \cap B) = P(A|B)\, P(B) = P(B|A)\, P(A)$

symmetry $= P(B \cap A)$

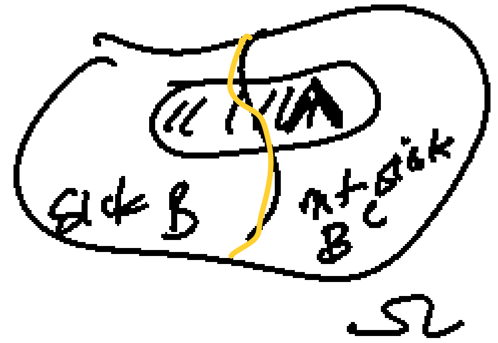
Example: Nationwide, Tuberculosis (TB) affects about 1 in every 15,000 people. Suppose that there's a TB scare in your town, and for simplicity assume that the rate of incidence of TB in your town is the same as the national average. Just to be safe, you go to the doctor to get tested for the disease. The doctor tells you that the test has a 1% false positive rate – which is to say that for every 100 healthy people, one will test positive. The doctor also reveals that the test has a 0.1% false negative rate – similarly, for every 1000 sick people, only one will test negative. Suppose that you test positive. What's the probability that you have TB?

$\frac{1}{15000}$ easy to find

Bayes' Theorem tell us that

$$\Pr(\text{sick}|\text{positive}) = \frac{\Pr(\text{sick})}{\underbrace{\Pr(\text{positive})}_{\text{ok}}} \cdot \underbrace{\Pr(\text{positive}|\text{sick})}_{\checkmark}.$$

Bayes' Theorem tell us that



$$\Pr(\text{sick}|\text{positive}) = \frac{\Pr(\text{sick})}{\Pr(\text{positive})} \cdot \Pr(\text{positive}|\text{sick}).$$

We use the partition "sick" and "not sick." So B is the event sick, B^c the event healthy (i.e., "not sick"), and A is the event of testing positive. We find

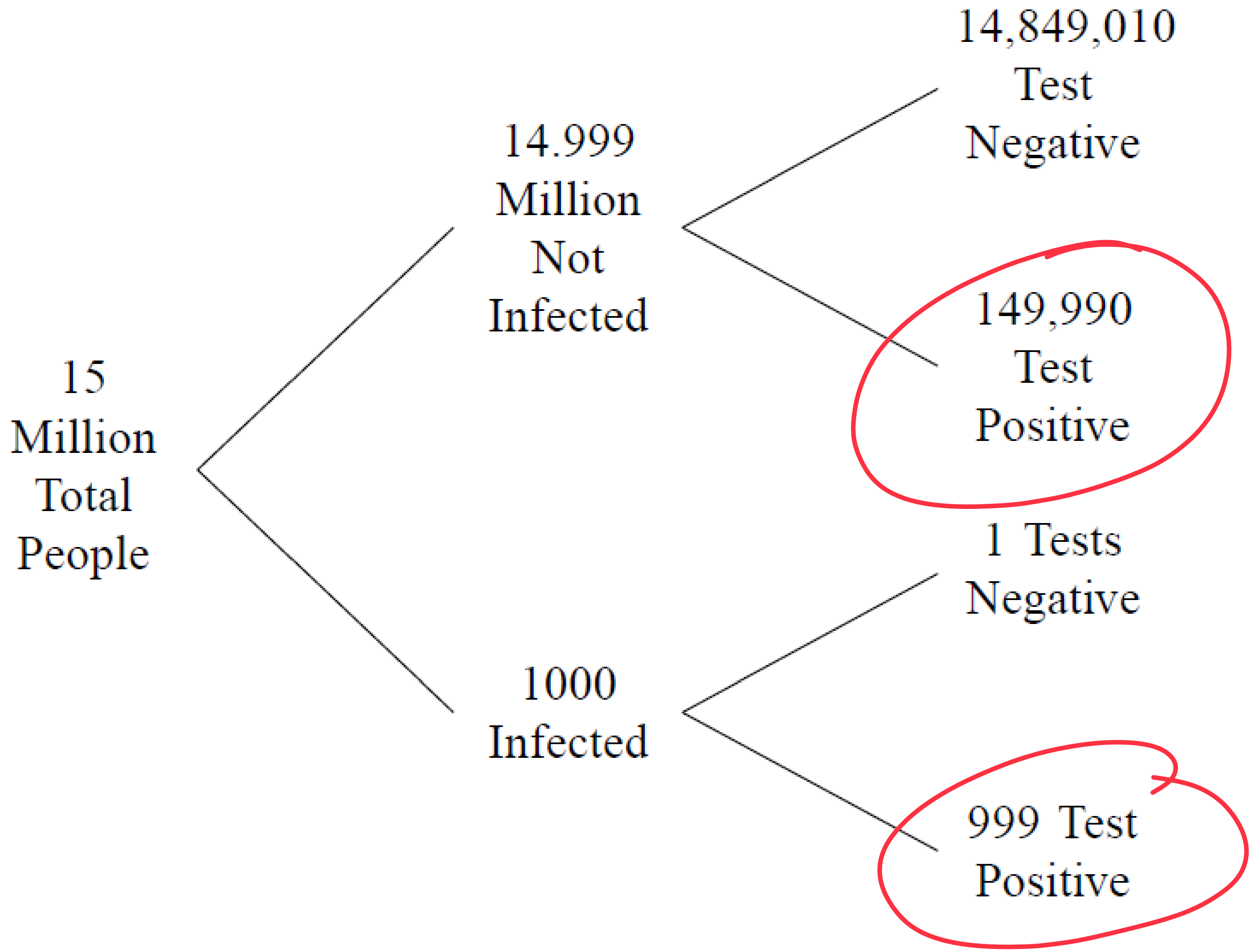
$$\begin{aligned}
\Pr(\text{positive}) &= \Pr(\text{positive}|\text{sick})\Pr(\text{sick}) + \Pr(\text{positive}|\text{healthy})\Pr(\text{healthy}) \\
&= 0.999 \cdot \frac{1}{15000} + 0.01 \cdot \frac{14999}{15000} \approx 0.01.
\end{aligned}$$

This gives

$$\Pr(\text{sick}|\text{positive}) = \frac{1/15000}{0.01} \cdot 0.999 \approx 0.0066.$$

Were you expecting the probability to be that low?

The expected counts approach can be seen graphically in the tree below.



Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 7: Inclusion-Exclusion Bounds, Derangements, Induction: https://youtu.be/e5H3XxPxdJk

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- Inclusion/Exclusion Bounds
- Partitions
- Derangements
- Basics of PDFs
- Random Variables: Continuous (FTC) vs Discrete
- Moments and Expected Values

General items.

- Rescale
- Taylor Trick (several variables)
- More coding: https://youtu.be/sSgjBysixdQ; code file here, pdf here

Inclusion/Exlcusion Bounds

At least one person has a one-suited hand in bridge

$$\text{Prob(North is one-suited)} = \frac{4}{\binom{52}{13}} = \frac{\binom{4}{1}\binom{13}{13}\cancel{\binom{39}{13}}\cancel{\binom{26}{13}}\cancel{\binom{13}{13}}}{\binom{52}{13}\cancel{\binom{39}{13}}\cancel{\binom{26}{13}}\cancel{\binom{13}{13}}}$$

So Prob at least one person is one-suited

$$= 4 * \text{Prob(North is one-suited)} - \text{Something Small}$$

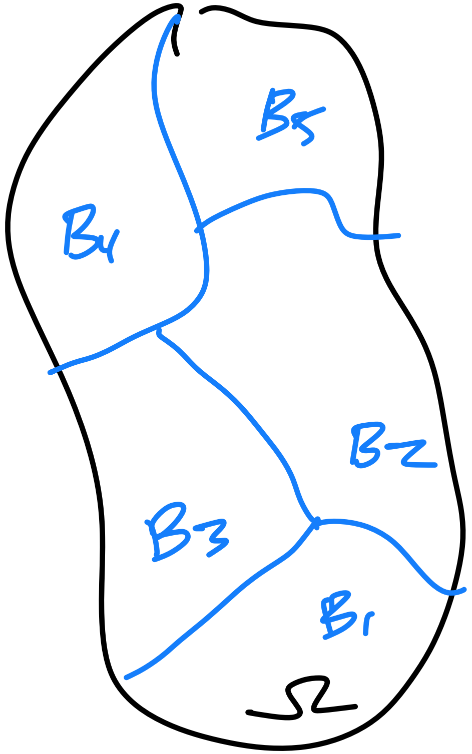
$(52 \text{ choose } 13)$ about $6.35 * 10^{11}$

```
Counting[n_] := Module[{},
    count = Binomial[4*n, n];
    Print["Counting to ", count];
    For[k = 1, k ≤ count, k++, x = 1]
   ];

Timing[Counting[8]]
Timing[Counting[9]]
Timing[Counting[10]]

Counting to 10 518 300

{3.4375, Null}

Counting to 94 143 280

{34.6406, Null}

Counting to 847 660 528

{346.422, Null}
```

Law of Total Probability: If $\{B_1, B_2, \dots\}$ form a partition for the sample space S (into at most countably many pieces), then for any $A \subset S$ we have

$$\Pr(A) = \sum_n \Pr(A|B_n) \cdot \Pr(B_n).$$

We should have $0 < \Pr(B_n) < 1$ for all n as the conditional probabilities aren't defined otherwise (note if a B_n has probability zero then it isn't needed, as that piece is hit by the factor $\Pr(B_n) = 0$, while if it is 1 then all the other factors are unnecessary).



Bayes' Theorem: Let $\{A_i\}_{i=1}^n$ denote a partition of the sample space. Then

$$\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\sum_{i=1}^n \Pr(B|A_i) \cdot \Pr(A_i)}.$$

Frequently one takes A to be one of the sets A_i.

denom is $P(B)$

$\Omega = \bigsqcup_{i=1}^{5} B_i$

disjoint union

$\sum_{i=1}^{5} P_r(B_i) = 1$

The At Least to Exactly Method: Let $N(k)$ be the number of ways for *at least* k things to happen, and let $E(k)$ be the number of ways for *exactly* k things to happen. Then $E(k) = N(k) - N(k+1)$. Equivalently,

$$\begin{aligned}&\text{Prob(exactly } k \text{ happen)}\\ =\ &\text{Prob(at least } k \text{ happen)} - \text{Prob(at least } k+1 \text{ happen)}.\end{aligned}$$

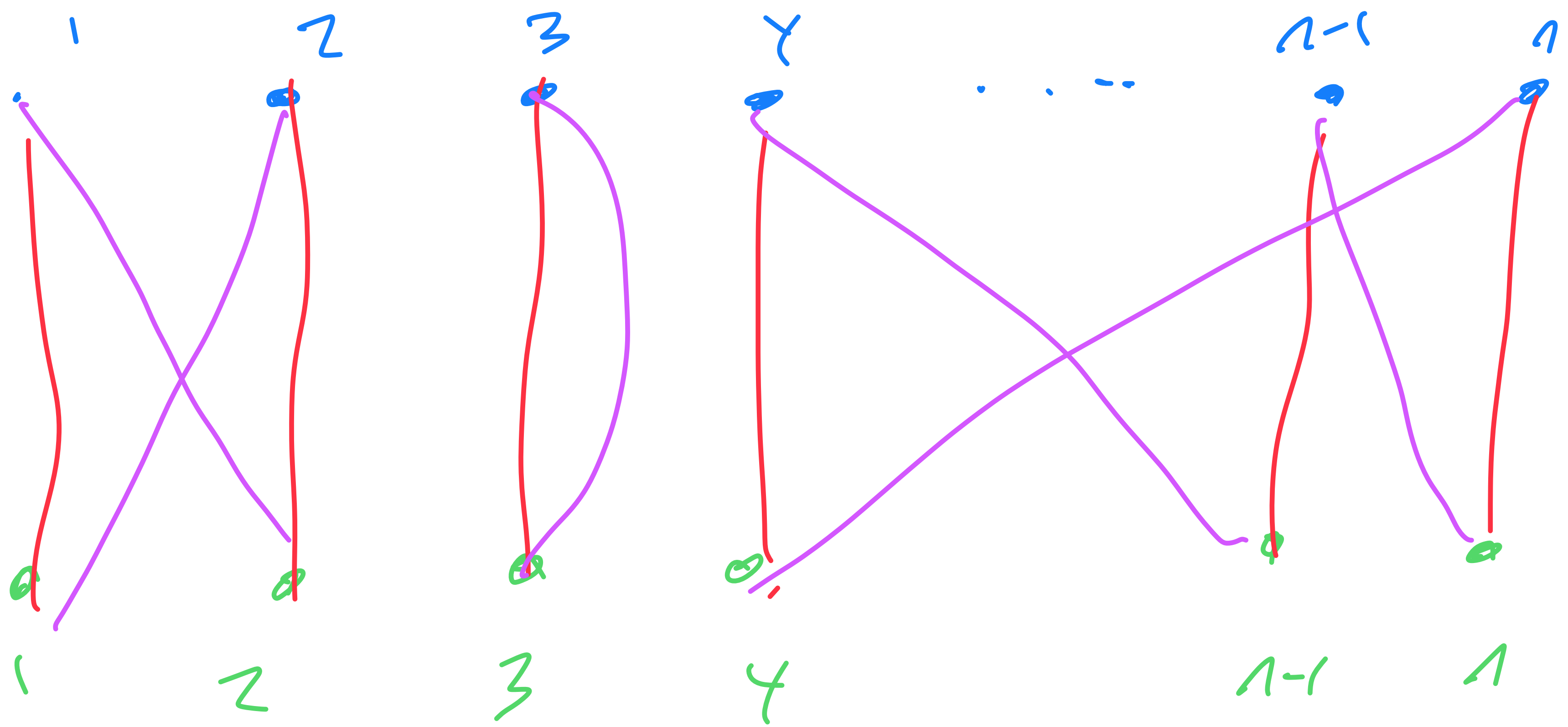
5.3.1 Counting Derangements

So, how many of the $n!$ orderings have no element returned to where it starts? This means the 1st element cannot be in the first spot, nor the 2nd element in the second spot, and so on. For example, $\{2, 3, 4, 1\}$ is a derangement as each number is moved, while $\{3, 2, 4, 1\}$ is not a derangement as 2 is in the second position.

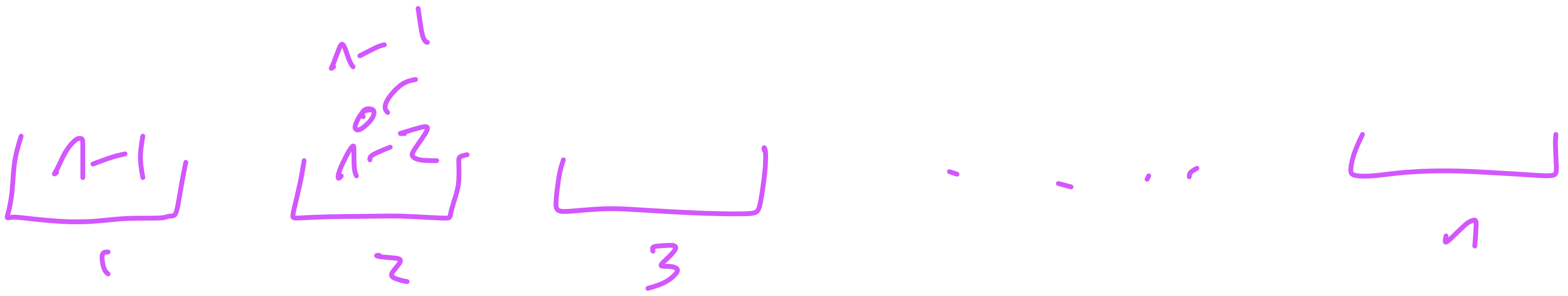
It turns out to be much easier to look at the related problem, where we count how many ways there are for at least one element to return to its starting point. Why is this easier? Remember the statement of the inclusion-exclusion principle (see §5.2.2). We show how to write an *at least* event in terms of intersections of events, and intersections are often easy to compute. To get the number of derangements, we just subtract the number of non-derangements from $n!$.

Derangements: Everything moves..... (Application: Graph Theory)

$\{1, 2, \ldots, n\}$ reorder to $\{k_1, k_2, \ldots, k_n\}$ require $k_\ell \neq \ell$



2-regular graph: each vertex has two edges from it



Cannot be a 1

n-1 if 2 is in first spot, else n-2

at least one is fixed
at most one is fixed
exactly one is fixed

do for a specific number/spot

Let A_k be the event k is fixed, don't care about anything else

$$\Pr(A_1) = \Pr(A_2) = \cdots = \Pr(A_n) = \frac{1 \cdot (n-1)!}{n!} = \frac{1}{n}$$

$$\Pr(A_{12}) = \Pr(A_{13}) = \cdots = \Pr(A_{n-1,n}) = \frac{1 \cdot 1 \cdot (n-2)!}{n!} = \frac{1}{n(n-1)}$$

↑ 1 and 2 fixed
do not care about rest

$$Pr(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i_1=1}^{n} Pr(A_{i_1}) - \sum_{1 \le i_1 < i_2 \le n} Pr(A_{i_1, i_2}) + \cdots$$

$$= n \frac{1}{n} - \binom{n}{2} \frac{1}{n(n-1)} + \binom{n}{3} \frac{1}{n(n-1)(n-2)} + \cdots + (-1)^{n-1} \binom{n}{n} \frac{1}{n!}$$

$$= 1 - \frac{n(n-1)}{2!\, n(n-1)} + \frac{n(n-1)(n-2)}{3!\, n(n-1)(n-2)} + \cdots + (-1)^{n-1} \frac{n!}{0!\, n! \cdot n!}$$

$$= -\frac{1}{0!} + \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + (-1)^{n-1} \frac{1}{n!} + \frac{1}{0!}$$

$$= 1 - \left(\frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n} \frac{1}{n!} \right)$$

$$\simeq 1 - e^{-1} \quad \text{as } e^{x} = \sum_{m=0}^{\infty} x^{m}/m!$$

$\Rightarrow$ Prob of a derangement converges to $1/e$ as $n \to \infty$

```
derangement[n_, numdo_] := Module[{},
  count = 0; (* set counter of successes to 0 *)
  (* prediction for finite n and n --> oo *)
  theory = Sum[(-1)^k/k!, {k, 0, n}];
  limit = 1/E;
  people = {}; (* this will be our list of people *)
  For[i = 1, i <= n, i++, people = AppendTo[people, i]];
  For[m = 1, m <= numdo, m++, (* main loop *)
   {
    mix = RandomSample[people]; (* randomly mix people *)
    found = 0; (* set found to 0, if becomes 1 someone fixed *)
    For[i = 1, i <= n, i++,
     {
      If[mix[[i]] == i,
        {
         found = 1;
         i = n + 1; (* exit loop: why keep computing! *)
         }];
      }];
    If[found == 1, count = count + 1]; (* if found=
    1 increase count *)
    }];
 Print["Theory is ", 100.  theory, "%."];
 Print["Limit is ", 100.  limit, "%."];
 (* we want prob derangement so it's 1 - prob someone fixed *)
 Print["Observe ", 100. - 100. count/numdo, "%."];
 ]
```

For example, if we run ten million trials with $n = 5$ we find

```
Theory is 36.6667%.
Limit is 36.7879%.
Observe 36.6783%.
```

$$n = 20,$$

```
Theory is 36.7879%.
Limit is 36.7879%.
Observe 36.8023%.
```

Coding

Sudoku:

1, 10, 100, 1000, ...

sums to 111,111,111

Multinomial Coefficients

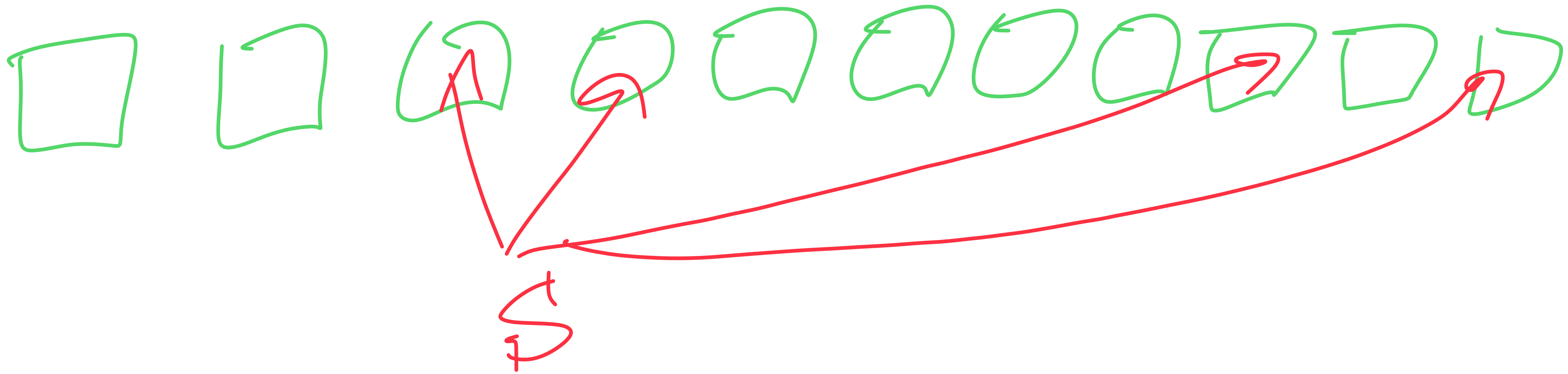
$$\underbrace{M_1 I_1 S_1 S_2 I_2 S_3 S_4 I_3 P_1 P_2 I_4}$$

$$\frac{11!}{4!\,4!\,2!\,1!}$$

$$\underline{NEWYORK}$$

$\underbrace{7!}$ distinct "words"

$$\frac{7!}{1!\,1!\,1!\cdots1!}$$



$$\binom{n}{k_1 \; k_2 \; \cdots \; k_\ell}$$

where $0 \le k_1, k_2, \ldots, k_\ell \le n$

and $k_1 + k_2 + \cdots + k_\ell = n$

$$\binom{n}{k} = \binom{n}{k}\binom{n-k}{n-k} \qquad (x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

$$\binom{n}{k \;\; n-k} = \frac{n!}{k!\,(n-k)!} = \binom{n}{k}$$

so reduces to binomial coeff

recovers

$$(x+y+z)^n = ((x+y)+z)^n$$

$$= \sum_{k=0}^{n} \binom{n}{k} (x+y)^k z^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} \sum_{l=0}^{k} \binom{k}{l} x^l y^{k-l} z^{n-k}$$

$$= \sum_{k=0}^{n} \sum_{l=0}^{k} \underbrace{\binom{n}{k}\binom{k}{l}}_{\frac{n!}{\cancel{k!}\,(n-k)!} \frac{\cancel{k!}}{l!\,(k-l)!}} x^l y^{k-l} z^{n-k}$$

$$= \frac{n!}{l!\,(k-l)!\,(n-k)!}$$

$$= \binom{n}{l,\ k-l,\ n-k}$$

Cookie Problem

C cookies: identical

P people: distinct

How many ways to divide <u>all</u> the cookies among the people?

$C = 10 \qquad P = 5$

A	B	C	D	E	
10	0	0	0	0	5
9	1	0	0	0	20
8	2	0	0	0	20
8	1	1	0	0	$5 \cdot \binom{4}{2} = 30$

$$\frac{5!}{1!\,1!\,3!} = 5 \cdot 4 = 20$$

$$\frac{5!}{1!\,2!\,2!} = 30$$

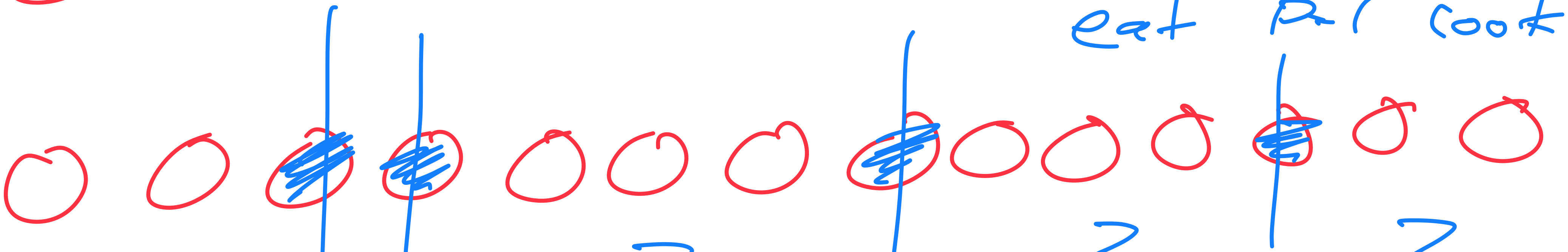
Cookie Monster, photo by the author from the National Bobblehead Hall Of Fame And Museum in Milwaukee, with thanks to the museum staff for their kindness in posing: https://www.bobbleheadhall.com/

Cookie Problem / Stars and Bars

$C = 10$ $\quad$ $P = 5$

$C + P - 1 = 14$ cookies

eat $P - 1$ cookies



2	0	3	3	2
A	B	C	D	E

$$\binom{14}{4} = 1001$$

Solved $x_1 + x_2 + x_3 + x_4 + x_5 = 10$

each x_i non-neg int; #solns is $\binom{14}{4}$

General: $x_1 + \cdots + x_p = C$ #solns is $\binom{C+p-1}{p-1}$

If everyone must get at least 1 cookie:

$$x_k = y_k + 1 \Rightarrow y_1 + \cdots + y_p = C - p$$

#solns is $\binom{C-p+p-1}{p-1}$

What if don't have to distribute all cookies?

$$\sum_{c=0}^{C} \binom{c+p-1}{p-1} = \binom{C+p+1-1}{p+1-1}$$

add the prof to get cookies

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 8: 3/6/25: Induction, PDF/CDF: https://youtu.be/N5jLFMOyW8A

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 8: March 4, 2025:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- Induction
- Basics of PDFs
- Random Variables: Continuous (FTC) vs Discrete
- Moments and Expected Values

General items.

- Rescale
- More coding: https://youtu.be/sSgjBysixdQ; code file here, pdf here

INDUCTION:

Statement $P(n)$

Base Case: Show $P(0)$ is true

Inductive Step: Show if $P(n)$ true then $P(n+1)$ is true.

Then can conclude $P(n)$ true for all $n \in \{0, 1, 2, \ldots\}$

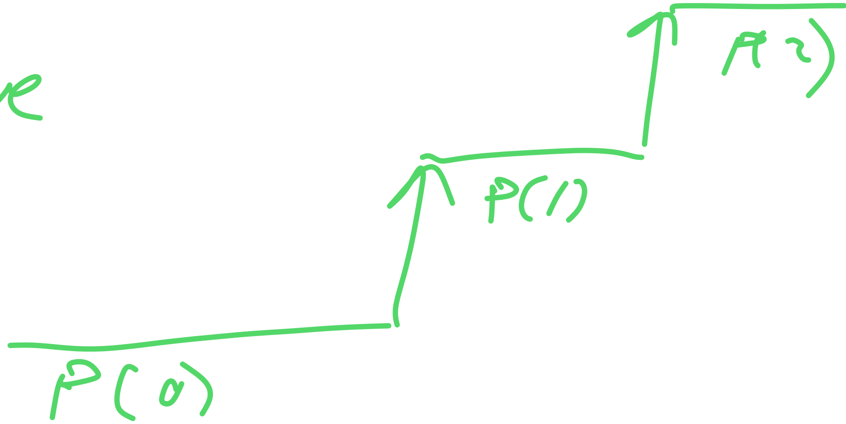
Staircase: $P(0)$ true

$P(0)$ true $\rightarrow$ $P(0+1)$ true

$\Longrightarrow$ $P(1)$ true

$P(1)$ true $\rightarrow$ $P(1+1)$ true

$\Longrightarrow$ $P(2)$ true



$0+1+2+\cdots+n = n(n+1)/2$: Proof by Induction

Statement $P(n)$: $0+1+2+\cdots+n = n(n+1)/2$

Base Case: is $P(0)$ true?

Is $0 \stackrel{?}{=} 0(0+1)/2$ YES

Inductive Step: Show if $P(n)$ true then $P(n+1)$ is true

Since $P(n)$ true can assume

$$0+1+\cdots+n = n(n+1)/2$$

add $n+1$

$$(0+1+\cdots+n)+(n+1) = \frac{n(n+1)}{2} + \frac{(n+1)2}{2} = \frac{(n+1)(n+2)}{2}$$

$$= \frac{(n+1)(n+1+1)}{2}$$

Shows $P(n+1)$ is true, done!

$P(n+1)$ is $\underbrace{0+1+\cdots+n}_{\text{This is } P(n)}+(n+1) \stackrel{?}{=} \frac{(n+1)(n+1+1)}{2}$

$$P(n):\ \underbrace{0^2+1^2+2^2+\cdots+n^2}_{\sum_{k=0}^{n} k^2 \approx \int_0^n x^2\,dx} = \underbrace{\frac{n(n+1)(2n+1)}{6}}_{\text{leading term is}}$$

$$\sum_{k=0}^{n} k^2 \approx \int_0^n x^2\,dx = \frac{x^3}{3}\Big|_0^n = \frac{n^3}{3}$$

leading term is

$$\frac{2n^3}{6} = \frac{n^3}{3}$$

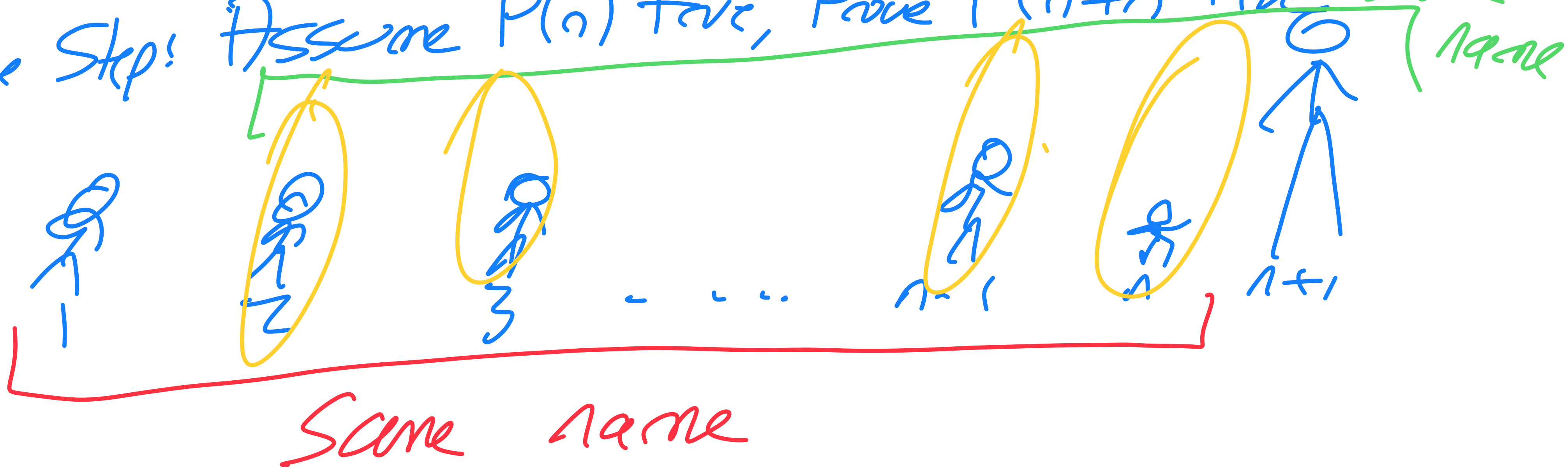
Thm: Everyone in the world is named Yanni

P(n): in any group of n people, all have same name.

~~Base Case: P(0) true?~~

Base Case: P(1) true? Yes!

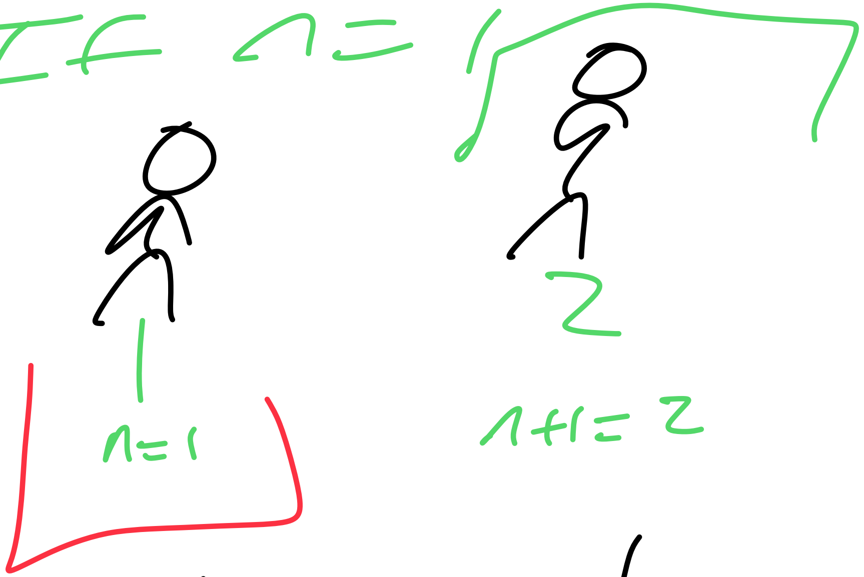
Inductive Step: Assume P(n) true, Prove P(n+1) true



all same name as person 2, so all same name.

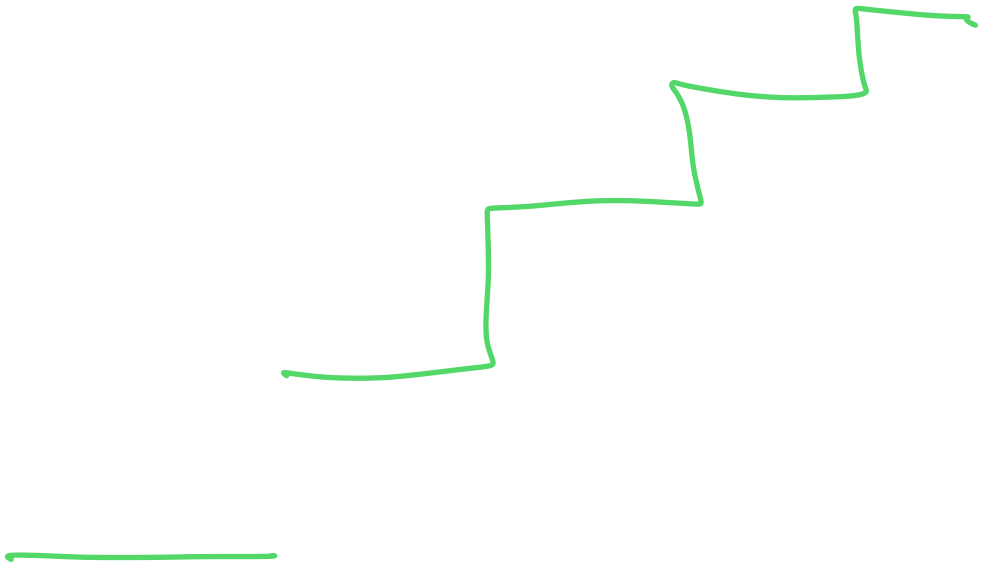
Issue: Bad Picture:

If $n=1$



disjoint!

Note: if $P(2)$ is true is $P(n)$ true for all $n \geq 2$? YES!



More on proofs:
https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/handouts/proofs.html

Probability density: Definition, Examples

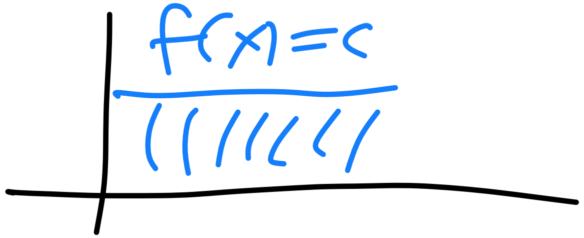
$f(x)$ is a density or a probability density function (pdf)

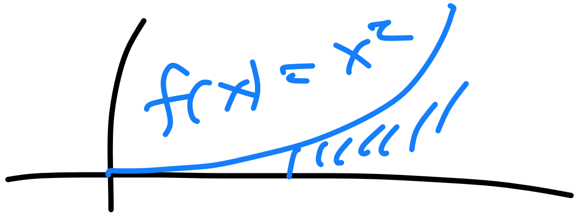
if $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (continuous case)

or $f(x_n) \geq 0$ and $\sum_{n=0}^{\infty} f(x_n) = 1$ (discrete case)

↳ in the discrete case either have a finite set or a countable set,

Fund Thm of Calc: $\int_a^b f(x)\,dx = F(b) - F(a)$, $F' = f$





use induction to do the Riemann sums

Simplest: Head / Tail

$p(H) = 1/2$ $\quad p(T) = \frac{1}{2}$ fair coin

$p(1) = \frac{1}{2}$ $\quad P(0) = \frac{1}{2}$

$P(1) = \frac{1}{2}$ $\quad P(-1) = \frac{1}{2}$

Binary, Bernouli: $\underline{X} \sim \text{Bern}(p)$

Mean $\text{Prob}(\underline{X} = 1) = p$

and $\text{Prob}(\underline{X} = 0) = 1-p$

\frac{\overline{\Xi}}{\Xi}

n indep tosses of a coin with prob p of heads, let X be the random variable of the # of heads.

$\text{Prob}(X = k)$ is 0 if $k \notin \{0, 1, \dots, n\}$ and it's $\binom{n}{k} p^k (1-p)^{n-k}$

$$\text{Prob}(\underset{p}{H}\,\underset{p}{H}\,\underset{1-p}{T}\,\underset{1-p}{T}\,\underset{p}{H}\,\underset{1-p}{T}\,\underset{1-p}{T}\,\underset{1-p}{T}\,\underset{p}{H}) = p^4(1-p)^5$$

Binomial Random Variable: $X \sim \text{Bin}(n, p)$

Let $X_k \sim \text{Bern}(p)$ is the R.V. of a head on toss k.

Then $X = X_1 + \cdots + X_n$

Exponential Random Variable with parameter λ (or ω)

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} \quad \underline{\underline{\text{or}}} \quad \omega e^{-\omega x} \qquad x \geq 0$$

Ave value / expected value / mean is

$$E[X] := \int_{-\infty}^{\infty} x f(x)\,dx \quad \text{or} \quad \sum_{n=0}^{\infty} x_n f(x_n)$$

$$E[X] = \int_0^\infty x \cdot \frac{1}{\lambda} e^{-x/\lambda}\,dx = \lambda \qquad E[X] = \int_0^\infty x\omega e^{-\omega x}\,dx$$

$$t = x/\lambda \text{ so } dt = \frac{dx}{\lambda} \quad \vdots \quad x: 0 \to \infty$$
$$: 0 \to \infty \qquad = \frac{1}{\omega}$$

$$= \int_0^\infty t e^{-t}\,dt$$

$$u = t \qquad dv = e^{-t}\,dt$$
$$du = dt \qquad v = -e^{-t}$$

$$= -te^{-t}\Big|_0^\infty + \int_0^\infty e^{-t}\,dt = 1$$

Probability density of Y = g(X) in terms of X and its density

X ~ Uniform(0,1) and Y = X^2.

$X \sim \text{Unif}(a,b)$ means $p(x) = \frac{1}{b-a}$ if $a \le x \le b$ and 0 otherwise

The cumulative distribution function (cdf) is the prob that X is at most x:

If f is the density F is the cdf: f_X and F_X

$$F_X(x) = \text{Prob}(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt$$

$$- F_X(-\infty)$$

This is zero so don't write

note $F_X(-\infty) = 0$

take deriv

$$F_X'(x) = \frac{d}{dx}\int_{-\infty}^{x} f_X(t)\,dt = \frac{d}{dx}\left[F_X(x) - \underbrace{F_X(-\infty)}_{=0}\right]$$

$$= \frac{d}{dX} F_X(x) = f_X(x)$$

$$\frac{F_X(x+h) - F_X(x)}{h} = \frac{1}{h}\int_{x}^{x+h} f_X(t)\,dt \approx \frac{1}{h} f_X(x)\,h$$

$X \sim \text{Unif}(0,1)$ so $f_X(x) = 1$ if $0 \le x \le 1$ and 0 otherwise

$Y = X^2$ what is $f_Y(y)$?

$X^2 \le y$

same as

$-\sqrt{y} \le X \le \sqrt{y}$

$$F_Y(y) = \text{Prob}(Y \le y) = \text{Prob}(X^2 \le y)$$

$$= \text{Prob}(-\sqrt{y} \le X \le \sqrt{y})$$

$$= \text{Prob}(X \le \sqrt{y})$$

$$= \int_0^{\sqrt{y}} f_X(t)\,dt = \int_0^{\sqrt{y}} 1\,dt = \sqrt{y}$$

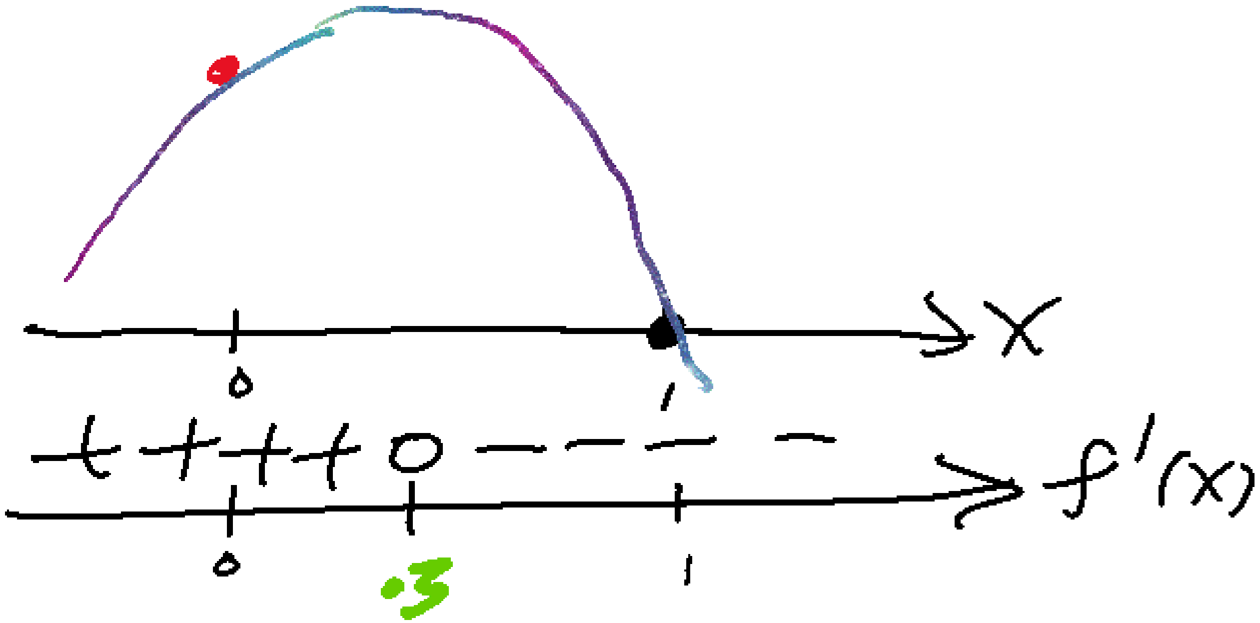
Shown $F_Y(y) = \sqrt{y}$ so $f_Y(y) = \frac{1}{2} y^{-1/2}$ for $0 \le y \le 1$

Check: $\int_0^1 \frac{1}{2} y^{-1/2} dy = \frac{1}{2} \frac{y^{1/2}}{1/2}\Big|_0^1 \quad \frac{1}{2}\left[\frac{1}{1/2} - \frac{0}{1/2}\right] = 1$

$\frac{1}{2} y^{-1/2} \geq 0$

Called the CDF Method

Example: $f(x) = 2 + 3x - 5x^2$ for x in [0,1]: Curve sketching



$$f(x) = 2 + 3x - 5x^2$$

$$f'(x) = 3 - 10x$$

$$f(0) = 2 \qquad f(1) = 0$$

Critical point: $f'(x) = 0$

so $3 - 10x = 0 \rightarrow x = \frac{3}{10}$

$$\int_0^1 (2 + 3x - 5x^2)\,dx = \left[2x + \frac{3x^2}{2} - \frac{5x^3}{3}\right]_0^1$$

$$= 2 + \frac{3}{2} - \frac{5}{3} = \frac{12 + 9 - 10}{6} = \frac{11}{6}$$

Not a density but... $\frac{6}{11} f(x)$ is a density

If $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)\,dx$ is finite $\neq 0$

Then $g(x) = \dfrac{f(x)}{\int_{-\infty}^{\infty} f(t)\,dt}$

is a probability density

$\rightarrow$ Renormalizing

Sum of two independent, fair die…. Uniform + Uniform = Triangle

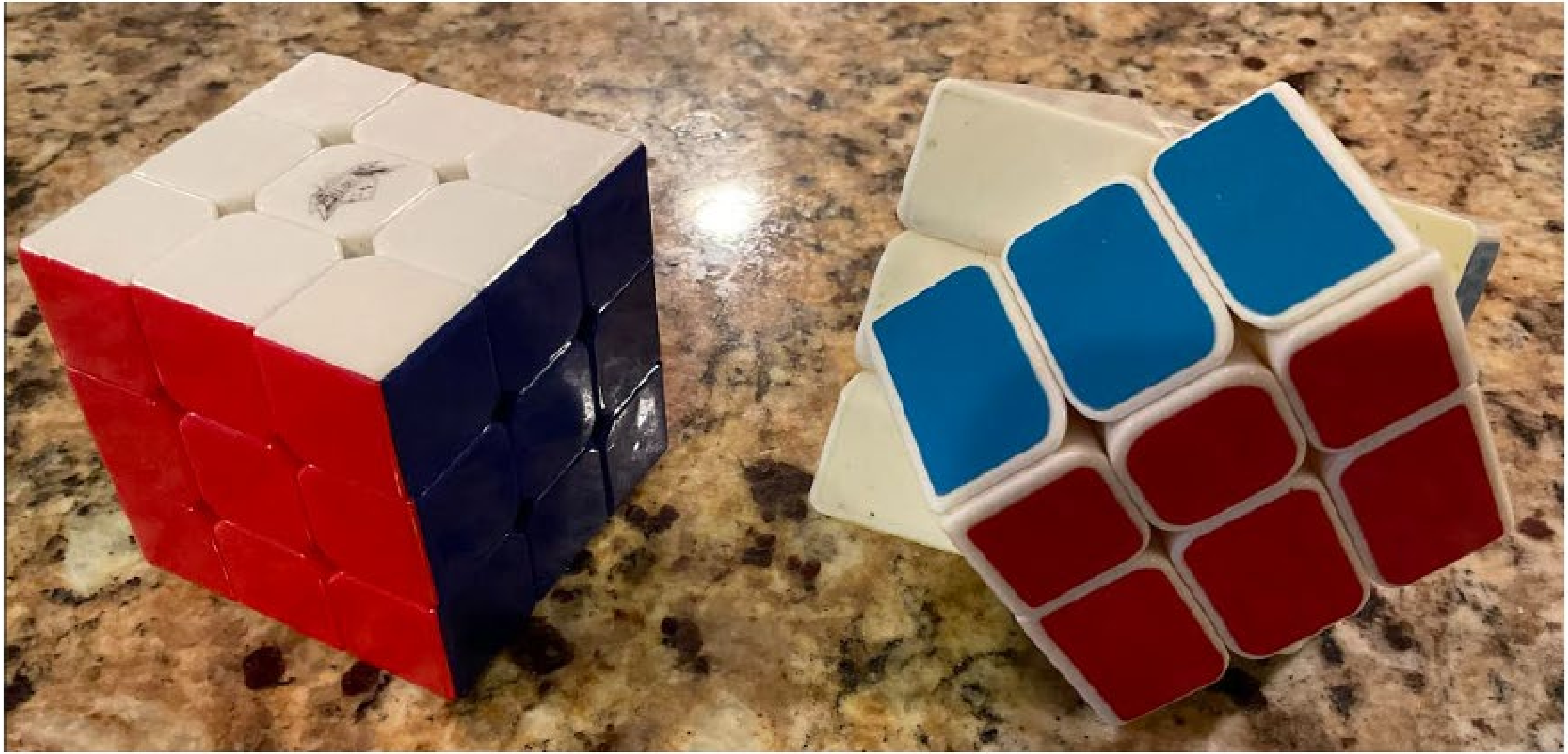
sum	
2	(1,1)
3	(1,2), (2,1)
4	(1,3), (2,2), (3,1)
5	
6	
7	
8	
9	
10	(4,6), (5,5), (6,4)
11	(5,6), (6,5)
12	(6,6)

Ex: write down a formula for sum of two indep fair die

$f_X(7) = 1/6$

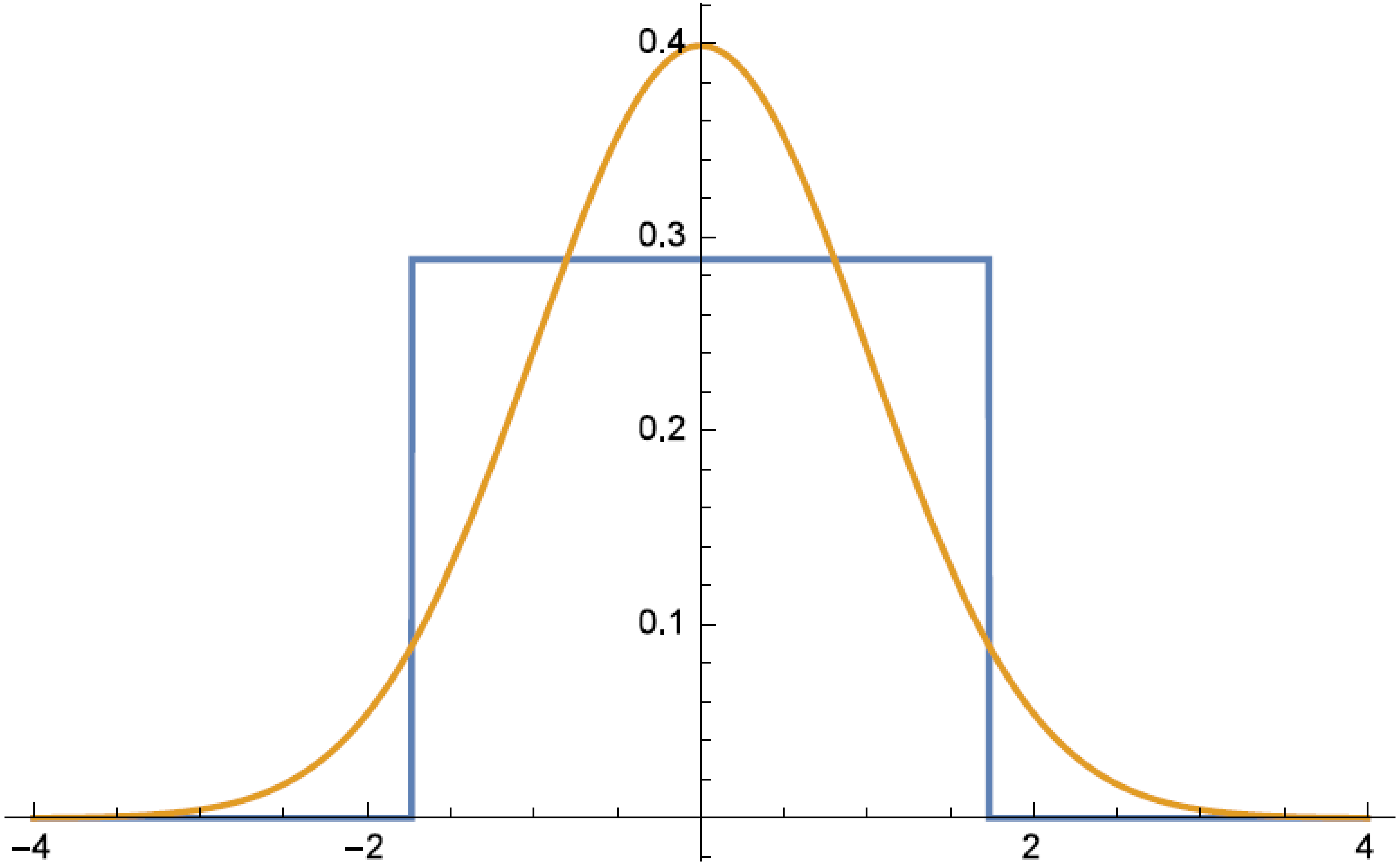
$f_X(2) = f_X(12) = \frac{1}{36}$

The Power of the Right Perspective



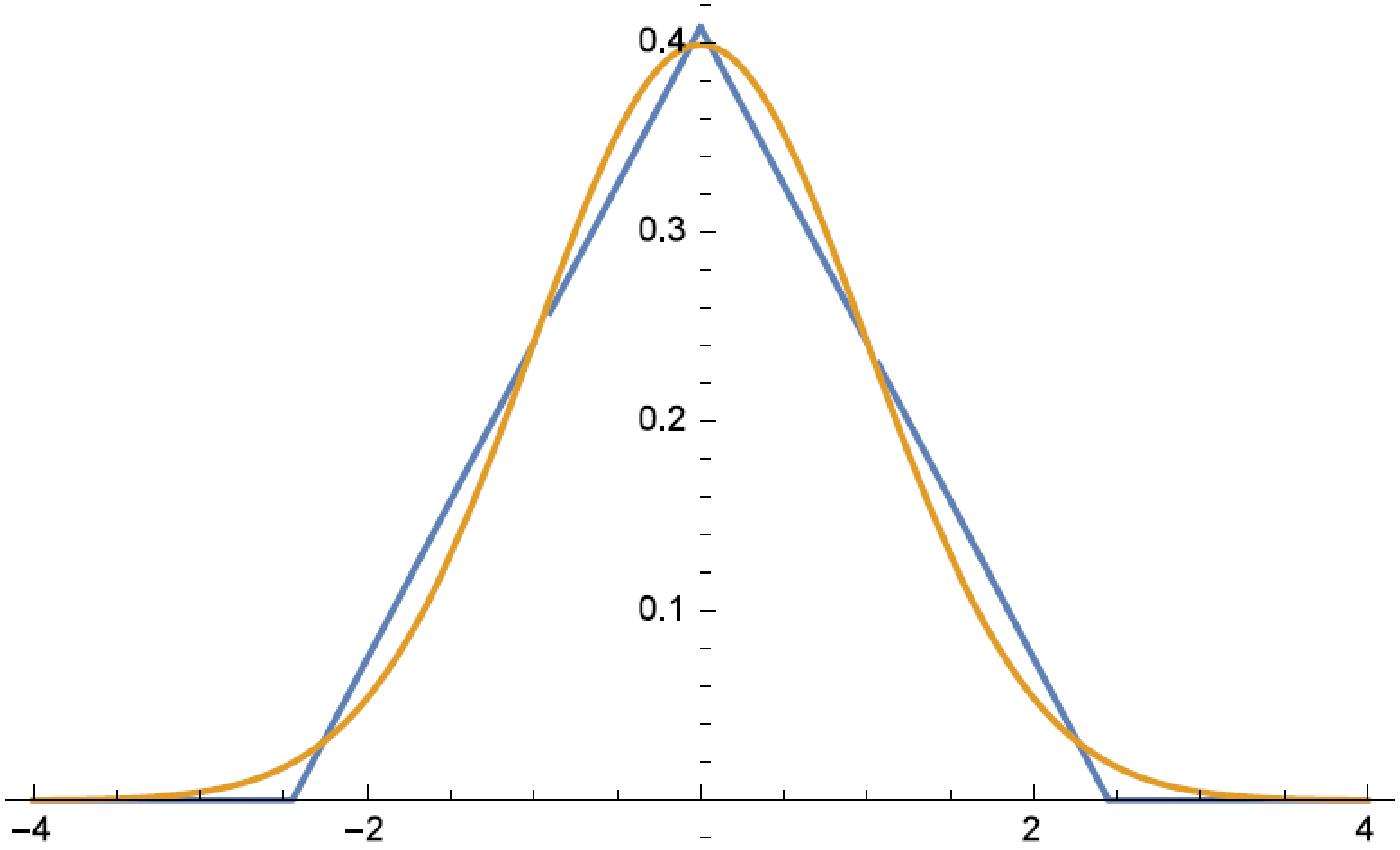
$X_i \sim \mathrm{Unif}(-1/2, 1/2)$ (adjusted to mean 0, variance 1)

$Y_1 = X_1/\sigma_{X_1}$ vs $N(0, 1)$.



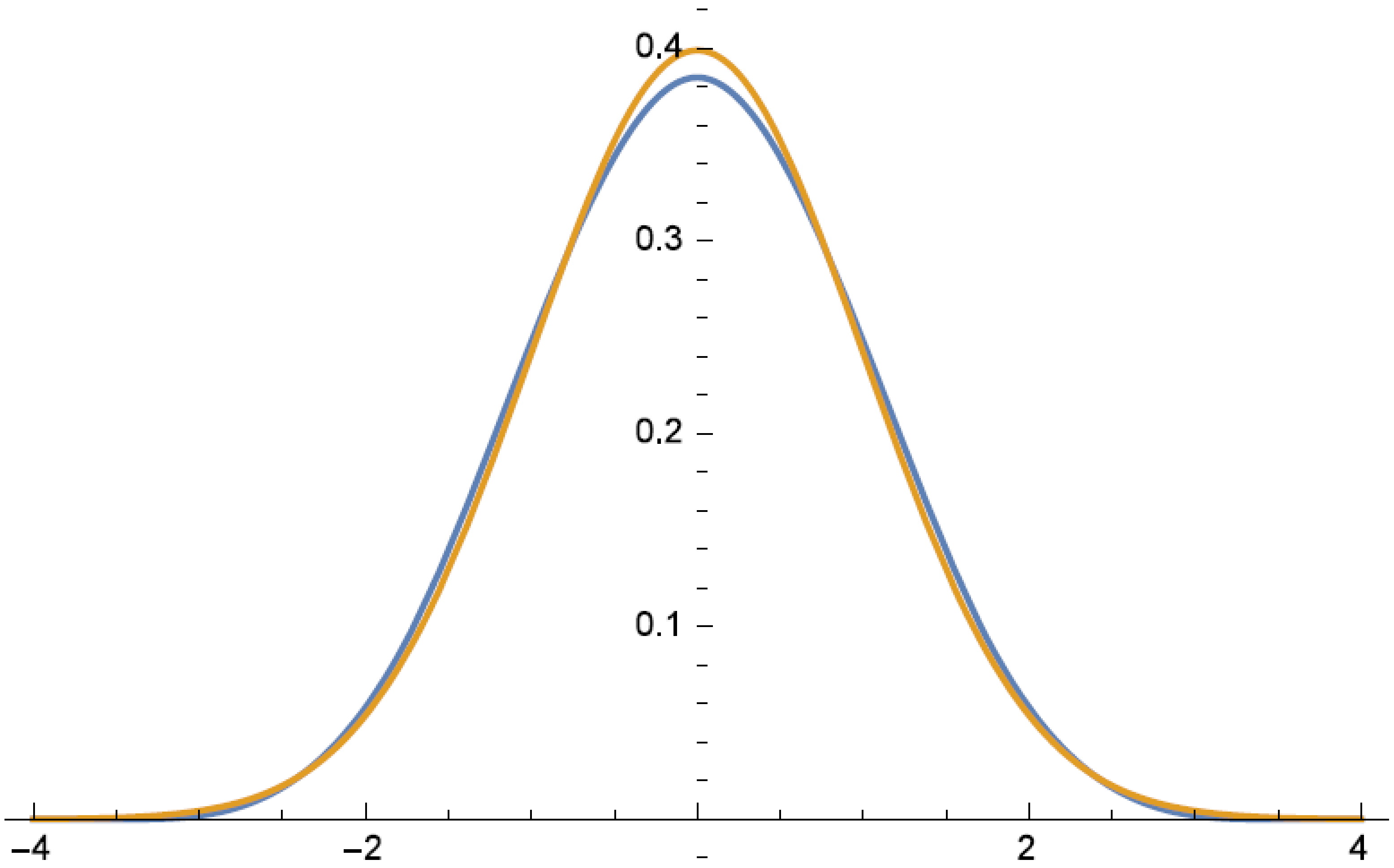
$X_i \sim \mathrm{Unif}(-1/2, 1/2)$ (adjusted to mean 0, variance 1)

$$Y_2 = (X_1 + X_2)/\sigma_{X_1+X_2} \text{ vs } N(0,1).$$



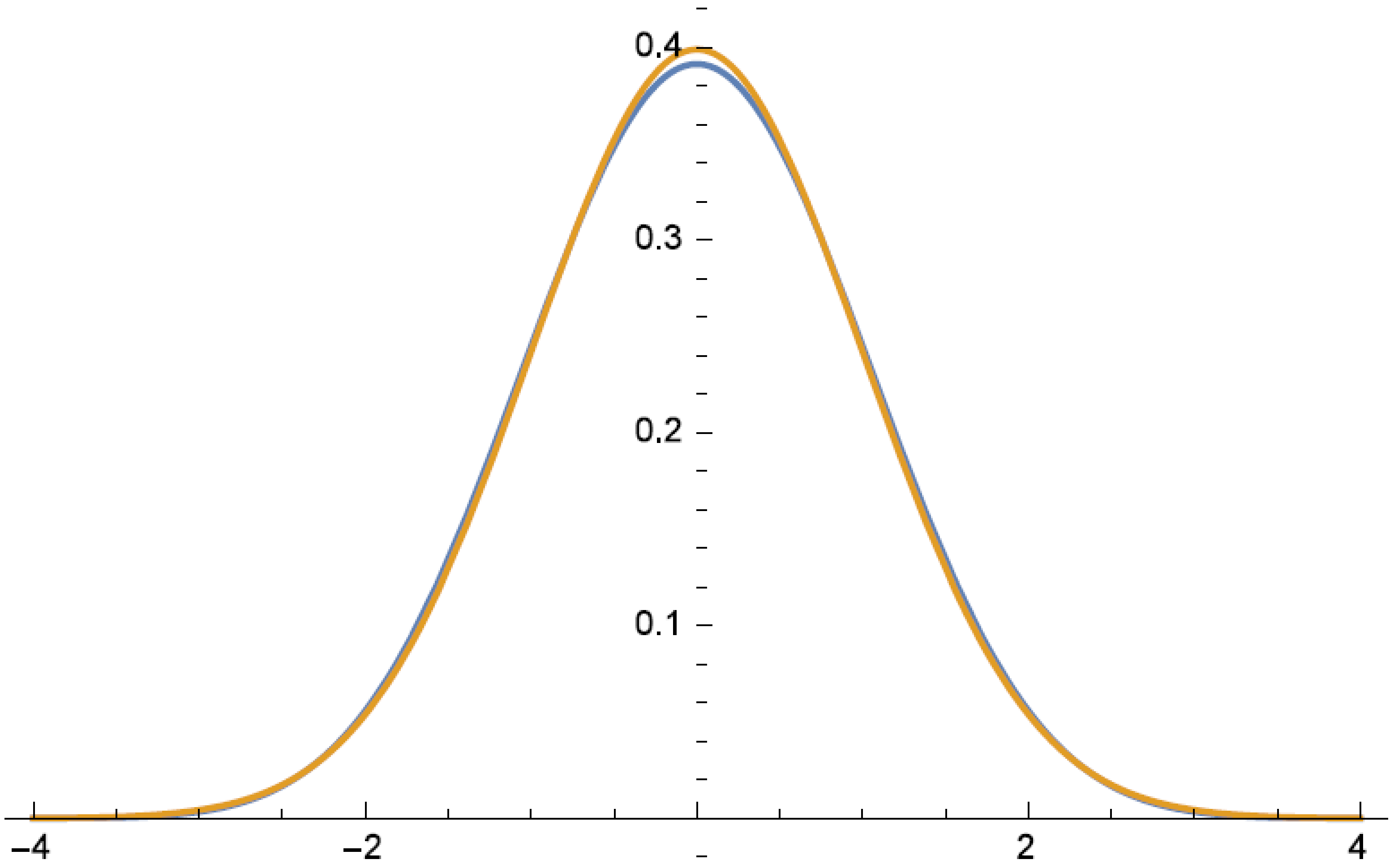
$X_i \sim \text{Unif}(-1/2, 1/2)$ (adjusted to mean 0, variance 1)

$Y_4 = (X_1 + X_2 + X_3 + X_4)/\sigma_{X_1+X_2+X_3+X_4}$ vs $N(0, 1)$.



$X_i \sim \mathrm{Unif}(-1/2, 1/2)$ (adjusted to mean 0, variance 1)

$Y_8 = (X_1 + \cdots + X_8)/\sigma_{X_1+\cdots+X_8}$ vs $N(0,1)$.



$X_i \sim \mathrm{Unif}(-1/2, 1/2)$ (adjusted to mean 0, variance 1)

Density of $Y_4 = (X_1 + \cdots + X_4)/\sigma_{X_1+\cdots+X_4}$.

$$\frac{\begin{cases} \frac{1}{27}\left(18 + 9\sqrt{3}\, y - \sqrt{3}\, y^3\right) & y == 0 \\ \frac{1}{18}\left(12 - 6\, y^2 - \sqrt{3}\, y^3\right) & -\sqrt{3} < y < 0 \\ \frac{1}{54}\left(72 - 36\sqrt{3}\, y + 18\, y^2 - \sqrt{3}\, y^3\right) & \sqrt{3} < y < 2\sqrt{3} \\ \frac{1}{54}\left(18\sqrt{3}\, y - 18\, y^2 + \sqrt{3}\, y^3\right) & y == \sqrt{3} \\ \frac{1}{18}\left(12 - 6\, y^2 + \sqrt{3}\, y^3\right) & 0 < y < \sqrt{3} \\ \frac{1}{54}\left(72 + 36\sqrt{3}\, y + 18\, y^2 + \sqrt{3}\, y^3\right) & -2\sqrt{3} < y \le -\sqrt{3} \\ 0 & \mathrm{True} \end{cases}}{\sqrt{3}}$$

(Don't even think of asking to see Y_8's!)

Moments and expected value of functions of random variables

k^{th} moment of X with density f_X is

$$E[X^k] := \int_{-\infty}^{\infty} x^k f_X(x)\,dx$$

mean, denote μ, is $E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$

Variance of X, denoted σ^2, is $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx$

algebra: $\sigma^2 = E[(X-\mu)^2] = E[X^2] - E[X]^2$

where $E[g(X)] := \int_{-\infty}^{\infty} g(x)\, f_X(x)\,dx.$

Standard Deviation is $\sigma = \sqrt{\sigma^2}$: Same units as mean.

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 9: 3/6/25: CDF Method, Cauchy Distribution, Joint Distribution, Marginal Distribution, Convolutions, Sums of Random Variables, Linearity of Expectation.
Video: https://youtu.be/gXubfwC0vJw

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 9: March 6, 2025:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- Review CDF Method
- Cauchy Distribution
- Joint PDF
- Linearity of Expectation
- Fermat Primes
- Buffon’s Needle

General items.

- Power of Linearity
- Avoiding brute force computations

CDF Method: Integrating without integrating….

$Y = X^2$, X non-negative random variable. $\text{Prob}(X < 0) = 0$

$$F_Y(y) = \text{Prob}(Y \le y)$$
$$= \text{Prob}(X^2 \le y)$$
so $-\sqrt{y} \le X \le y$ as $X \ge 0$
$$= \text{Prob}(X \le \sqrt{y})$$
CDF of X being at most $x = \sqrt{}$
$$= F_X(\sqrt{y})$$
Showed pdf is the deriv of the cdf

$$F_Y'(y) = \frac{d}{dy}\left[F_X(\sqrt{y})\right]$$
"of" Think chain rule
$$(A(B(x)))' = A'(B(x)) \cdot B'(x)$$
$$f_Y(y) = F_X'(\sqrt{y}) \cdot (\sqrt{y})'$$
$$= f_X(\sqrt{y}) \, \frac{1}{2} y^{-1/2}$$

do not need closed form expression for CDF!

Divine Inspiration: Review Trig

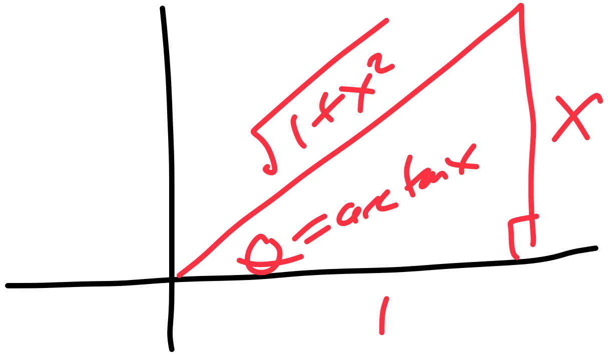
Imagine f, g inverses so $A(x) = f(g(x)) = x$

Then $A'(x) = f'(g(x))\, g'(x) = 1 \Rightarrow g'(x) = \dfrac{1}{f'(g(x))}$

Consider: $\tan(\arctan x) = x \qquad \Rightarrow \arctan'(x) = \dfrac{1}{\tan'(\arctan x)}$

$\tan x = \dfrac{\sin x}{\cos x} \Rightarrow \tan' x = \dfrac{\cos x \cos x - \sin x(-\sin x)}{\cos^2 x} = \dfrac{1}{\cos^2 x} = \sec^2 x$

Thus $\arctan'(x) = \cos^2(\arctan x)$



$\cos \qquad \cos(\arctan x) = \dfrac{1}{\sqrt{1+x^2}}$

So $\cos^2(\arctan x) = \dfrac{1}{1+x^2} = \arctan'(x)$

$$\int \frac{1}{1+x^2}\, dx = \arctan x$$

Cauchy Distribution

Study $f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ non-neg; does it integrate to 1?

$$\int_{-\infty}^{\infty} \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{2}{\pi} \int_0^{\infty} \frac{1}{1+x^2} dx = \frac{2}{\pi} \int_0^{\infty} \arctan'(x) dx$$

$$= \frac{2}{\pi} \Big[\underbrace{\arctan(\infty)}_{\frac{\pi}{2}} - \underbrace{\arctan(0)}_{=0} \Big] = 1$$

What is your mean?

What's your standard deviation?

Mean: $\displaystyle\int_{-\infty}^{\infty} x \cdot \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \underbrace{\frac{x}{1+x^2}}_{\text{odd}} dx = 0$

NO!

Need $\int_{-\infty}^{\infty} |x f_X(x)|\,dx$ to be finite: not the case here!

Consider $\lim_{A,B\to\infty} \int_{-A}^{B} x \frac{1}{\pi}\frac{1}{1+x^2}\,dx$ (if $A=B$ get 0)

If $B=2A$: $\lim_{A\to\infty} \int_{-A}^{2A} x \frac{1}{\pi}\frac{1}{1+x^2}\,dx = \frac{1}{\pi}\lim_{A\to\infty}\int_{A}^{2A}\frac{x}{1+x^2}\,dx$

$$\approx \frac{1}{\pi}\lim_{A\to\infty}\int_{A}^{2A}\frac{1}{x}\,dx$$

$$= \frac{1}{\pi}\lim_{A\to\infty}\left(\ln 2A - \ln A\right)$$

$$= \frac{1}{\pi}\lim_{A\to\infty}\ln 2 \neq 0$$

Diff of logs is the log of the quotient

$\log \alpha - \log \beta = \log\frac{\alpha}{\beta}$

2nd Moment $\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\pi}\frac{1}{1+x^2}\,dx = \infty$

Joint probability density function. Let $X_1, X_2, \ldots, X_n$ be continuous random variables with densities $f_{X_1}, f_{X_2}, \ldots, f_{X_n}$. Assume each X_i is defined on a subset of $\mathbb{R}$ (the real numbers). The joint density function of the tuple $(X_1, \ldots, X_n)$ is a non-negative, integrable function $f_{X_1,\ldots,X_n}$ such that, for every nice set $S \subset \mathbb{R}^n$ we have

$$\text{Prob}\,((X_1, \ldots, X_n) \in S) \;=\; \int \cdots \int_S f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \cdots dx_n,$$

and

$$\begin{aligned} f_{X_i}(x_i) \;&=\; \int_{x_1=-\infty}^{\infty} \cdots \int_{x_{i-1}=-\infty}^{\infty} \int_{x_{i+1}=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} \\ &\qquad f_{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \prod_{\substack{j=1 \\ j\neq i}}^{n} dx_j. \end{aligned}$$

We call f_{X_i} the **marginal density** of X_i, and obtain it by integrating out the other $n-1$ variables.

The n random variables $X_1, \ldots, X_n$ are independent if and only if

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \;=\; f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

For discrete random variables, replace the integrals with sums.

(2015 lecture with detailed joint PDF example: http://youtu.be/gQzorseWuVc)

00000

X #H Y #H

indep

	Prob($Y=0$)	Prob($Y=1$)	Prob($Y=2$)	
Prob($X=0$)	1/32	2/32	1/32	1/8
Prob($X=1$)	3/32	6/32	3/32	3/8
Prob($X=2$)	3/32	6/32	3/32	3/8
Prob($X=3$)	1/32	2/32	1/32	1/8
	1/4	2/4	1/4	

Table 9.2: The joint density of (X, Y), where X is the number of heads in the first 3 tosses and Y is the number of heads in the last 2 tosses of 5 independent tosses of fair coins.

000,00

U #H V #H

	Prob($V=0$)	Prob($V=1$)	Prob($V=2$)	
Prob($U=0$)	1/16	1/16	0/16	1/8
Prob($U=1$)	2/16	3/16	2/16	3/8
Prob($U=2$)	1/16	3/16	2/16	3/8
Prob($U=3$)	0/16	1/16	1/16	1/8
	1/4	2/4	1/4	

Table 9.3: The joint density of (U, V), where U is the number of heads in the first 3 tosses and V is the number of heads in the last 2 tosses of 5 tosses of fair coins.

Convolutions and CDF Method

X, Y with pdfs f_X and f_Y, form $Z = X + Y$ with pdf f_Z

$$F_Z(z) = \text{Prob}(Z \le z) = \text{Prob}(X + Y \le z)$$

$$= \int_{x=-\infty}^{\infty} f_X(x)\, \text{Prob}(Y \le z - x)\, dx$$

$$F_Z(z) = \int_{x=-\infty}^{\infty} f_X(x)\, F_Y(z-x)\, dx$$

$'$ means w.r.t z

$$F_Y'(z-x) \cdot (z-x)'$$

$$f_Z(z) = \frac{d}{dz} \int_{x=-\infty}^{\infty} f_X(x) F_Y(z-x)\, dx = \int_{x=-\infty}^{\infty} f_X(x) \left[\frac{d}{dz} F_Y(z-x)\right] dx$$

$$= \int_{x=-\infty}^{\infty} f_X(x)\, f_Y(z-x) \cdot 1\, dx = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx$$

Convolution: $(f * g)(z) = \int_{-\infty}^{\infty} f(x)\, g(z-x)\, dx$

Theorem 9.5.1 (Linearity of Expectation) *Let $X_1, \dots, X_n$ be random variables, let $g_1, \dots, g_n$ be functions such that $\mathbb{E}[|g_i(X_i)|]$ exists and is finite, and let $a_1, \dots, a_n$ be any real numbers. Then*

$$\mathbb{E}[a_1 g_1(X_1) + \cdots + a_n g_n(X_n)] = a_1 \mathbb{E}[g_1(X_1)] + \cdots + a_n \mathbb{E}[g_n(X_n)].$$

Note the random variables are not *assumed to be independent. Also, if $g_i(X_i) = c$ (where c is a fixed number) then $\mathbb{E}[g_i(X_i)] = c$.*

If X has density f_X then $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$

$g(x) = x$ mean

$g(x) = (x-\mu)^2$ variance where $\mu = \mathbb{E}[X]$

write $\mathbb{E}[(X-\mu)^2] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sigma^2$

study standard deviation $\sigma = \sqrt{\sigma^2}$; same units as mean

Proof of Linearity of Expectation

Consider $E[a_1 X_1 + a_2 X_2]$

$$= \int_{x_1=-\infty}^{\infty} \int_{x_2=-\infty}^{\infty} (a_1 x_1 + a_2 x_2) f_{X_1, X_2}(x_1, x_2)\, dx_1\, dx_2$$

$$= a_1 \int_{x_1=-\infty}^{\infty} x_1 \left[\int_{x_2=-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2)\, dx_2 \right] dx_1 + a_2 \int_{x_2=-\infty}^{\infty} x_2 \left[\int_{x_1=-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2)\, dx_1 \right] dx_2$$

$$= a_1 \int_{x_1=-\infty}^{\infty} x_1 f_{X_1}(x_1)\, dx_1 + a_2 \int_{x_2=-\infty}^{\infty} x_2 f_{X_2}(x_2)\, dx$$

$$= a_1 E[X_1] + a_2 E[X_2] \quad \blacksquare$$

General: Similar calculation or $(X_1 + X_2) + X_3$ (grouping)

Application of Linearity of Expectation: Bernoulli/Binomial

$X_1, \dots, X_n$ tosses of a coin with prob p of heads

$X_k \sim \mathrm{Bern}(p)$, let $X = X_1 + \dots + X_n \sim \mathrm{Bin}(n,p)$

$$E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n] = nE[X_1]$$

$$E[X_1] = 1 \cdot p + 0 \cdot (1-p) = p$$

Thus $E[X] = E[X_1 + \dots + X_n] = np$

$$E[X] = \sum_{k=0}^{n} k \cdot \mathrm{Prob}(X = k) = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

Fermat Primes

. 3, 5, 17,

2^n+1

$\uparrow$ n

$F_n = 2^{2^n}+1$

primes $\leq X$ is approx $\frac{X}{\log X}$

prob a number is prime is about $1/\log X$ if number $\approx X$

$$X_n = \begin{cases} 1 & \text{with prob } 1/\log F_n \\ 0 & \text{with prob } 1 - 1/\log F_n \end{cases}$$

$$\sum E[X_n] = \sum \frac{1}{\log(2^{2^n}+1)} \approx \sum \left(\frac{1}{\log(2^{2^n})}\right)$$

$$= \sum \frac{1}{2^n \log 2} < \infty$$

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 10: 3/11/25: Buffon's Needle, Simpson's Paradox, Crackerjack Box
Video: https://youtu.be/iXmE8jvOYAQ

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 10: March 11, 2025:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- Buffon’s Needle
- Simpson's paradox
- Crackerjack Problem

General items.

- Power of Linearity
- Avoiding brute force computations

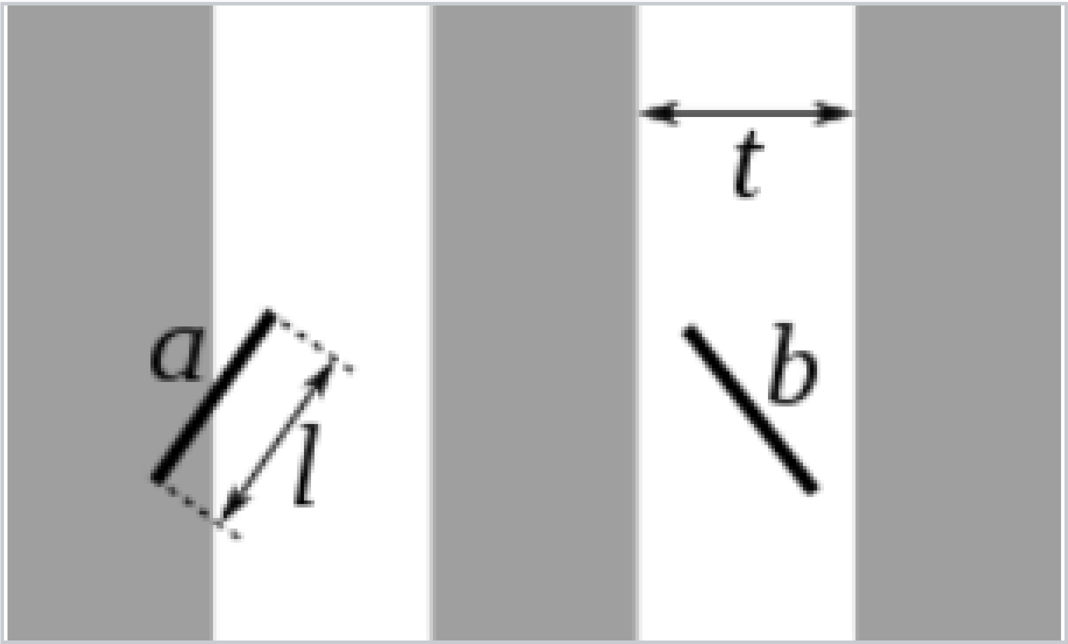
Buffon's needle problem

https://en.wikipedia.org/wiki/Buffon%27s_needle_problem

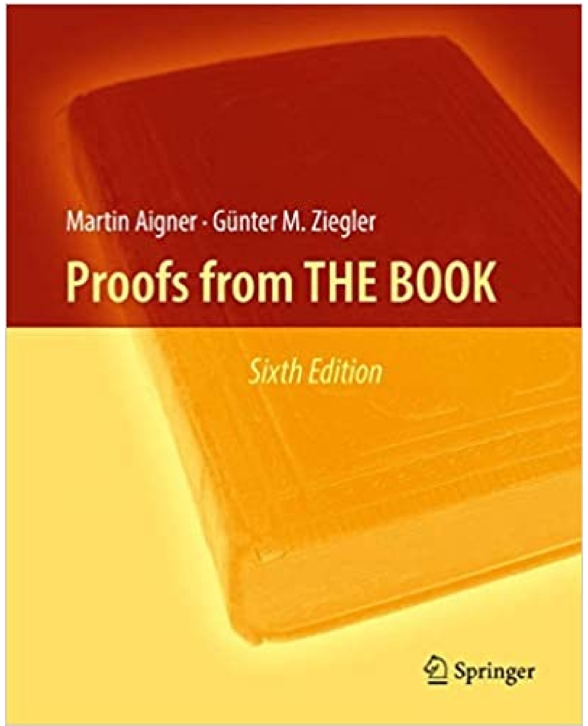
From Wikipedia, the free encyclopedia

In mathematics, **Buffon's needle problem** is a question first posed in the 18th century by Georges-Louis Leclerc, Comte de Buffon:[1]

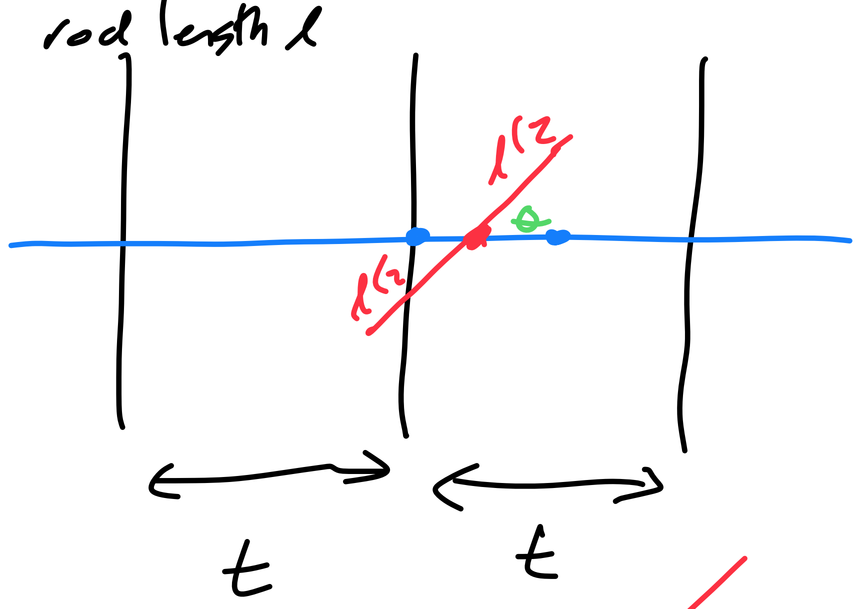
> Suppose we have a floor made of parallel strips of wood, each the same width, and we drop a needle onto the floor. What is the probability that the needle will lie across a line between two strips?



The *a* needle lies across a line, while the *b* needle does not.



https://pub.math.leidenuniv.nl/~finkelnbergh/seminarium/stelling_van_Buffon.pdf



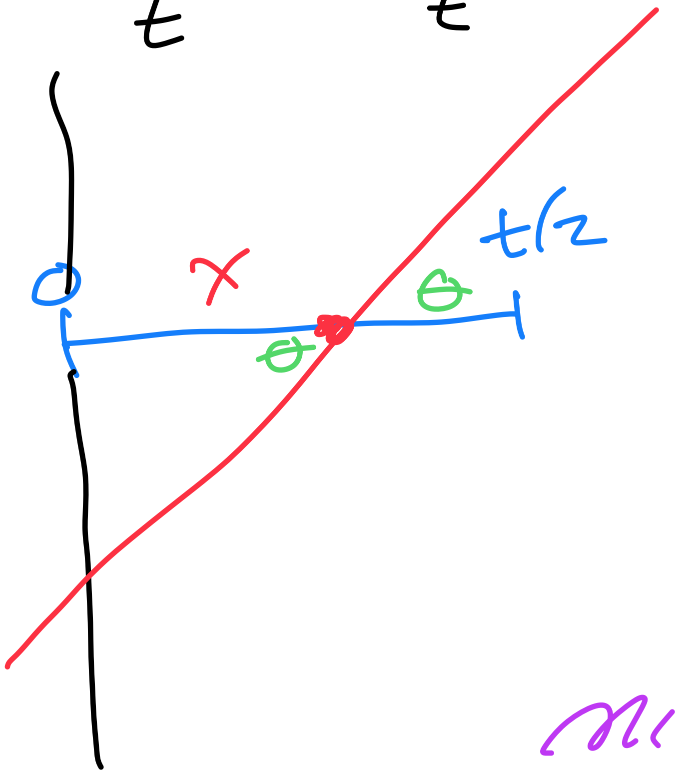
wlog, x-coord of rod center is b/w 0 and $t/2$: $X \sim \mathrm{Unif}(0, t/2)$

↳ density $\frac{1}{t/2} = \frac{2}{t}$

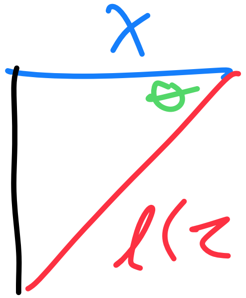
Wlog, angle is b/w 0 and $\pi/2$

So $\Theta \sim \mathrm{Unif}(0, \pi/2)$

↳ density $\frac{1}{\pi/2} = \frac{2}{\pi}$



Keep hitting till have a right △ of hypotenuse length $\ell/2$



$\cos\theta = \frac{x}{\ell/2}$ or $\frac{2x}{\ell}$

So $\theta = \arccos(2x/\ell)$

missed $x \leq \ell/2$

$$\theta = \arccos(2x/\ell)$$

should be $\ell/2$

$$\int_{x=0}^{t/2} \int_{\theta=0}^{\arccos(2x/\ell)} \frac{4}{\pi t}\, d\theta\, dx = \frac{4}{\pi t} \int_{x=0}^{t/2} \arccos\left(\frac{2x}{\ell}\right) dx$$

$u = \frac{2x}{\ell}$ so $du = \frac{2}{\ell} dx$ or $dx = \frac{\ell}{2} du$: $\quad x: 0 \to t/2$ ($\ell/2$), $\quad u: 0 \to t/\ell$ (1)

$$= \frac{4}{\pi t} \frac{\ell}{2} \int_{u=0}^{t/\ell \; (1)} \arccos(u)\, du = \frac{2}{\pi} \frac{\ell}{t} \int_{u=0}^{t/\ell \; (1)} \arccos(u)\, du$$

(only depends on ratio ℓ/t)

Try $u = \cos w$ Then $du = -\sin w\, dw$ $\quad u: 0 \to t/\ell$ (1) $\Rightarrow w: \pi/2 \to \cancel{\arccos(t/\ell)}$ 0

$$= \frac{2}{\pi} \frac{\ell}{t} \int_{w=\cancel{\arccos(t/\ell)}\; 0}^{\pi/2} w \sin w\, dw$$

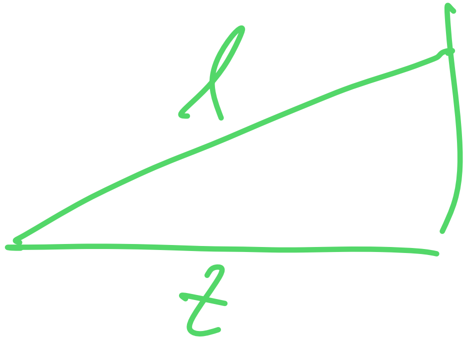
$$\frac{2}{\pi}\frac{l}{t}\int_{w=\cancel{\arccos(t/l)}\;0}^{\pi/2} w\sin w\,dw = AB\Big|_{\sim}^{\sim} - \int_{\sim}^{\sim} B\,dA$$

$$A = w \qquad dB = \sin w\,dw$$

$$dA = dw \qquad B = -\cos w$$

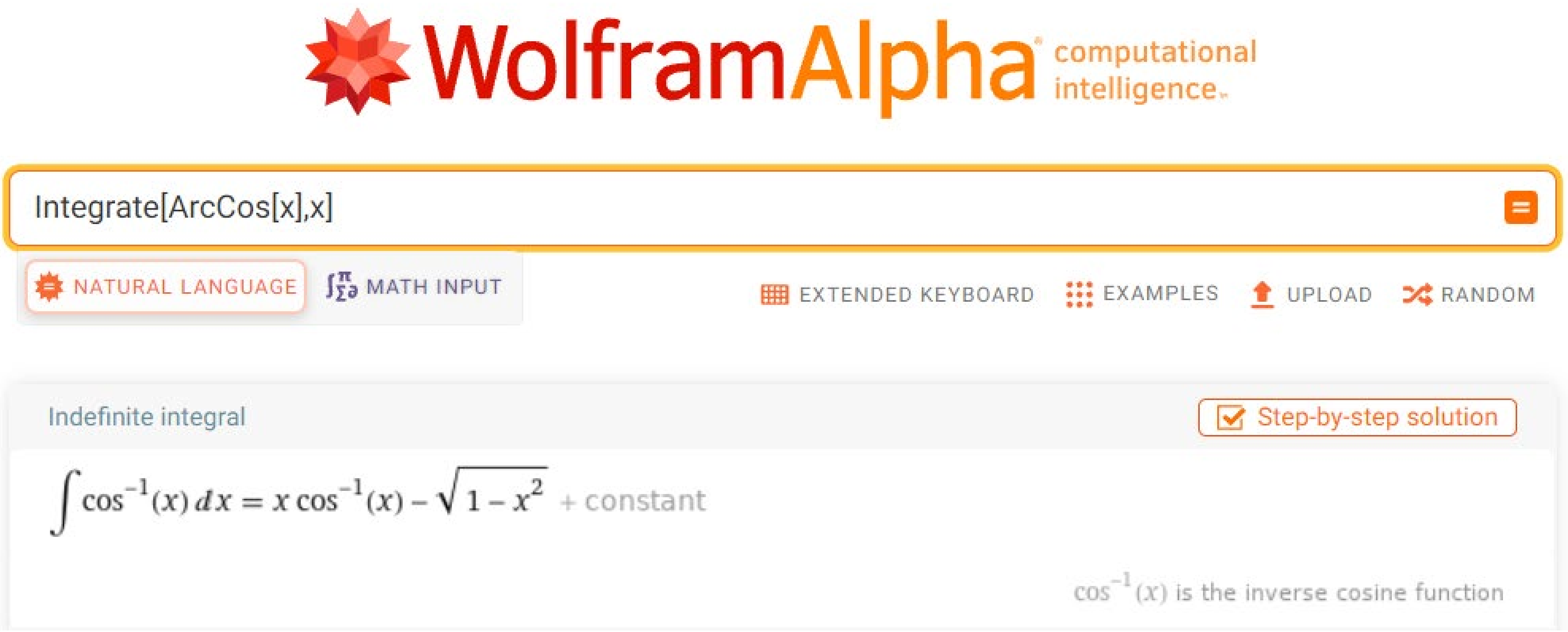
$$= \frac{2}{\pi}\frac{l}{t}\left[w\cos w\Big|_{\pi/2}^{\cancel{\arccos(t/l)}\;0} + \int_{\cancel{\arccos(t/l)}\;0}^{\pi/2} \cos w\,dw \right]$$

$$= \frac{2}{\pi}\frac{l}{t}\left[\cancel{\arccos\left(\frac{t}{l}\right)\frac{t}{l}}\;0 + \sin w\Big|_{\cancel{\arccos(t/l)}\;0}^{\pi/2} \right]$$



$$= \frac{2}{\pi}\frac{l}{t}\left[\cancel{\frac{t}{l}\arccos\left(\frac{t}{l}\right)} + 1 + 0 \right]$$

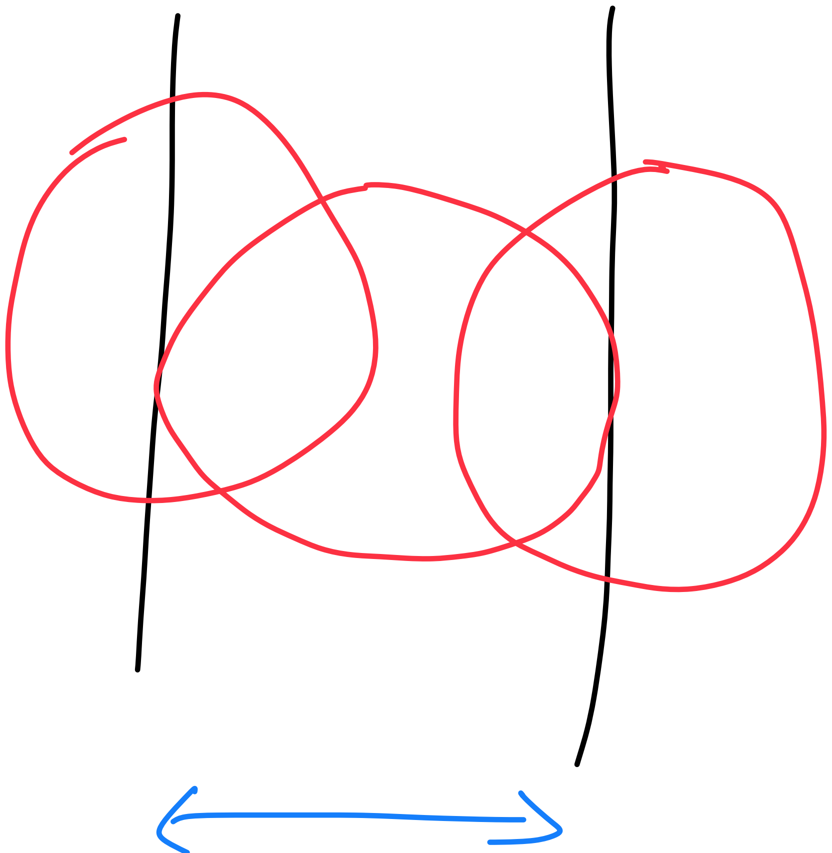
$$= \frac{2}{\pi}\frac{l}{t}$$



But maybe it is better to switch the order of integration.... What would you get?

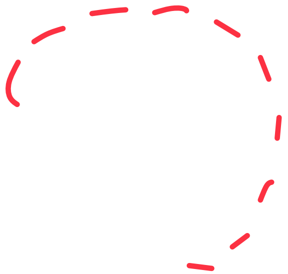
Here is the ‘proof from the book’ link:
https://pub.math.leidenuniv.nl/~finkelnbergh/seminarium/stelling_van_Buffon.pdf

Proof from the book



t

perimeter: $2\pi\, t/2$



https://en.wikipedia.org/wiki/Simpson%27s_paradox

Simpson's Paradox

Derek Jeter: .250 (1995) and .314 (1996)

David Justice: .253 (1995) and .321 (1996)

BUT Jeter's two year average was .310, exceeding Justice's .270.

Is this surprising?

If yes, how can this be true?

Batter \ Year	1995		1996		Combined	
Derek Jeter	12/48	.250	183/582	.314	195/630	**.310**
David Justice	104/411	**.253**	45/140	**.321**	149/551	.270

Kidney stone treatment [edit]

Another example comes from a real-life medical study[17] comparing the success rates of two treatments for kidney stones.[18] The table below shows the success rates (the term *success rate* here actually means the success proportion) and numbers of treatments for treatments involving both small and large kidney stones, where Treatment A includes open surgical procedures and Treatment B includes closed surgical procedures. The numbers in parentheses indicate the number of success cases over the total size of the group.

Stone size \ Treatment	Treatment A	Treatment B
Small stones	*Group 1* **93% (81/87)**	*Group 2* 87% (234/270)
Large stones	*Group 3* **73% (192/263)**	*Group 4* 69% (55/80)
Both	78% (273/350)	**83% (289/350)**

The paradoxical conclusion is that treatment A is more effective when used on small stones, and also when used on large stones, yet treatment B appears to be more effective when considering both sizes at the same time. In this example, the "lurking" variable (or confounding variable) causing the paradox is the size of the stones, which was not previously known to researchers to be important until its effects were included.[*citation needed*]

Cracker Jack Problem:

Total of T toys you can get.
Each box equally likely to have any of the T toys.
Each box prize independent of the others.
How many boxes do you expect to open before seeing all?

- Should ↑ with T
- Get first, get second, ... $T + \dots + T$
  or T^2 (kind of upper bound)
- lower bound T

Expect wait b/w T and T^2

How long to wait?

$\underbrace{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx}}_{\text{at least one of each of the other } T-1 \text{ prizes}}\ \underbrace{\hphantom{xxxxxxxxx}}_{T\text{th prize is obtained at } n\text{th box}}$

wlog have prizes $1, 2, \ldots, T-1$

$X_1 = \#\text{prize } 1, \ \ldots, \ X_{T-1} = \#\text{prize } T-1$

$X_1 + \cdots + X_{T-1} = n-1$ and each $X_i \geq 1$ so $X_i = Y_i + 1$

$Y_1 + \cdots + Y_{T-1} = n-1-(T-1)$ $\quad Y_i \geq 0$

orderings, multinomials, ... HELP!

Linearity of Expectation

X_1 = # boxes till see first prize haven't seen: 1

X_2 = # boxes till see the next new prize

$\vdots$

X_k = # boxes till see next new prize, have seen $k-1$ prizes so far

$\vdots$

X_T

Define: $X = X_1 + X_2 + \cdots + X_T$

$$E[X] = 1 + E[X_2] + \cdots + E[X_T]$$

X_k = # boxes to open ~~to see~~ next new prize, have already seen exactly $k-1$ of T.

Prob next box is new is $P_k = \frac{T-(k-1)}{T}$

Prob takes one box is P_k

Prob takes two boxes is $(1-P_k)P_k$

" " three boxes is $(1-P_k)^2 P_k$

Basically waiting is a geom random variable

$$E[X_k] = \sum_{n=1}^{\infty} \underset{\text{Value}}{n} \cdot \underbrace{(1-P_k)^{n-1} P_k}_{\text{Prob takes } n \text{ boxes}} = P_k \sum_{n=0}^{\infty} n(1-P_k)^{n-1}$$

$$= \frac{P_k}{1-P_k} \sum_{n=1}^{\infty} n((1-P_k))^n$$

Can extend to $n=0$

Must find $\sum_{n=0}^{\infty} n r^n = 0 + r + 2r^2 + 3r^3 + \cdots$

Know $\sum_{n=0}^{\infty} r^n = 1 + r + r^2 + r^3 + \cdots = \frac{1}{1-r}$

$r\sum_{n=0}^{\infty} r^n = r + r^2 + r^3 + \cdots = r\frac{1}{1-r}$

$r^2\sum_{n=0}^{\infty} r^n = r^2 + r^3 + \cdots = r^2\frac{1}{1-r}$

$\vdots$

$1 + 2r + 3r^2 + 4r^3 + \cdots = \frac{1}{1-r}\frac{1}{1-r}$

Multiply by r

$r + 2r^2 + 3r^3 + \cdots = \frac{r}{(1-r)^2}$

Lemma: $\displaystyle\sum_{n=0}^{\infty} n r^n = \frac{r}{(1-r)^2}$

$$E[X_k] = \frac{p_k}{1-p_k} \sum_{n=0}^{\infty} n\left(1-p_k\right)^n = \frac{p_k}{1-p_k} \cdot \frac{1-p_k}{p_k^2} = \frac{\cancel{1-p_k}\ 1}{p_k}$$

but $p_k = \dfrac{T-(k-1)}{T}$ $\quad\cancel{\text{so } \dfrac{1-p_k}{p_k} = \dfrac{k-1}{T} \cdot \dfrac{T}{T-(k-1)}}$

$$p_k = 1 - \frac{k-1}{T}$$

$$E[X_k] = \frac{1}{p_k} = \frac{T}{T-(k-1)}$$

$$\text{So } E[X] = \frac{T}{T-0} + \frac{T}{T-1} + \frac{T}{T-3} + \frac{T}{T-4} + \dots + \frac{T}{1}$$

$$= T\overbrace{\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{T}\right)}^{\text{Harmonic Sum}} \simeq T \log T$$

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 11: 3/13/25: Review, Variances/Covariances, Portfolios, Coding, Differentiating Identities:
Video: https://youtu.be/mGQb4ZTTu0c
https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 11: March 11, 2025:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- Review (CDF, Convolution, Special Distributions)
- Variance / Covariance
- Expectation Results
- Portfolio Application
- Coding
- Differentiating Identities

General items.

- Power of Independence

The Method of the Cumulative Distribution Function. Let X be a random variable with density f_X whose density is non-zero on some interval I, and let $Y = g(X)$ where $g : I \to \mathbb{R}$ is a differentiable function with inverse h. Assume the derivative of g is either always positive or always negative in I, except at finitely many points where it may vanish. To find the density f_Y:

1. Identify the interval I where the random variable X is defined.

2. Prove the function g has a derivative that is always positive or always negative (except, of course, at potentially finitely many points).

3. Determine the inverse function $h(y)$, where $g(h(y)) = y$ and $h(g(x)) = x$.

4. Determine $h'(y)$, either by directly differentiating h or using the relation $h'(y) = 1/g'(h(y))$.

5. The density of Y is $f_Y(y) = f_X(h(y))|h'(y)|$.

Definition 10.1.1 *The* ***convolution*** *of* independent *continuous random variables X and Y on $\mathbb{R}$ with densities f_X and f_Y is denoted $f_X * f_Y$, and is given by*

$$(f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(t) f_Y(z-t) dt.$$

If X and Y are discrete, we have

$$(f_X * f_Y)(z) = \sum_n f_X(x_n) f_Y(z - x_n);$$

note of course that $f_Y(z - x_n)$ is zero unless $z - x_n$ is one of the values where Y has positive probability (i.e., one of the special points y_m).

The convolution of two random variables has many wonderful properties, including the following theorem.

Theorem 10.1.2 *Let X and Y be continuous or discrete* independent *random variables on $\mathbb{R}$ with densities f_X and f_Y. If $Z = X + Y$, then*

$$f_Z(z) = (f_X * f_Y)(z).$$

Further, convolution is commutative: $f_X * f_Y = f_Y * f_X$.

Mean, variance. Let X be either a continuous or a discrete random variable with density f_X.

1. The **mean** (or **average value** or **expected value**) of X is the first moment. We denote it by $\mathbb{E}[X]$ or μ_X (if the random variable is clear, we often suppress the subscript X and write μ). Explicitly,

$$\mu = \begin{cases} \int_{-\infty}^{\infty} x \cdot f_X(x)dx & \text{if } X \text{ is continuous} \\ \sum_n x_n \cdot f_X(x_n) & \text{if } X \text{ is discrete.} \end{cases}$$

2. The **variance** of X, denoted σ_X^2 or $\mathrm{Var}(X)$, is the second centered moment, or equivalently the expected value of $g(X) = (X - \mu_X)^2$; again, we often suppress the subscript X if the random variable is clear, and write σ^2. Writing it out in full,

$$\sigma_X^2 = \begin{cases} \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)dx & \text{if } X \text{ is continuous} \\ \sum_n (x_n - \mu_X)^2 f_X(x_n) & \text{if } X \text{ is discrete.} \end{cases}$$

   As $\mu_X = \mathbb{E}[X]$, after some algebra (see Lemma 9.5.3) one finds

$$\sigma^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

   This relates the variance to the first two moments of X, and is useful for many calculations. The **standard deviation** is the square-root of the variance, or $\sigma_X = \sqrt{\sigma_X^2}$.

3. **Technical caveat:** in order for the mean to exist, we want $\int_{-\infty}^{\infty} |x| f_X(x)dx$ (in the continuous case) or $\sum_n |x_n| f_X(x_n)$ (in the discrete case) to be finite.

USUALLY:

$\mathbb{E}[X^2] \neq \mathbb{E}[X]^2$

Covariance. Let X and Y be random variables. The covariance of X and Y, denoted by σ_{XY} or $\mathrm{Cov}(X,Y)$, is

$$\sigma_{XY} := \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right].$$

Note $\mathrm{Cov}(X,X)$ equals the variance of X. Also, if $X_1,\ldots,X_n$ are random variables and $X = X_1+\cdots+X_n$, then

$$\mathrm{Var}(X) = \sum_{i=1}^{n}\mathrm{Var}(X_i) + 2\sum_{1\le i<j\le n}\mathrm{Cov}(X_i,X_j).$$

$$\sum_{i\neq j} \underbrace{\mathrm{Cov}(X_i, X_j)}_{\sigma_{X_i X_j}}$$

The Bernoulli Distribution: X has a **Bernoulli distribution** with parameter $p \in [0,1]$ if $\mathrm{Prob}(X=1) = p$ and $\mathrm{Prob}(X=0) = 1-p$. We view the outcome 1 as a **success**, and 0 as a **failure**. We write $X \sim \mathrm{Bern}(p)$. We also call X a **binary indicator random variable**.

$$E[X] = \sum_n n \,\mathrm{Prob}(X=n)$$

$$= 0 \cdot (1-p) + 1 \cdot p = p$$

$$E[X^2] = \sum_n n^2 \,\mathrm{Prob}(X=n)$$

$$= 0^2 \cdot (1-p) + 1^2 \cdot p = p$$

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = p - p^2 = p(1-p) \in (0,1)$$

largest at $p = 1/2$: $f(p) = p(1-p) = p - p^2$ so $f'(p) = 1 - 2p$

Theorem: Simplify, Simplify

$f'(p) = 0 \Rightarrow p = 1/2$

$f''(p) = -2$ so it's a max

Lemma 9.5.2 *Let X be a random variable with mean μ_X and variance σ_X^2. If a and b are any fixed constants, then for the random variable $Y = aX + b$ we have*

$$\mu_Y = a\mu_X + b \quad \text{and} \quad \sigma_Y^2 = a^2\sigma_X^2.$$

Lemma 9.5.3 *Let X be a random variable. Then*

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Theorem 9.6.1 *If X and Y are independent random variables, then*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

A particularly important case is

$$\mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[X - \mu_X]\mathbb{E}[Y - \mu_Y] = 0.$$

X, Y indep

$\rightarrow X - \mu_X, Y$ indep

$\rightarrow X - \mu_X, Y - \mu_Y$ indep

Proof: $Z = XY$ $\quad E[Z] = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} xy\, f_{X,Y}(x,y)\, dx\, dy$

As indep: $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$

Thus $E[XY] = \underbrace{\int_{y=-\infty}^{\infty} y\, f_Y(y)\, dy}_{E[Y]} \underbrace{\int_{x=-\infty}^{\infty} x\, f_X(x)\, dx}_{E[X]}$ ∎

The Binomial Distribution: Let n be a positive integer and let $p \in [0,1]$. Then X has the **binomial distribution** with parameters n and p if

$$\operatorname{Prob}(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k \in \{0, 1, \ldots, n\} \\ 0 & \text{otherwise.} \end{cases}$$

We write $X \sim \operatorname{Bin}(n,p)$. The mean of X is np and the variance is $np(1-p)$.

$X = X_1 + \cdots + X_n$ each $X_i \sim \operatorname{Bern}(p)$ and indep

$E[X] = E[X_1] + \cdots + E[X_n] = np$

$\operatorname{Var}(X) = \operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n) = np(1-p)$

Var of sum is sum of variances as independent.

$\sigma^2 = \sigma_1^2 + \sigma_2^2 \implies \sigma = \sigma_1 + \sigma_2$? NO!

Theorem 9.6.2 (Means and Variances of Sums of Random Variables) *Let $X_1, \ldots, X_n$ be random variables with means $\mu_{X_1}, \ldots, \mu_{X_n}$ and variances $\sigma^2_{X_1}, \ldots, \sigma^2_{X_n}$. If $X = X_1 + \cdots + X_n$, then*

$$\mu_X = \mu_{X_1} + \cdots + \mu_{X_n}.$$

If the random variables are independent, then we also have

$$\sigma^2_X = \sigma^2_{X_1} + \cdots + \sigma^2_{X_n} \quad \text{or} \quad \mathrm{Var}(X) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

In the special case when the random variables are independent and identically distributed (so all the means equal μ and all the variances equal σ^2), then

$$\mu_X = n\mu \quad \text{and} \quad \sigma^2_X = n\sigma^2.$$

Imagine know for two terms.

$$\mathrm{Var}((X_1 + X_2) + X_3)$$
$$= \mathrm{Var}(X_1 + X_2) + \mathrm{Var}(X_3)$$
$$= (\mathrm{Var}(X_1) + \mathrm{Var}(X_2)) + \mathrm{Var}(X_3)$$
$$= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3)$$

Proof: $\mathrm{Var}(X+Y) = E[((X+Y) - (\mu_X + \mu_Y))^2]$

$$= E[((X - \mu_X) + (Y - \mu_Y))^2]$$
$$= E[(X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2]$$

independent

$$= E[(X-\mu_X)^2] + 2E[X-\mu_X]E(Y-\mu_Y] + E[(Y-\mu_Y)^2]$$
$$= \mathrm{Var}(X) + 0 + \mathrm{Var}(Y) \quad \blacksquare$$

Lemma: $E[X - \mu_X] = 0$

Proof: Equals $E[X] - E[\mu_X]$

$= \mu_X - \mu_X$

$= 0$ $\blacksquare$

Covariance. Let X and Y be random variables. The covariance of X and Y, denoted by σ_{XY} or $\mathrm{Cov}(X,Y)$, is

$$\sigma_{XY} := \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right].$$

Note $\mathrm{Cov}(X,X)$ equals the variance of X. Also, if $X_1,\ldots,X_n$ are random variables and $X = X_1 + \cdots + X_n$, then

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1\le i<j\le n} \mathrm{Cov}(X_i, X_j).$$

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\
&= \mathbb{E}[XY - \mu_Y X - \mu_X Y + \mu_Y \mu_X] \\
&= \mathbb{E}[XY] - \mu_X \mathbb{E}[X] - \mu_X \mathbb{E}[Y] + \mathbb{E}[\mu_X \mu_Y] \\
&= \mathbb{E}[XY] - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mu_X\mu_Y .
\end{aligned}$$

Application: Amazingly, Theorem 9.6.2 plays a big role in designing optimal **investment portfolios**! Here's a brief sketch of how it's used. Imagine we have two stocks with variable returns. Let X_1 denote our return from the first, and X_2 from the second. For simplicity, let's assume they both cost \$1 per share, both have an average return of \$3, and both have a variance of \$2. Our goal is to build a portfolio that makes as much money as possible, with as little risk as possible. *If* the two stocks are independent, we can minimize risk by investing in each!

$0 \le w \le 1$ invest w % in X_1 and $(1-w)$ % in X_2

$$X = wX_1 + (1-w)X_2$$

$$E[X] = E[wX_1 + (1-w)X_2] = E[wX_1] + E[(1-w)X_2]$$

$$= wE[X_1] + (1-w)E[X_2] = w \cdot \$3 + (1-w)\$3 = \$3$$

$$Var(X) = Var(\underbrace{wX_1} + \underbrace{(1-w)X_2}) = Var(wX_1) + Var((1-w)X_2)$$

$$= w^2 Var(X_1) + (1-w)^2 Var(X_2) = [w^2 + (1-w)^2]\$2$$

$Var(X) = \left[w^2 + (1-w)^2\right] 2$: find w to minimize (TANSTAAFL)

$f(w) = w^2 + (1-w)^2$

find critical points and endpoints

$f'(w) = 2w + 2(1-w)(-1)$ so $f'(w) = 0 \Rightarrow w = 1/2$

$f(1/2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$f(0) = f(1) = 1$

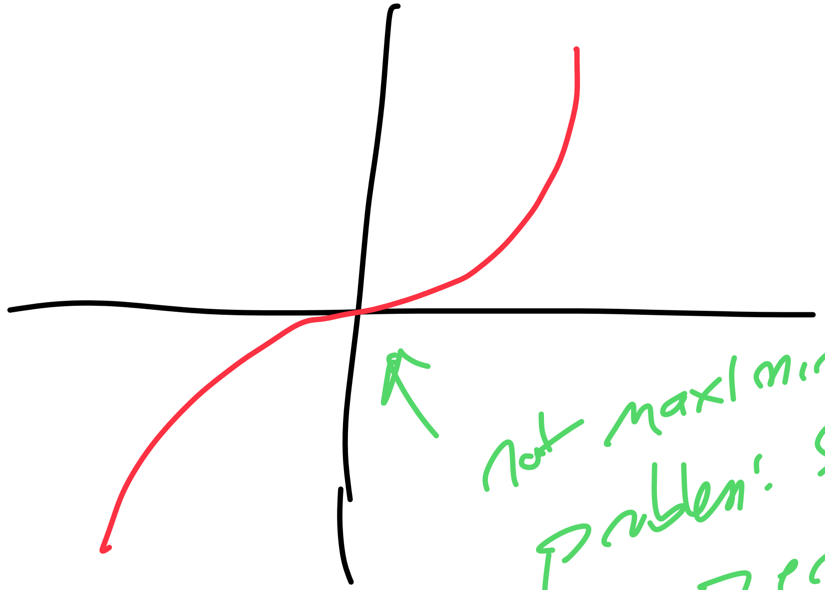
Minimum at $w = 1/2$

If indep, can $\downarrow$ variance by a combination

$f(x) = x^3$

$f'(x) = 3x^2$ so critical points just $x=0$

$-1 \leq x \leq 1$



not max/min
problem: $f''(x)$ is zero at $x=0$, need to look at $f'''(0)$

Method of Differentiating Identities: Let $\alpha, \beta, \gamma, \ldots, \omega$ be some parameters. Assume

$$\sum_{n=n_{\min}}^{n_{\max}} f(n;\alpha,\beta,\ldots,\omega) \;=\; g(\alpha,\beta,\ldots,\omega),$$

where f and g are differentiable functions with respect to α. Then

$$\sum_{n=n_{\min}}^{n_{\max}} \frac{\partial f(n;\alpha,\beta,\ldots,\omega)}{\partial\alpha} \;=\; \frac{\partial g(\alpha,\beta,\ldots,\omega)}{\partial\alpha},$$

provided that f has sufficient decay to justify the interchange of summation and differentiation.

Cracker Jack Toy Problem

T toys, each box indep of others, each toy = likely in any box.
How many boxes to open till see each toy?

Set structure: open box if not in set add to set,
keep doing till set of size T.

Trick: boxes 1, 10, 100, ... keep track of sum
↳ can do with mod games! to see if have 100,
mod by 1000 and see if at least 100.
Array: toy[k] is 1 if get toy k, and see if
all positive/product is positive.

Let t_{max} be the largest toy number seen.
WLOG first toy seen is 1, so $t_{max} = 1$ now

↳ open boxes till see something > 1.
call whatever you see prize 2, set $t_{max} = 2$.

↳ open boxes till see something > 2.
call that toy 3, now set $t_{max} = 3$

Everytime open box ↑ counter by 1, Do until $t_{max} = T$.

totalcount = totalcount + counter; vartotalcount = vartotalcount + counter^2

↳ divide by number of iterations to get average.

n = 1 to numdo

{

Find # boxes to get all toys

↑ counter by that

↑ varcounter by square of that

}

Calculate mean by $\div$ by numdo,

Similarly get var from $E[X^2] - E[X]^2$

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 12: 3/18/25: Coding, Differentiating Identities
Video: https://youtu.be/E6o3AUEbGxM

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 11: March 11, 2025:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- Coding
- Differentiating Identities
- Computing Means and Variances

General items.

- Milking equalities

```
crackerjackprizeproblem[numprizes_, numdo_] := Module[{},
  xsum = 0; xxsum = 0; (* to compute mean and variance *)
  For[n = 1, n ≤ numdo, n++,
   {
    numboxes = 1; (* start off with one box, as first box is new! *)
    count = 1; (* while loop opens boxes till get a new toy, relabel that to numbox+1 *)
    While[numboxes < numprizes,
     {
      count = count + 1; (* opened another box *)
      b = RandomInteger[{1, numprizes}]; (* value of toy *)
      If[b > numboxes, (* thus new toy found *) numboxes = numboxes + 1; ]; (* end of if b loop *)
     }]; (* end of while loop *)
    xsum = xsum + count; xxsum = xxsum + count^2;
   }]; (* end of for loop *)
  obsmean = xsum / numdo; predvar = Sum[ (1 - (k / numprizes)) (numprizes / k)^2, {k, 1, numprizes - 1}];
  Print["Mean is ", 1.0 obsmean, "."];
  Print["Predicted mean is ", 1.0 numprizes HarmonicNumber[numprizes], "."];
   (*Print["Variance is ",  (1.0 xxsum/numdo) - obsmean^2, "."];*) ×
   Print["Standard Deviation is ", Sqrt[(1.0 xxsum / numdo) - obsmean^2], "."];
  Print["Predicted Standard Deviation is ", Sqrt[1.0 predvar], "."];
 ]
```

```
Timing[crackerjackprizeproblem[10, 100000]]
```

Mean is 29.3183.

Predicted mean is 29.2897.

Standard Deviation is 11.2561.

Predicted Standard Deviation is 11.211.

{5.53125, Null}

```
Timing[crackerjackprizeproblem[10, 1000000]]
```

Mean is 29.3037.

Predicted mean is 29.2897.

Standard Deviation is 11.2204.

Predicted Standard Deviation is 11.211.

{55.8281, Null}

```
Timing[crackerjackprizeproblem[1000, 1000]]
```

Mean is 7416.54.

Predicted mean is 7485.47.

Standard Deviation is 1280.27.

Predicted Standard Deviation is 1279.24.

{12.2188, Null}

```
Timing[crackerjackprizeproblem[1000, 10000]]
```

Mean is 7495.48.

Predicted mean is 7485.47.

Standard Deviation is 1287.97.

Predicted Standard Deviation is 1279.24.

{152.281, Null}

```
arraycrackerjackprizeproblem[numprizes_, numdo_] := Module[{},
  xsum = 0; xxsum = 0; (* to compute mean and variance *)
  For[i = 0, i ≤ 2 numprizes^2 + 1, i++, arraynumboxes[i] = 0]; (* array to store *)
  For[n = 1, n ≤ numdo, n++,
   {
    numboxes = 1; (* start off with one box, as first box is new! *)
    count = 1; (* while loop opens boxes till get a new toy, relabel that to numbox+1 *)
    While[numboxes < numprizes,
     {
      count = count + 1; (* opened another box *)
      b = RandomInteger[{1, numprizes}]; (* value of toy *)
      If[b > numboxes, (* thus new toy found *) numboxes = numboxes + 1; ]; (* end of if b loop *)
     }]; (* end of while loop *)
    If[count ≤ 2 * numprizes^2, arraynumboxes[count] = arraynumboxes[count] + 1,
     { Print["Exceeded upper bound!!!"]; arraynumboxes[0] = arraynumboxes[0] + 1;}]; (* stores in array; if too big put in 0 *)
    xsum = xsum + count; xxsum = xxsum + count^2;
   }]; (* end of for loop *)
  obsmean = xsum / numdo; predvar = Sum[ (1 - (k / numprizes)) (numprizes / k)^2, {k, 1, numprizes - 1}];
  Print["Mean is ", 1.0 obsmean, "."];
  Print["Predicted mean is ", 1.0 numprizes HarmonicNumber[numprizes], "."];
  (*Print["Variance is ",  (1.0 xxsum/numdo) - obsmean^2, "."];*)
  Print["Standard Deviation is ", Sqrt[(1.0 xxsum / numdo) - obsmean^2], "."];
  Print["Predicted Standard Deviation is ", Sqrt[1.0 predvar], "."];
  plotlist = {};
  For[i = numprizes, i ≤ 4 * numprizes * Log[numprizes], i++, plotlist = AppendTo[plotlist, {i, 1.0 arraynumboxes[i] / numdo}]];
  Print[ListPlot[plotlist]];
 ]
```

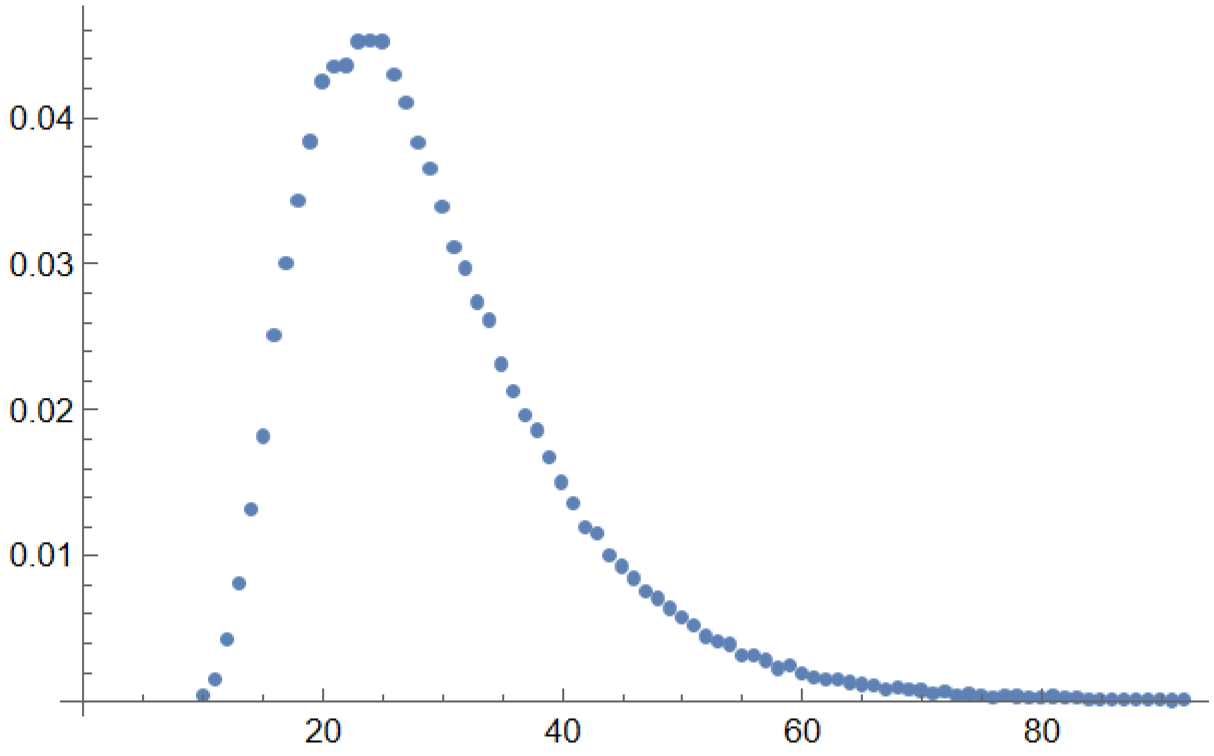
```
Timing[arraycrackerjackprizeproblem[10, 100 000]]
```

Mean is 29.2638.

Predicted mean is 29.2897.

Standard Deviation is 11.1746.

Predicted Standard Deviation is 11.211.



{5.84375, Null}

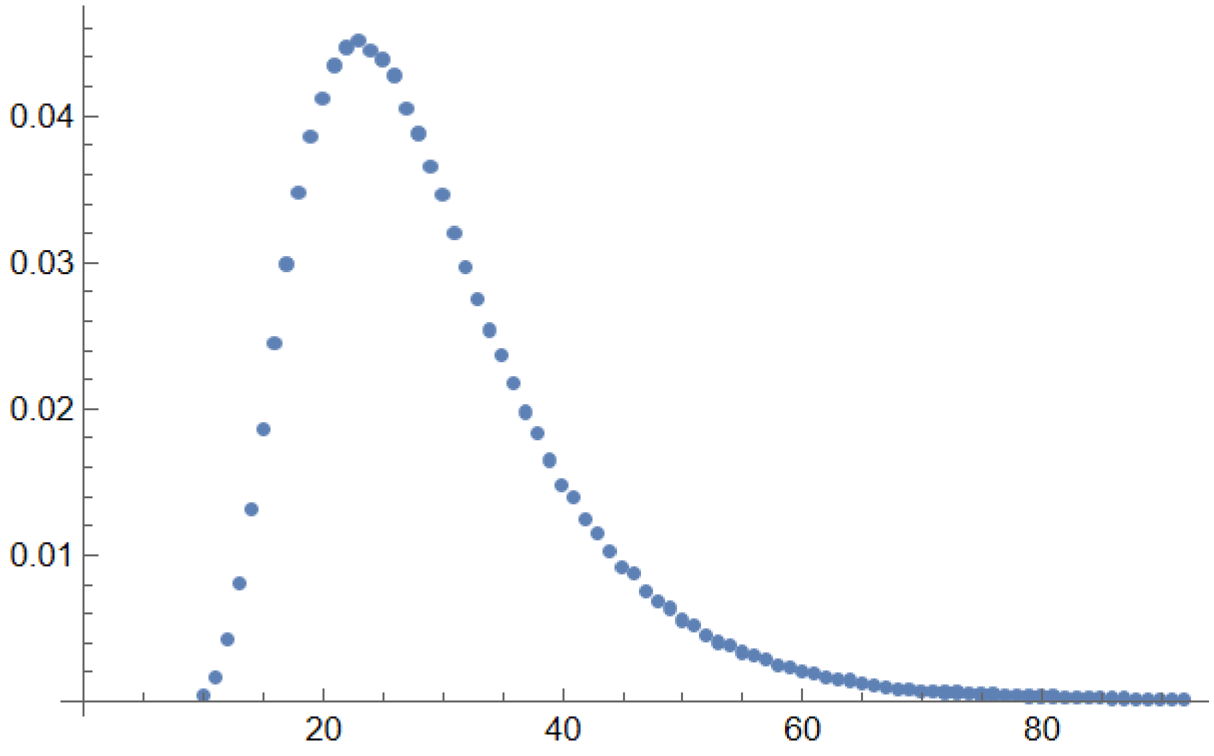
```
Timing[arraycrackerjackprizeproblem[10, 1 000 000]]
```

Mean is 29.3099.

Predicted mean is 29.2897.

Standard Deviation is 11.2184.

Predicted Standard Deviation is 11.211.



{58.7969, Null}

```
Timing[crackerjackprizeproblem[1000, 10000]]

Mean is 7464.73.

Predicted mean is 7485.47.

Standard Deviation is 1279.71.

Predicted Standard Deviation is 1279.24.

{131.016, Null}

Timing[arraycrackerjackprizeproblem[1000, 10000]]

Mean is 7471.15.

Predicted mean is 7485.47.

Standard Deviation is 1253.26.

Predicted Standard Deviation is 1279.24.

{136.641, Null}
```

Method of Differentiating Identities: Let $\alpha, \beta, \gamma, \ldots, \omega$ be some parameters. Assume

$$\sum_{n=n_{\min}}^{n_{\max}} f(n;\alpha,\beta,\ldots,\omega) = g(\alpha,\beta,\ldots,\omega),$$

where f and g are differentiable functions with respect to α. Then

$$\sum_{n=n_{\min}}^{n_{\max}} \frac{\partial f(n;\alpha,\beta,\ldots,\omega)}{\partial\alpha} = \frac{\partial g(\alpha,\beta,\ldots,\omega)}{\partial\alpha},$$

provided that f has sufficient decay to justify the interchange of summation and differentiation.

$$\int_{x=3}^{13} f(x,y)\,dx = g(y)$$

Binomial Thm: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$

↳ $x = p$ and $y = 1-p$ This is $\text{Bin}(n,p)$: $\underline{X}$ = # heads in n indep tosses

Term $\binom{n}{k} p^k (1-p)^{n-k}$ is the prob of exactly k heads with n tosses (indep) of a coin with prob p of heads

$\underline{X} \sim \text{Bin}(n,p)$ $\quad \underline{X} = \underline{X}_1 + \cdots + \underline{X}_n$, each $\underline{X}_k \sim \text{Bern}(p)$

Want to compute $\sum_{k=0}^{n} k \cdot \text{Prob}(\underline{X} = k)$

↳ need to study $\sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$

↳ maybe $\binom{n}{k} = \frac{n!}{k(k-1)!\,(n-k)!}$ and play algebra games

Know $\displaystyle\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = (x+y)^n$

need to bring down a factor of k, then set $\begin{cases} x = p \\ y = 1-p \end{cases}$

$x\dfrac{d}{dx}$ $\displaystyle\sum_{k=0}^{n} \binom{n}{k} k\, x^k y^{n-k} = n(x+y)^{n-1} x$

$\displaystyle\Rightarrow \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$ (set $x=p$, $y=1-p$)

$= n(p+1-p)^{n-1} p$

$= np$

Variance: do $x\dfrac{d}{dx}\left(x\dfrac{d}{dx}\left(\sim \right)\right)$ gives $E[X^2]$ when $x=p$, $y=1-p$

$$\sum_{k=0}^{n} k \binom{n}{k} x^k y^{n-k} = n(x+y)^{n-1} x$$

$$x\frac{d}{dx} \sum_{k=0}^{n} k^2 \binom{n}{k} x^k y^{n-k} = x \frac{d}{dx}\left[n(n-1)(x+y)^{n-2} x + n(x+y)^{n-1}\right]$$

Set $x = p$, $y = 1-p$ - . . . .

Avoiding Product Rule: $(x+y)^{n-1}(x+y-y)$

$$= (x+y)^n - (x+y)^{n-1} y$$

Normal Random Variable: $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$

wlog: do for $\mu=0, \sigma=1$ and then change variables

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} \frac{1}{2\pi} e^{-(x^2+y^2)/2} dx\,dy$$

$$= \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} \frac{1}{2\pi} e^{-r^2/2} r\,dr\,d\theta$$

$$= \frac{1}{2\pi} 2\pi \, e^{-r^2/2} \Big|_{\infty}^{0} = 1$$

Know $$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\, dx = \sigma$$

$\sigma^a \frac{d}{d\sigma}$

$$\int_{-\infty}^{\infty} \sigma^a \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \frac{d}{d\sigma}\left[-\frac{(x-\mu)^2}{2}\sigma^{-2}\right] dx = \sigma^a$$

$$= \int_{-\infty}^{\infty} \sigma^a \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left[-\frac{(x-\mu)^2}{2}\right]\left[-2\sigma^{-3}\right] dx = \sigma^a$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 \underbrace{\frac{1}{\sqrt{2\pi\sigma^{6-2a}}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}_{\text{if } 6-2a=2, \text{ density of the normal}} dx = \sigma^a$$

variance!

is σ^2 if $6-2a=2$ as that makes $a=2$

Exponential: $f(x) = \frac{1}{\lambda} e^{-x/\lambda}$ $\quad 0 \le x < \infty$ and 0 otherwise

mean: $E[X] = \int_0^\infty x \frac{1}{\lambda} e^{-x/\lambda} dx$

$$u = x \qquad dv = e^{-x/\lambda} dx/\lambda$$

$$du = dx \qquad v = -e^{-x/\lambda}$$

$$E[X] = uv\Big|_0^\infty - \int_0^\infty v\,du$$

$$= -x e^{-x/\lambda}\Big|_0^\infty + \int_0^\infty e^{-x/\lambda} dx$$

$$= 0 - \lambda e^{-x/\lambda}\Big|_0^\infty = \lambda$$

$$\boxed{\frac{x}{e^{x/\lambda}} \le \frac{x}{\frac{1}{2!}(x/\lambda)^2}}$$

L'Hopital:

$$\lim_{x\to\infty} \frac{x}{e^{x/\lambda}} = \lim_{x\to\infty} \frac{1}{\frac{1}{\lambda} e^{x/\lambda}}$$

$$= 0$$

$$E[X^2] = \int_0^\infty x^2 \frac{1}{\lambda} e^{-x/\lambda} dx$$

Know $\int_0^\infty \underbrace{\frac{1}{\lambda} e^{-x/\lambda} dx = 1}$ Same as $\int_0^\infty e^{-x/\lambda} dx = \lambda$

$\lambda^a \frac{d}{d\lambda}$

$$\int_0^\infty \lambda^a e^{-x/\lambda} \frac{d}{d\lambda}\left[-x\lambda^{-1}\right] dx = \lambda^a$$

$$\int_0^\infty \lambda^a x e^{-x/\lambda} \lambda^{-2} dx = \lambda^a$$

$$\int_0^\infty x \underbrace{\lambda^{a-2} e^{-x/\lambda}} = \lambda^a$$

take $a = 1$

This is the exp density!

Mean

So $E[X] = \lambda$

Find deriv of $x \ln 5$

Find deriv of $\ln(2x) - \ln x$

$$= \ln \frac{2x}{x} = \ln 2$$

Thoreau: Simplify, Simplify

Know $\int_0^\infty x\lambda^{a-2} e^{-x/\lambda} dx = \lambda^a$ $\quad a=1$

Gives $\int_0^\infty x \frac{1}{\lambda} e^{-x/\lambda} dx = \lambda$

Now have $\int_0^\infty x e^{-x/\lambda} dx = \lambda^2$

$\lambda^b \frac{d}{d\lambda}$

$$\int_0^\infty \lambda^b x e^{-x/\lambda} \frac{d}{d\lambda}\left[-x\lambda^{-1}\right] dx = \lambda^b \cdot 2\lambda$$

$$\int_0^\infty \underbrace{x^2 \lambda^{b-2} e^{-x/\lambda} dx}_{\text{take } b=1 \text{ get exp distr}} = 2\lambda^{b+1}$$

$$E[X^2] = 2\lambda^2$$

So $Var(X) = E[X^2] - E[X]^2 = 2\lambda^2 - \lambda^2 = \lambda^2$

$X \sim Exp(\lambda)$ so density is $\frac{1}{\lambda} e^{-x/\lambda}$

Mean λ meters

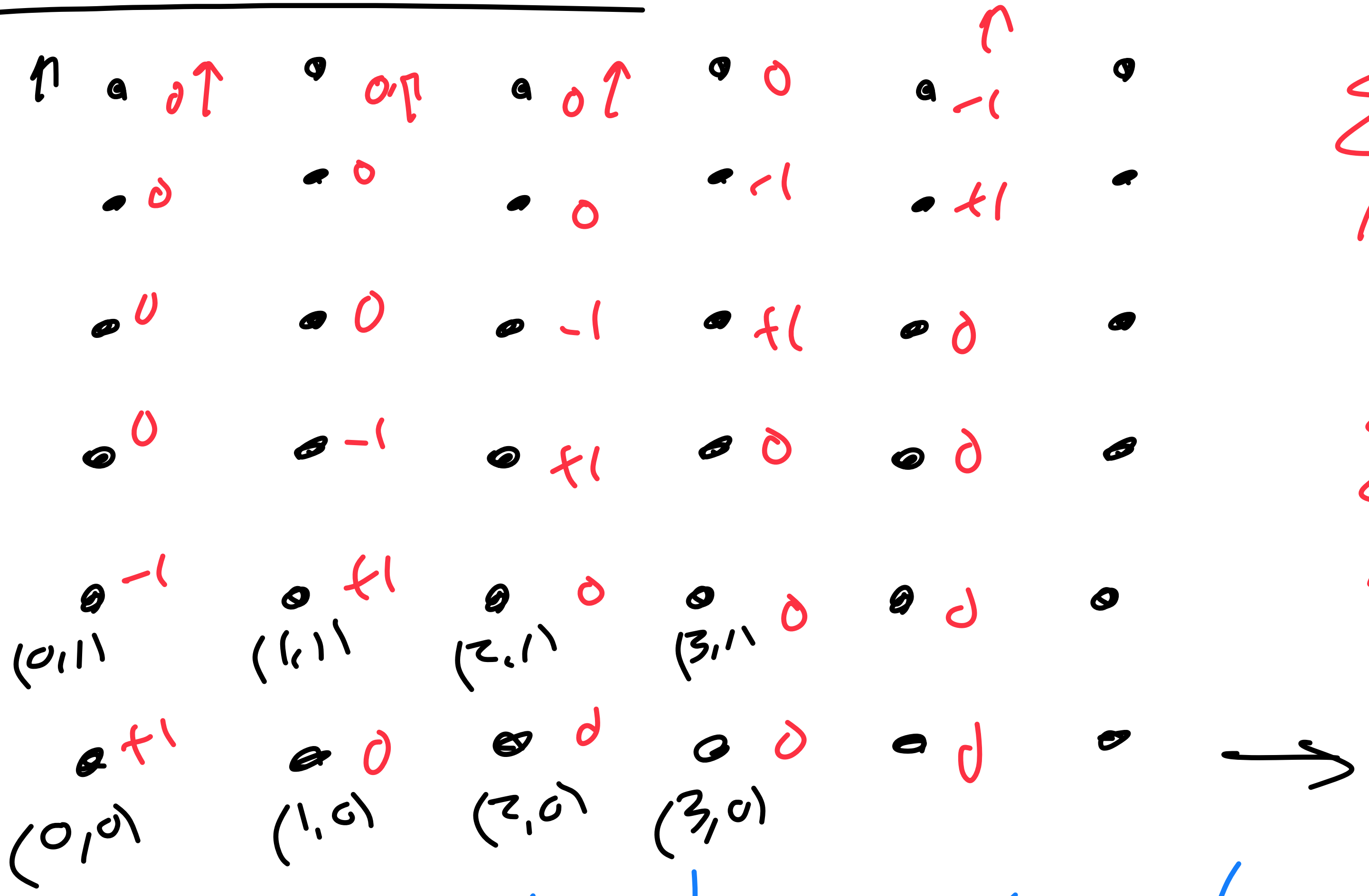
Var λ^2 meters^2 units work

Stdev λ meters

Is it $\frac{x}{\text{mean}}$ or $\frac{x}{\text{stdev}}$ in the argument?

In Normal it was $\frac{(x-\text{mean})^2}{\text{stdev}^2}$

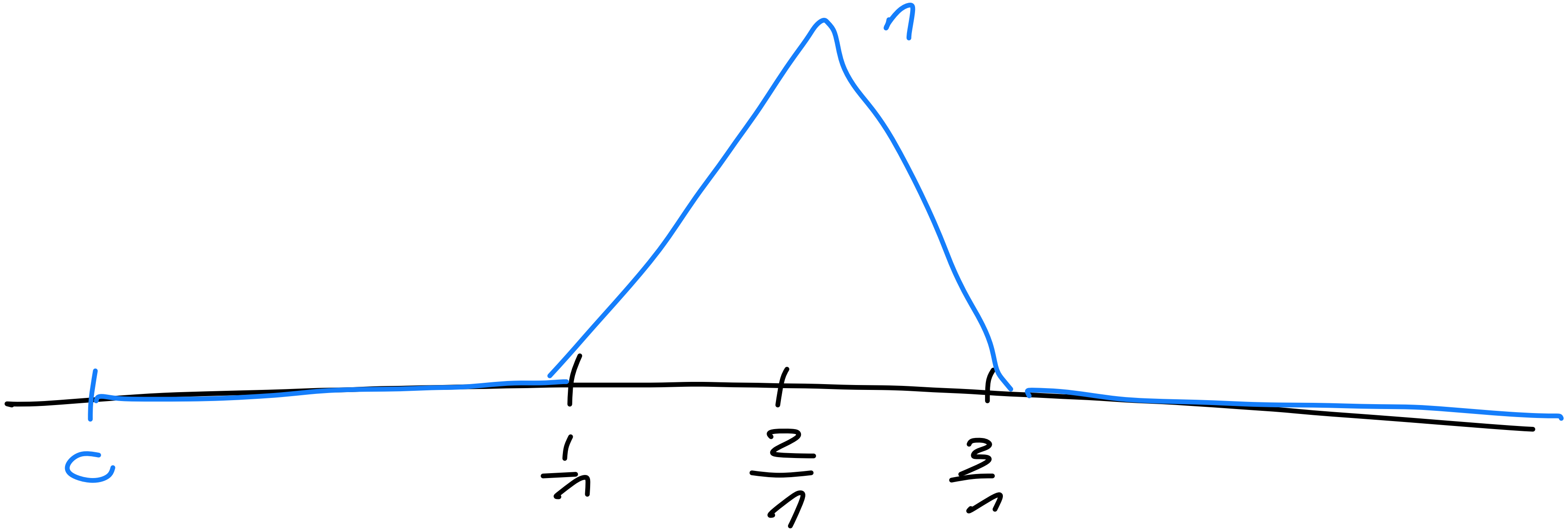
Dangers of interchange



$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{mn} \neq \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn}$$

vertical then horizontal get 0

horizontal then vertical get 1



$f_n(x) = 0$ if $x < \frac{1}{n}$ or $x > 3/n$

and triangle with height n otherwise

centered at $2/n$

$\int_0^\infty f_n(x)\,dx = 1$ so $\lim\limits_{n\to\infty} \int_0^\infty f_n(x)\,dx = 1$

But $\lim\limits_{n\to\infty} f_n(x) = 0$ so $\int_0^\infty \lim\limits_{n\to\infty} f_n(x)\,dx = 0$

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 13: 3/19/25: Differentiating Identities and Applications, Darth Vader, Double Sixes and Marriage Problems
Video: https://youtu.be/BRTWawmZJLs
https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 11: March 11, 2025:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- Differentiating Identities
- Darth Vader Problem, Double Sixes Problem
- Marriage / Secretary Problem

General items.

- Building on previous results
- Gaining feel from subset of data

Differentiating Identities: Geometric Random Variables

Waiting for the first success: p = prob of success

X_p = # (indep) tosses coin with prob p till get success : $X_k \sim \text{Bern}(p)$

Extreme cases: $p=1$: $X_p = 1$ $\quad$ $p=0$: $X_p =$ "∞"

Guess: as $p \uparrow 1$, $E[X_p] \downarrow 1$ and as $p \downarrow 0$, $E[X_p] = \uparrow$ to ∞

Maybe $E[X_p] = 1/p$

What is $\text{Prob}(X_p = n) = (1-p)^{n-1} p$ $\quad$ $\sum_{n=1}^{\infty} \text{Prob}(X_p = n) = 1$

What is $E[X_p] = \sum_{n=1}^{\infty} n \, \text{Prob}(X_p = n)$

$$= \sum_{n=1}^{\infty} n (1-p)^{n-1} p = \frac{p}{1-p} \sum_{n=0}^{\infty} n (1-p)^n \quad \text{if } p \neq 1$$

$$= \frac{p}{1-p} \sum_{n=1}^{\infty} n (1-p)^n$$

Need $\sum_{n=1}^{\infty} n(1-p)^n$ study $\sum_{n=1}^{\infty} n r^n$

Consider $1+r+r^2+\cdots = \sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$ if $|r|<1$

$r\frac{d}{dr}$

$$\sum_{n=1}^{\infty} n r^n = r\left[-(1-r)^{-2}(-1)\right] = \frac{r}{(1-r)^2}$$

If $r=1-p \Rightarrow \sum_{n=1}^{\infty} n(1-p)^n = \frac{1-p}{p^2}$

So $E[X_p] = \frac{p}{1-p} * \frac{1-p}{p^2} = \frac{1}{p}$

The Darth Vader Problem

Only the Emperor is less forgiving than Darth Vader; one mistake and you are dead! No one seems to fail him twice….

If your probability of failing him on a task is p, how many tasks till you die?

The Darth Vader Problem

If your probability of failing him on a task is p,

how many tasks till you die?

Could be unlucky and fail at the first task and die.

Could be very lucky and never fail and live a long, long time….

- What is the probability your first failure is on your first task? p
- What is the probability your first failure is on your second task? $(1-p)\,p$
- What is the probability your first failure is on your third task? $(1-p)^2\,p$
- What is the probability your first failure is on your n^{th} task? $(1-p)^{n-1}\,p$

The EXPECTED VALUE of a random variable is the sum of the product of each value it takes on times the probability it takes on that value.

For this problem: $1 * Prob(first\ fail\ at\ 1) + 2 * Prob(first\ fail\ at\ 2) + \cdots$

$$1 * p + 2 * (1-p)p + 3 * (1-p)2\, p + \cdots + n * (1\text{-}p)^{n\text{-}1}\, p + \cdots \qquad p$$

Lower bound is 1

Lower bound is $p + (1-p)p + (1-p)^2 p + \cdots = p\left[1 + (1-p) + (1-p)^2 + \cdots\right]$

$$= p \frac{1}{1-(1-p)} = 1$$

Improve: $p + 2(1-p) + 2(1-p)^2 p + 2(1-p)^3 p + \cdots$

The Darth Vader Problem: LOWER BOUND

If your probability of failing a task is p, how many tasks till you die?

The EXPECTED VALUE of a random variable is the sum of the product of each value it takes on times the probability it takes on that value.

$$S(p) = p(1 + 2 * (1 - p) + 3 * (1 - p)^2 \ + 4(1 - p)^3 + \cdots)$$

The Darth Vader Problem: UPPER BOUND

If your probability of failing a task is p, how many tasks till you die?

The EXPECTED VALUE of a random variable is the sum of the product of each value it takes on times the probability it takes on that value.

$$S(p) = p(1 + 2 * (1-p) + 3 * (1-p)^2 + 4(1-p)^3 + \cdots)$$

Upper bound: $p(1 + 2(1-p) + (2(1-p))^2 + \cdots)$ $r = 2(1-p)$

$= p \frac{1}{1-2(1-p)} = p \frac{1}{2p-1}$

$p=1 \Rightarrow 1$ good

$p=0 \Rightarrow 0$??

$p=\varepsilon_{small} \rightarrow \frac{\varepsilon}{2\varepsilon-1}$ = negative

only if $|r|<1$

or $2(1-p)<1$

$1-p<\frac{1}{2}$

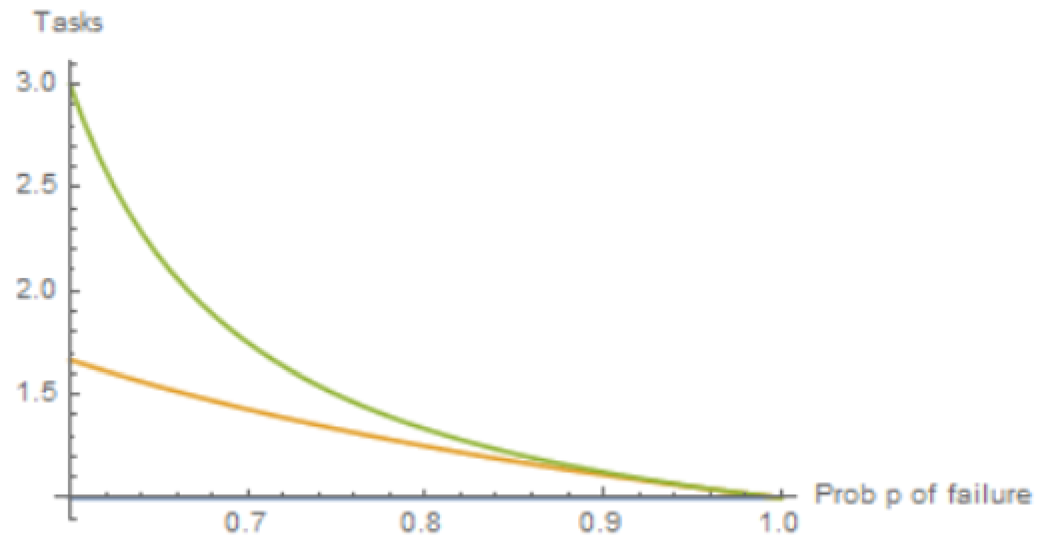
$\frac{1}{2}<p$

always check conds met

$$S(p) = p(1 + 2 * (1 - p) + 3 * (1 - p)^2 + 4(1 - p)^3 + \cdots)$$

Bounds: $If\ (1 - p) < \frac{1}{2}\ so\ p > \frac{1}{2}$

$Then\ 1 \le S(p) \le \frac{p}{1 - 2(1-p)}$.



The Darth Vader Problem

Probability of failing a task is p, how many tasks till you die?

$$S(p) = p(1 + 2 * (1-p) + 3 * (1-p)^2 + 4(1-p)^3 + \cdots)$$

Let $q = 1-p$. Note this is $p(1 + 2q + 3q^2 + 4q^3 + \cdots)$.

We can rewrite: It is

$p(1 + q + q^2 + q^3 + \cdots) + p(q + q^2 + q^3 + \cdots) + p\,(q^2 + q^3 + q^4 + \cdots) + \cdots$

Each is a geometric series with ratios q, q, q, … but different starting terms.

$S(p) = p\,(1 + q + q^2 + \cdots) + pq\,(1 + q + q^2 + \cdots) + pq^2\,(1 + q + q^2 + \cdots) + \cdots$

$S(p) = (p + pq + pq^2 + pq^3 + \cdots)\frac{1}{1-q} = p\,(1 + q + q^2 + q^3 + \cdots)\frac{1}{1-q} = p\frac{1}{1-q}\frac{1}{1-q}$

Thus S(p) = 1/p as claimed! And without calculus!

The Sixes Game

Probability of failing a task is p, how many tasks till you die?

Answer: Expect 1/p

Equivalently, if the probability of a success is p, the number of tasks or tries you need before the first success is 1/p.

We can use this to study a new game!

The sixes game: you roll a fair die until you get a 6. How many rolls do you expect before this happens?

expected # rolls is 6

The Double Sixes Game

You have two fair die.

On each turn you can roll one or both of the die.

The goal is to have both show a 6.

Thus once one of the die lands on a 6 you can stop rolling it.

Questions:

- How many rolls do you expect before you have double sixes?
- What is the probability you win on your first turn? On your second? On your n^{th}?

Can we use the Darth Vader Theorem here? Why or why not? no!

sometimes need both sixes, sometimes just one

The Double Sixes Game: Upper/Lower Bounds

lower bounds $\quad 1 \le 6 \le$

forget about 2nd die

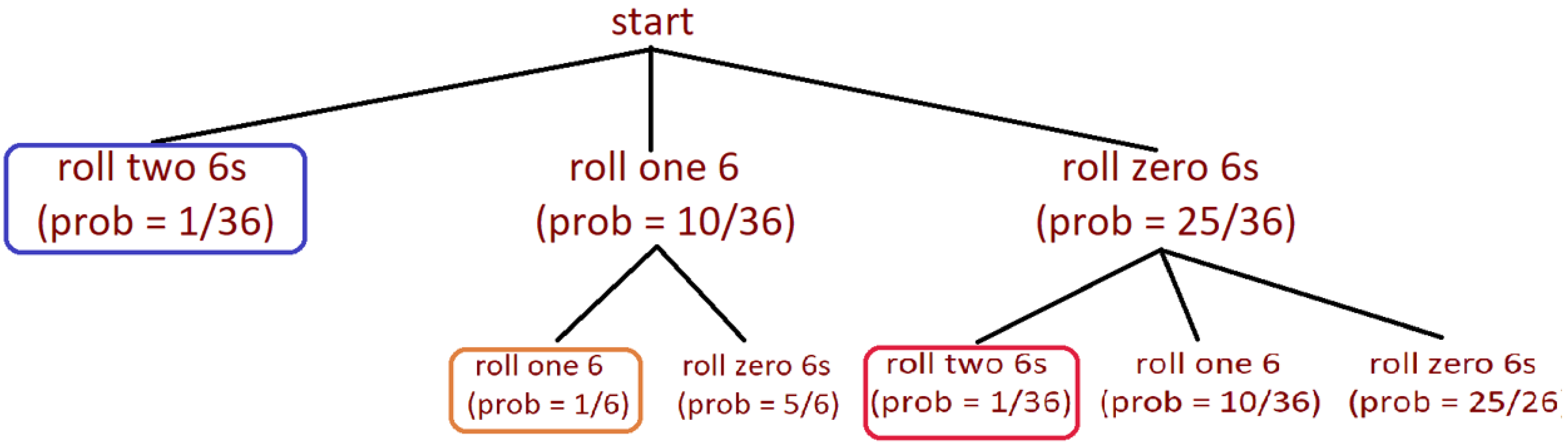
upper bounds $\quad 36 \ge 12 \ge$

do one then the other

The Double Sixes Game

You have two fair die. On each turn you can roll one or both of the die.

The goal is to have both show a 6. Thus once one of the die lands on a 6 you can stop rolling it.



Prob(win first roll) = 1/36. Prob(win second roll) = 10/36*1/6 + 25/36*1/36 = 85/1296

The Double Sixes Game

You have two fair die. On each turn you can roll one or both of the die.

Want both to show a 6. Once one of the die lands on a 6 you can stop rolling it.

The Law of Complementary Events: If the probability something happens is p, then the probability it does not happen is 1-p.

The Law of Double Counting: The probability A or B happens is the sum of the probability each happens minus the probability they both happen: Prob(A or B) = Prob(A) + Prob(B) – Prob(A and B).

What is the probability we win by the n^{th} turn?

It is 1 minus the probability we have NOT won.

What is the probability we haven't won? It is....

A: 1st die never a 6 in n rolls

B: same for second

$$\left(\frac{5}{6}\right)^n + \left(\frac{5}{6}\right)^n - \left(\frac{25}{36}\right)^n$$

The Double Sixes Game

You have two fair die. On each turn you can roll one or both of the die.

The goal is to have both show a 6. Thus once one of the die lands on a 6 you can stop rolling it.

The Law of Complementary Events: If the probability something happens is p, then the probability it does not happen is 1-p.

What is the probability we win BY the n^{th} turn? $1 - 2*(5/6)^n + (25/36)^n$.

It is 1 minus the probability we have NOT won.

What is the probability we haven't won? It is $(5/6)^n + (5/6)^n - (25/36)^n$.

So…, what is the probability we win ON the n^{th} turn?

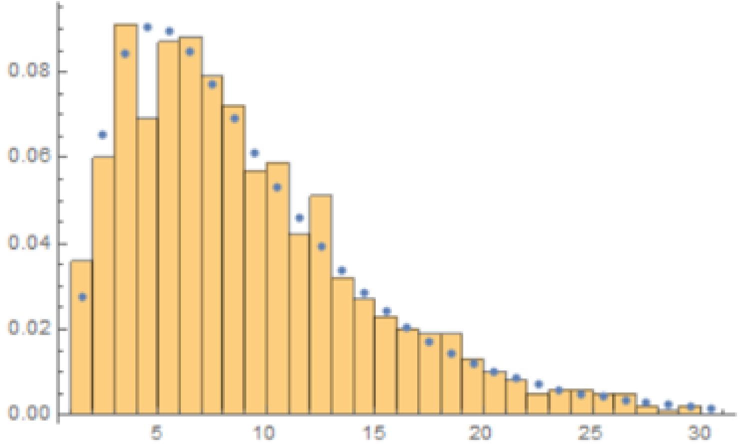
It is the probability we win BY the n^{th} turn MINUS the probability we win BY the $(n-1)^{st}$ turn! $(1 - 2*(5/6)^n + (25/36)^n) - (1 - 2*(5/6)^{n-1} + (25/36)^{n-1})$

The Double Sixes Game

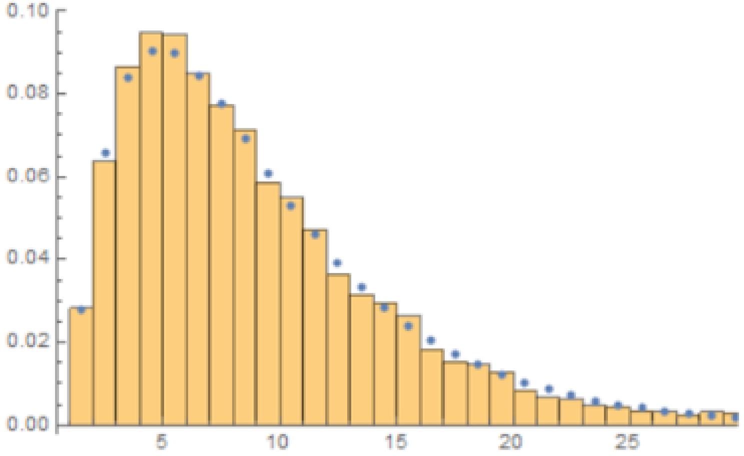
You have two fair die. On each turn you can roll one or both of the die.

The goal is to have both show a 6. Thus once one of the die lands on a 6 you can stop rolling it.

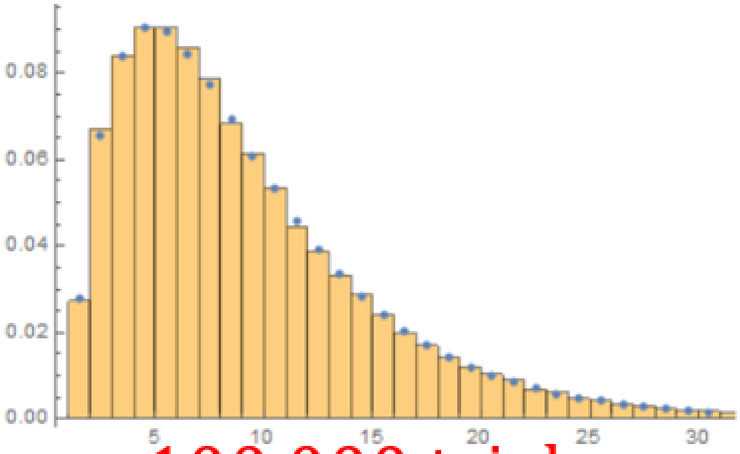
Probability win on n^{th} turn: $(2/6)(5/6)^{n-1} - (11/36)(25/36)^{n-1}$.



100 trials



10,000 trials



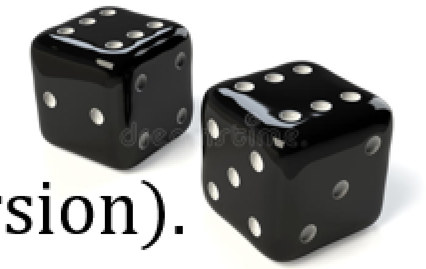
100,000 trials

The Double Sixes Game: Code

Mathematica code to simulate

```
In[68]:= f[n_] := 2 (5 / 6) ^n - (25 / 36) ^n
         g[n_] := 1 - f[n] (* probability succeed by n *)
         success[n_] := g[n] - g[n - 1];
         (* probability succeed at n *)

In[71]:= doublesixes[numdo_] := Module[{},
           count = {};
           For[m = 1, m ≤ numdo, m++,
            {
             firstdie = 0; seconddie = 0; rolls = 0;
             While[firstdie + seconddie < 12,
              {
               rolls = rolls + 1;
               die1 = RandomInteger[{1, 6}];
               die2 = RandomInteger[{1, 6}];
               If[die1 == 6, firstdie = 6];
               If[die2 == 6, seconddie = 6];
              }];
             count = AppendTo[count, rolls];
            }];
           theory = {};
           For[k = 1, k ≤ 30, k++, theory = AppendTo[theory, {k + .5, success[k]}]];
           Print[Show[Histogram[count, Automatic, "Probability"], ListPlot[theory]]];
          ]
```

The Double Sixes Game: Expected Value



Need the FULL strength of the Darth Vader Theorem (friendly version).

The Darth Vader Theorem: If the probability of a success is p, then the expected number of trials until a success is 1/p. Furthermore:
$S(p) = p(1 + 2 * (1-p) + 3 * (1-p)^2 + 4(1-p)^3 + \cdots) = 1/p.$

Notation: $\sum_{n=1}^{\infty} a_n \quad means \;\; a_1 + a_2 + a_3 + \cdots$ (using a Greek Sigma for Sum)

We have $\sum_{n=1}^{\infty} n((2/6)(5/6)^{n-1} - (11/36)(25/36)^{n-1})$.

Equals $2 * \frac{1}{6} \sum_{n=1}^{\infty} n\,(1 - 1/6)^{n-1}$ - $\frac{25}{11}\,\frac{11}{36} \sum_{n=1}^{\infty} n\,(1 - 11/36)^{n-1}$.

What is the first term? $2 * \frac{1}{1/6}$ What is second? $\frac{1}{11/36}$.

Answer is $2 * 6 \;-\; \frac{36}{11} = \frac{96}{11}$ (or about 8.7 rolls until you get both sixes).

Review: Big Takeaways

The Darth Vader Theorem: If the probability of a success is p, then the expected number of trials until a success is 1/p. Furthermore:
$S(p) = p(1 + 2 * (1-p) + 3 * (1-p)^2 + 4(1-p)^3 + \cdots) = 1/p.$

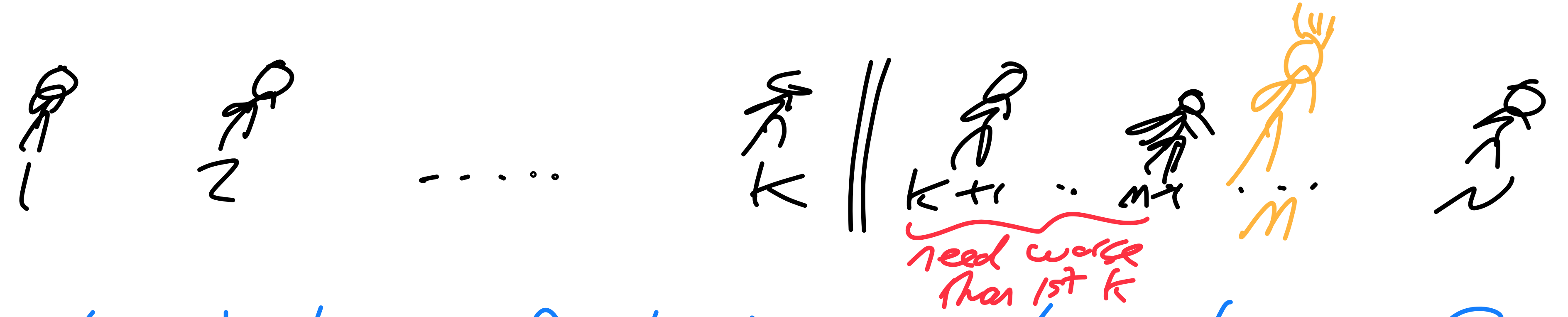
The Law of Complementary Events: If the probability something happens is p, then the probability it does not happen is 1-p.

The Law of Double Counting: The probability A or B happens is the sum of the probability each happens minus the probability they both happen: Prob(A or B) = Prob(A) + Prob(B) – Prob(A and B).

The Power of Algebra: Sometimes have to do a bit of algebraic manipulations to make what you have look like something you know.

Marriage / Secretary Problem

https://www.youtube.com/watch?v=dafvzF66vzY&t=135s



Fix k, look at first k people, choose first person we see who is better than best of the first k. (Use to estimate the field)

Let m be spot of maximum value: if $m \leq k$, lose

What is the prob the best of the first $m-1$ in first k?

↳ It is $\dfrac{k}{m-1}$

Conditional Prob

$$\text{Prob}(W_m) = \sum_{m=k+1}^{N} \text{Prob}\left(\begin{matrix}\text{Best of first} \\ m-1 \text{ is in the} \\ \text{first } k\end{matrix} \,\middle|\, \begin{matrix}\text{Best of} \\ \text{all is} \\ \text{at } m\end{matrix}\right) \underbrace{\text{Prob}\left(\begin{matrix}\text{Best of} \\ \text{all is} \\ \text{at } m\end{matrix}\right)}_{1/N}$$

$$= \frac{1}{N} \sum_{m=k+1}^{N} \frac{k}{m-1}$$

$$= \frac{k}{N} \sum_{m=k+1}^{N} \frac{1}{m-1} = \frac{k}{N}\left[\frac{1}{k} + \frac{1}{k+1} + \frac{1}{k+2} + \cdots + \frac{1}{N-1}\right]$$

$$= \frac{k}{N}\left[\left(1 + \frac{1}{2} + \cdots + \frac{1}{N-1}\right) - \left(1 + \frac{1}{2} + \cdots + \frac{1}{k-1}\right)\right] = \frac{k}{N}\left[H_{N-1} - H_{k-1}\right]$$

$$H_\ell = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{\ell}$$

Harmonic Numbers: $H_\ell = 1 + \frac{1}{2} + \cdots + \frac{1}{\ell} \simeq \log(\ell)$

Note $H_\ell \simeq \int_1^\ell \frac{1}{x}\,dx = \ln x\Big|_1^\ell = \ln(\ell)$

$$H_{N-1} - H_{K-1} \simeq H_N - H_K = \log N - \log K \text{ or } \log \frac{N}{K}$$

$\text{Prob}(\text{Win using Strategy } K) = \frac{K}{N} \log \frac{N}{K}$ $\quad K = xN$

$$f(x) = x \log\left(\frac{1}{x}\right) = -x \log x$$

$f(x) = -x \log x$ endpoints $x \to 0$ and $x \to 1$

as $x \to 0$ prob goes to 0, as $x \to 1$ prob $\to 0$

$f'(x) = -\left[1 \cdot \log x + x \frac{1}{x}\right]$ set $= 0$ for critical points

so $f'(x) = 0 \implies \log x + 1 = 0 \implies \log x = -1 \implies x = \frac{1}{e}$

$x = e^{-1}$

$k = xN = \frac{1}{e} N$

$\text{Prob}(\text{win with } k = \frac{1}{e} N) = f(\frac{1}{e}) = -\frac{1}{e} \log \frac{1}{e} = \frac{1}{e} > \frac{1}{3}$

$$\lim_{x\to 0} x\log x \qquad x \text{ as } \frac{1}{n} \text{ and } x\to 0 \text{ same as } n\to\infty$$

$$= \lim_{n\to\infty} \frac{1}{n}\log\left(\frac{1}{n}\right)$$

$$= -\lim_{n\to\infty} \frac{\log(n)}{n} \underset{\text{L'Hopital}}{=} -\lim_{n\to\infty} \frac{1/n}{1} = 0$$

So yes, $\lim_{x\to 0} x\log x = 0$

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 14: 4/08/25: Differentiating Identities: (Gaussian, Exponential, Geometric, Negative Binomial), Sums of Uniform Random Variables, Sums of Gaussian Random Variables, Cauchy Distribution, Gregory-Leibnitz Formula:
Video: https://youtu.be/FjfRqQnJlLo

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 14: April 8, 2025:

- Differentiating Identities (Gaussian, Exponential, Geometric, Negative Binomial)
- Sums of Uniform, Gaussian Random Variables
- Cauchy Distribution and Gregory-Leibnitz Formula:

General items.

- Care in interchanging order….
- Integrating without integrating!
- Lectures from previous years for more examples/details:

Lecture 19: 10/29/21: Differentiating Identities: (Gaussian, Exponential, Geometric, Negative Binomial): https://youtu.be/oYRoyqV2jAI (slides)

Lecture 20: 11/01/21: Sums of Uniform Random Variables, Sums of Gaussian Random Variables, Cauchy Distribution, Gregory-Leibnitz Formula: https://youtu.be/qW-3bHAwdPU (slides)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Gaussian!

$$f(x;\beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & x \geq 0, \\ 0 & x < 0. \end{cases} \qquad f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Exponential

$$\Pr(X = k) = (1-p)^{k-1} p$$

Geometric

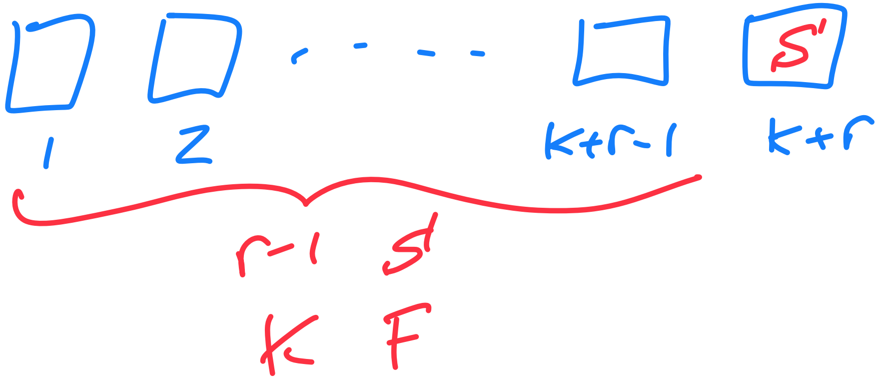
$k \in \{1, 2, 3, \dots\}$ counting attempts

$$f(k; r, p) \equiv \Pr(X = k) = \binom{k+r-1}{r-1} (1-p)^k p^r$$

Negative Binomial

counting number of failures

$k \in \{0, 1, 2, \dots\}$



$$\binom{k+r-1}{r-1} (1-p)^k p^r$$

strings

Negative binomial distribution

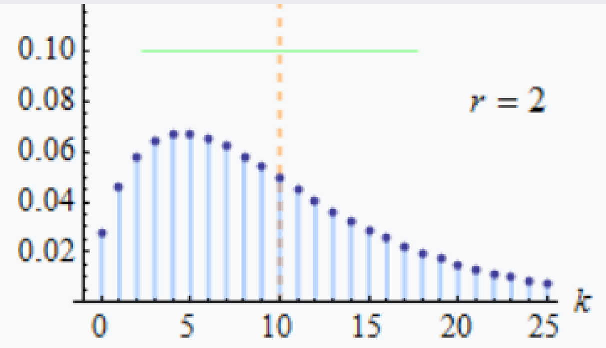
Contents hide

From Wikipedia, the free encyclopedia

In probability theory and statistics, the **negative binomial distribution** is a discrete probability distribution that models the number of failures in a sequence of independent and identically distributed Bernoulli trials before a specified/constant/fixed number of successes r occur.[2] For example, we can define rolling a 6 on some dice as a success, and rolling any other number as a failure, and ask how many failure rolls will occur before we see the third success ($r = 3$). In such a case, the probability distribution of the number of failures that appear will be a negative binomial distribution.

Different texts (and even different parts of this article) adopt slightly different definitions for the negative binomial distribution. They can be distinguished by whether the support starts at *k* = 0 or at *k* = *r*, whether *p* denotes the probability of a success or of a failure, and whether *r* represents success or failure,[1] so identifying the specific parametrization used is crucial in any given text.

Probability mass function



The orange line represents the mean, which is equal to 10 in each of these plots; the green line shows the standard deviation.

Notation	$\mathrm{NB}(r,\, p)$

An alternative formulation is to model the number of total trials (instead of the number of failures). In fact, for a specified (non-random) number of successes (r), the number of failures ($n - r$) is random because the number of total trials (n) is random. For example, we could use the negative binomial distribution to model the number of days n (random) a certain machine works

Negative Binomial: The mean by differentiating identities, linearity of expectation, recurrences.

$P(\text{exactly } k \text{ failures}) = \binom{k+r-1}{r-1}(1-p)^k p^r$

Know $\displaystyle\sum_{k=0}^{\infty}\binom{k+r-1}{r-1}(1-p)^k p^r = 1$

$\displaystyle\sum_{k=0}^{\infty}\binom{k+r-1}{r-1}(1-p)^k = p^{-r}$

$(1-p)\dfrac{d}{dp}$

$\displaystyle\sum_{k=0}^{\infty}\binom{k+r-1}{r-1}(-k)(1-p)^k = -r\,p^{-r-1}(1-p)$

$\displaystyle\sum_{k=0}^{\infty} k\cdot\binom{k+r-1}{r-1}(1-p)^k p^r = r\,\frac{1-p}{p}$

mean # of failures before r successes is $r\,\dfrac{1-p}{p} = \dfrac{r}{p} - r$

mean # of attempts before r successes is $\dfrac{r}{p}$

View as a sum of r indep Geom(p)

Linearity of Expectation

If $X_{r,p}$ is the random var to have r success with prob p of success, counting attempts, then it is the sum of r indep Geom(p) random vars waiting for first success

$$X_{r,p} = G_{1,p} + G_{2,p} + \cdots + G_{r,p}$$

By Lin of Expectation:

$$E[X_{r,p}] = r E[G_{k,p}] = \frac{r}{p}$$

Recurrence Relation Proof

$\mu_{r,p}$ = ave # of attempts before r successes
with prob p of success each time.

Claim/Conjecture that $\mu_{r,p} = r/p$

Know $\mu_{1,p} = 1/p$ (Geometric Case)

Proof by Induction

- Base Case Done
- Assume $\mu_{r,p} = r/p$

Then $\mu_{r+1,p} = \underbrace{(\mu_{r,p}+1)\,p}_{\text{start with success}} + \underbrace{(\mu_{r+1,p}+1)(1-p)}_{\text{start with failure}}$

$$\mu_{r+1,p} = (\mu_{r,p}+1)\,p + (\mu_{r+1,p}+1)(1-p)$$

Start with success   start with failure

↑ I know by assump this is r/p

$$\mu_{r+1,p} = \left(\frac{r}{p}+1\right)p + (1-p)\,\mu_{r+1,p} + (1-p)$$

$$p \cdot \mu_{r+1,p} = r + p + 1 - p$$

$$\mu_{r+1,p} = \frac{r+1}{p}$$ as desired!

Convolutions: If X, Y indep R.V. with densities f_X and f_Y, if $Z = X+Y$

Then $f_Z(z) = \int_{-\infty}^{\infty} f_X(t)\, f_Y(z-t)\, dt$

$$f_Z(z) = (f_X * f_Y)(z)$$

Sums of Unif random variables (indep)

$X_1, X_2 \sim \mathrm{Unif}(0,1)$

densities $p(X) = 1$ if $0 \le x \le 1$ and 0 otherwise

$X = X_1 + X_2$ condition!

$$p_X(x) = (p_{X_1} * p_{X_2})(x) = (p * p)(x)$$

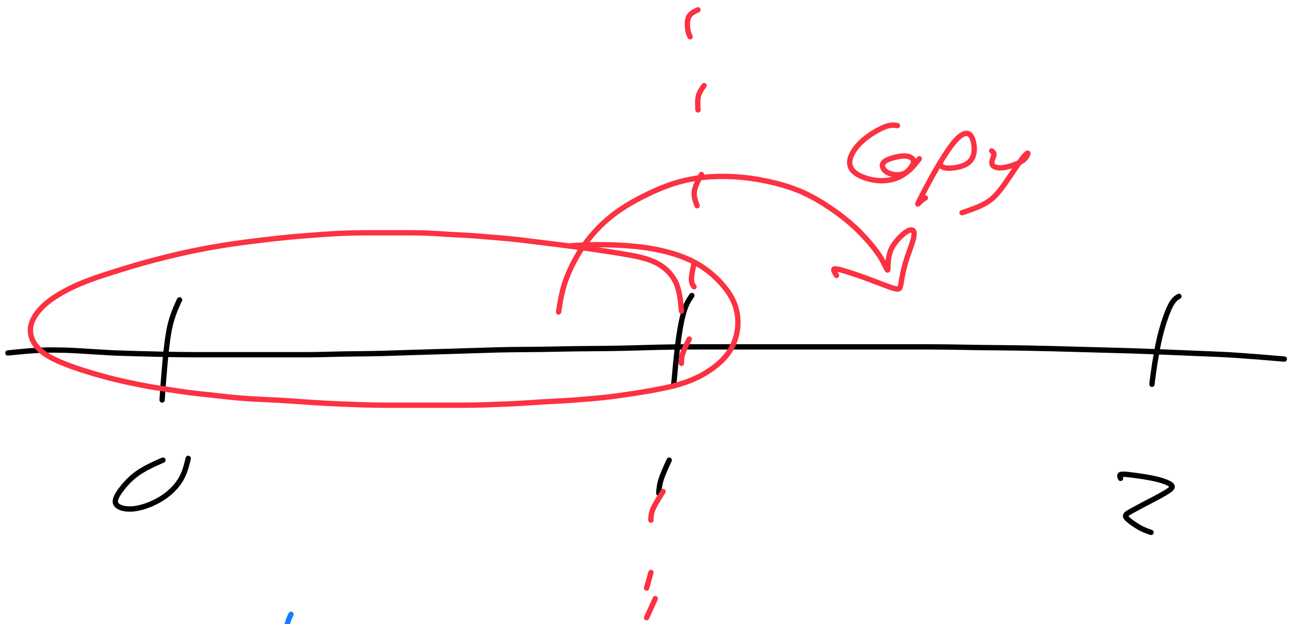
$$= \int_{-\infty}^{\infty} \underbrace{p(t)}_{\text{forces } 0 \le t \le 1} p(x-t)\, dt$$

$$= \int_0^1 1\, \underbrace{p(x-t)}_{\substack{0 \le x-t \le 1 \\ t \le x \le 1+t}} dt$$

Symmetry: Claim $X = X_1 + X_2$ is symm about 1

If X_k is Unif$(0,1)$ so is $Y_k = 1 - X_k$

$X_1 + X_2$ look at $Y_1 + Y_2 = 2 - (X_1 + X_2)$



Prob of sum of $\frac{1}{4}$ = Prob of a sum of $2 - \frac{1}{4} = \frac{7}{4}$

$$p_X(x) = \int_0^1 1 \; p(x-t)\, dt$$

Wlog, $0 \le x \le 1$

$$0 \le x - t \le 1$$

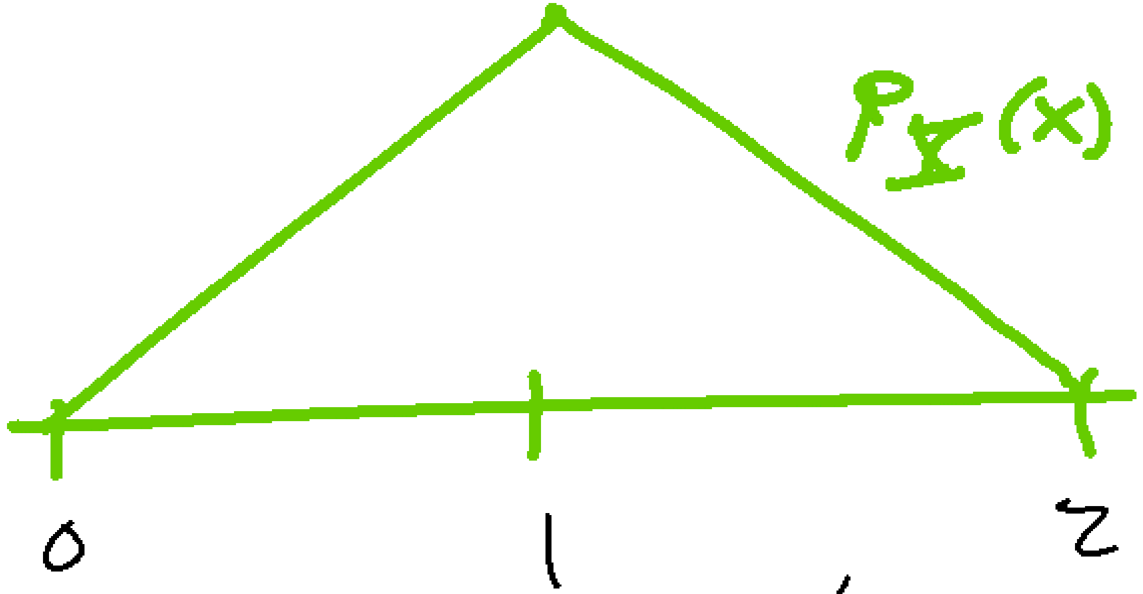
$$t \le x \le 1+t$$

trivial if $x \le 1$

Get trivially $x \le 1+t$

integral restricted to $t \le x$

$$= \int_{t=0}^{x} 1 \cdot 1 \, dt = x$$



Check

(1) non-neg ✓

(2) area is 1 ✓

Caveat: region of integration!

<u>Gaussian</u>: wlog $\mu = 0, \sigma = 1$

X_1, X_2 indep $N(0,1)$

$X = X_1 + X_2$? $E[X] = 0 = E[X_1] + E[X_2]$

$Var(X) = Var(X_1) + Var(X_2) = 1 + 1 = 2$

so $StDev(X) = \sqrt{2}$

Stable distribution: same shape when sum indep elements

Generalize

$X_1 \sim N(0, \sigma_1^2) \qquad X_2 \sim N(0, \sigma_2^2)$

$$\underbrace{\underbrace{(Y_1 + Y_2 + Y_3 + \cdots + Y_{n-1})}_{\text{Gaussian}} + \underbrace{Y_n}_{\text{Gaussian}}}_{\text{Gaussian}}$$

$$\left(\left((Y_1 + Y_2) + Y_3\right) + Y_4\right) + Y_5 \qquad \text{Grouping } (\ )$$

$X_i \sim N(0, \sigma_i^2)$ $\quad P_i(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-x^2/2\sigma_i^2}$

$$\int_{-\infty}^{\infty} P(t)\,P(x-t)\,dt = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-t^2/2\sigma_1^2}\, e^{-(x-t)^2/2\sigma_2^2}\,dt$$

$\Rightarrow$ If can show x-dependence is that of a Gaussian, must be $N(0, \sigma_1^2+\sigma_2^2)$

Theory of Normalization Constants

$$\begin{cases} A e^{-x/3} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

If know this a prob distribution, x-dependence is that of Exp(3). So $A = 1/3$.

$$\int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} e^{-t^2/2\sigma_1^2} e^{-(x-t)^2/2\sigma_2^2}\,dt$$

Exponential is $-\frac{1}{2\sigma_1^2}t^2 - \frac{1}{2\sigma_2^2}(x-t)^2 = -ax^2 - b(t+cx)^2$

$$-\frac{1}{2\sigma_1^2}t^2 - \frac{1}{2\sigma_2^2}(x^2 - 2xt + t^2) \iff -a\cdot x^2 - b(t^2 + 2ctx + c^2x^2)$$

$$\left(-\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}\right)t^2 + \frac{xt}{\sigma_2^2} - \frac{x^2}{2\sigma_2^2} \qquad (-a - bc^2)x^2 - 2bcxt - bt^2$$

Match t^2: forces b on us

match xt: forces c on us

match x^2: forces a

$$\int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} e^{-ax^2} e^{-b(t+cx)^2} dt \qquad \underline{-ax^2 - b(t+cx)^2}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} e^{-ax^2} \int_{-\infty}^{\infty} e^{-b(t+cx)^2} dt \qquad u = t + cx$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} e^{-ax^2} \underbrace{\int_{-\infty}^{\infty} e^{-bu^2} du}_{\text{some number, say } B}$$

$$= \frac{B}{2\pi\sigma_1\sigma_2} e^{-ax^2}$$

x-dependence is that of a Gaussian!

So $a = \frac{1}{2(\sigma_1^2+\sigma_2^2)}$

$$\frac{B}{2\pi\sigma_1\sigma_2} = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}$$

Gaussians are Stable!

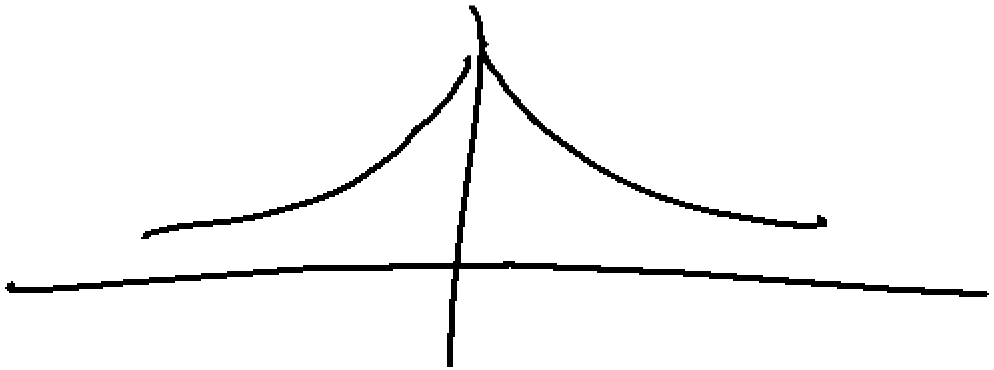
Gaussian + Gaussian = Gaussian

Uniform is not Stable!

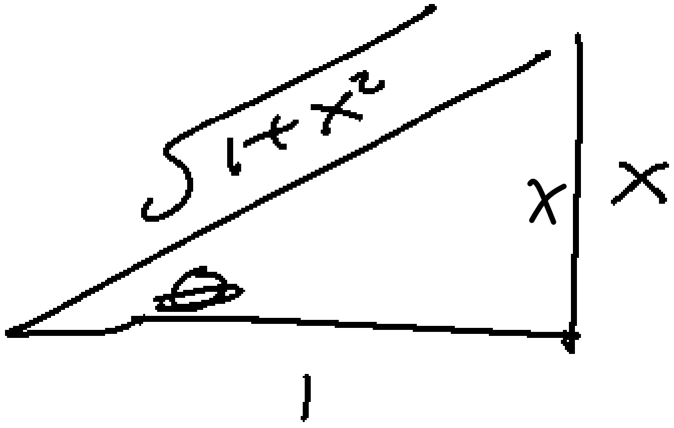
What else is Stable?

Cauchy Distribution

$$p(x) = \frac{1}{\pi} \; \frac{1}{1+x^2}$$



non-neg ✓ $\int$ to 1 ✓



$$\tan(\arctan x) = x$$

$$\tan'(\arctan x) \cdot \arctan'(x) = 1$$

$$\arctan'(x) = \frac{1}{\tan'(\arctan x)}$$

$$= \cos^2(\arctan x) = \frac{1}{1+x^2}$$

$$\int_0^1 \frac{1}{(1+x^2)}\,dx = \arctan(x)\Big|_0^1 = \frac{\pi}{4} - 0 = \frac{\pi}{4}$$

$$\int_0^1 \frac{1}{1+x^2}\,dx = \int_0^1 \frac{1}{1-(-x^2)}\,dx \qquad r = -x^2$$

$$= \int_0^1 \left[1 + (-x^2) + (-x^2)^2 + (-x^2)^3 + \cdots\right] dx$$

$$= \int_0^1 (1 - x^2 + x^4 - x^6 + x^8 - \cdots)\,dx$$

$$= \int_0^1 1\,dx - \int_0^1 x^2\,dx + \int_0^1 x^4\,dx - \int_0^1 x^8\,dx + \cdots$$

overestimate

$$= 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots = \frac{\pi}{4}$$

underestimate

Gregory Leibniz Formula

•In Economics, the standard random walk hypothesis seems to have lost most of its supporters, though there are variants (and I'm not familiar with all); see also the efficient market hypothesis and technical analysis, and all the links there. (There are also many good links on the wikipedia page on Eugene Fama). Two famous books (with different conclusions) are Malkiel's A random walk down wall street and Mandelbrot-Hudson's The (mis)behavior of markets (a fractal view of risk, ruin and reward). Some interesting papers if you want to read more:

- **Mandelbrot: Variation on certain speculative prices (a must read!)**
- Fama: Mandelbrot and Stable Paretian Hypothesis
- Fama: Random Walks Stock Prices
- For more on randomness, check out The Black Swan by Taleb (amazon.com page here, wikipedia page here).
- For more on fractal geometry, click here. See the Koch snowflake; another popular one is the Cantor set. See here for fractal dimensions. To actually compute pictures of items like the Mandelbrot set, one needs to iterate polynomials. This can lead to the fascinating subject of efficient algorithms; when I wrote such programs years ago on what would now be considered `slow' computer, I had to use Horner's algorithm to get things to run in a reasonable time.

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 15: 4/10/25: Sabermetrics:
Slides for talk: https://web.williams.edu/Mathematics/sjmiller/public_html/math/talks/PythagWLTalk_WNE2024.pdf and https://web.williams.edu/Mathematics/sjmiller/public_html/math/talks/pythagintrostats_williamsalumns_sanfran2024.pdf
Video: https://youtu.be/EZq9XKLSzPA

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 16: 4/15/25: Gamma Function, Chi-Square Distribution, Markov inequality:
Video: https://youtu.be/0IqoNdzPI84

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 16: April 15, 2021:

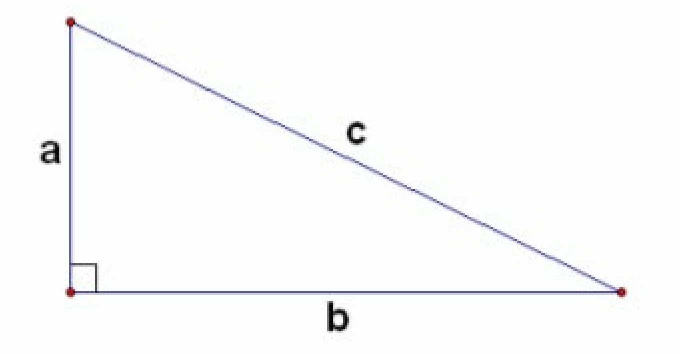
https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- Pythagorean Theorem (not the formula!)
- Gamma function
- Chi-Square Distribution
- Markov's Inequality
- Chebyshev's Inequality
- Divide and Conquer
- Newton's Method

General items.

- Dimensional Analysis
- The more you assume, the more you can deduce…

Geometry Gem: Pythagorean Theorem



Theorem (Pythagorean Theorem)

Right triangle with sides a, b and hypotenuse c, then $a^2 + b^2 = c^2$.

Most students know the statement, but the proof?

Why are proofs important? Can help see big picture.

Geometric Proofs of Pythagoras

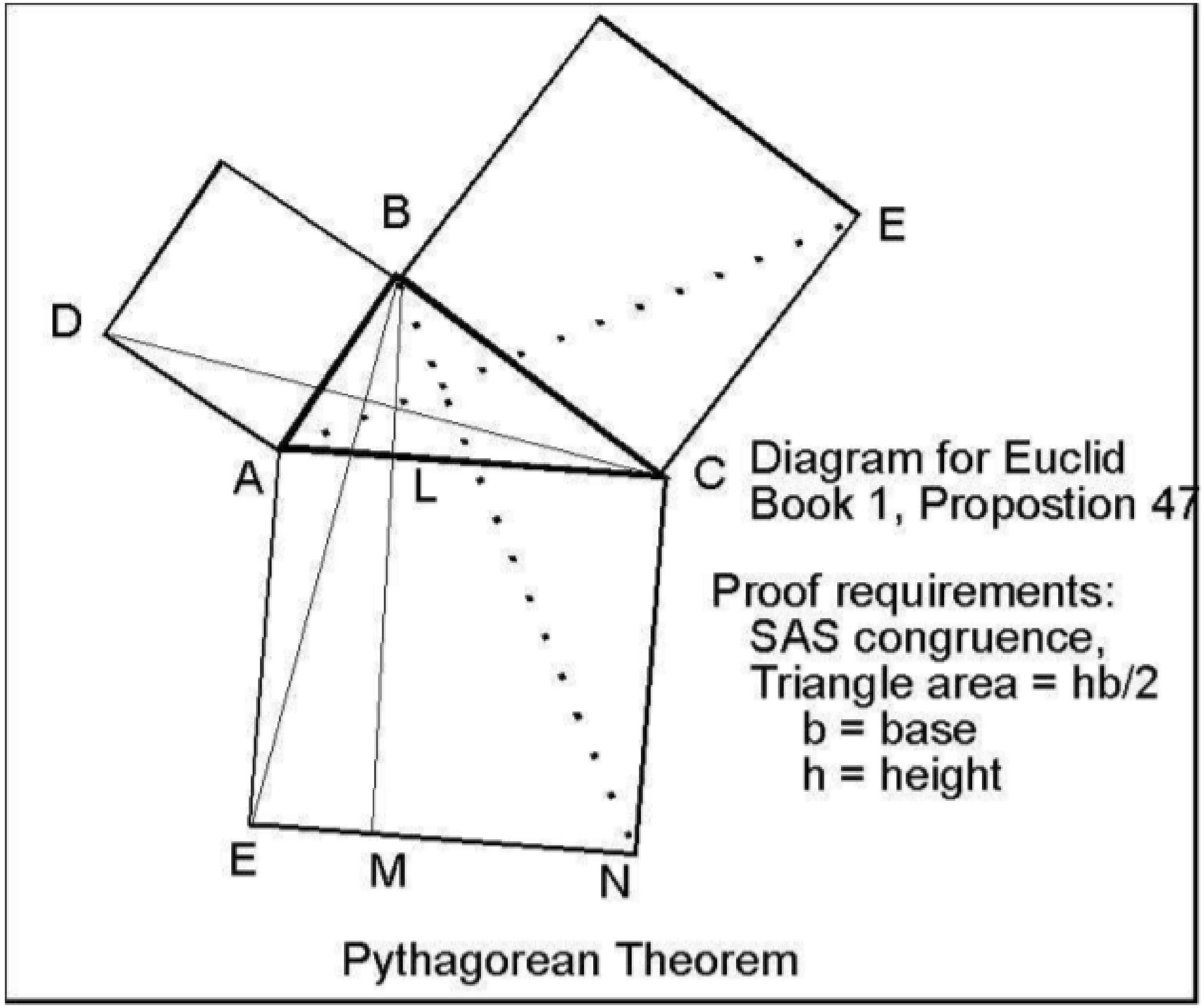


Figure: Euclid's Proposition 47, Book I. Why these auxiliary lines? Why are there equalities?

Geometric Proofs of Pythagoras

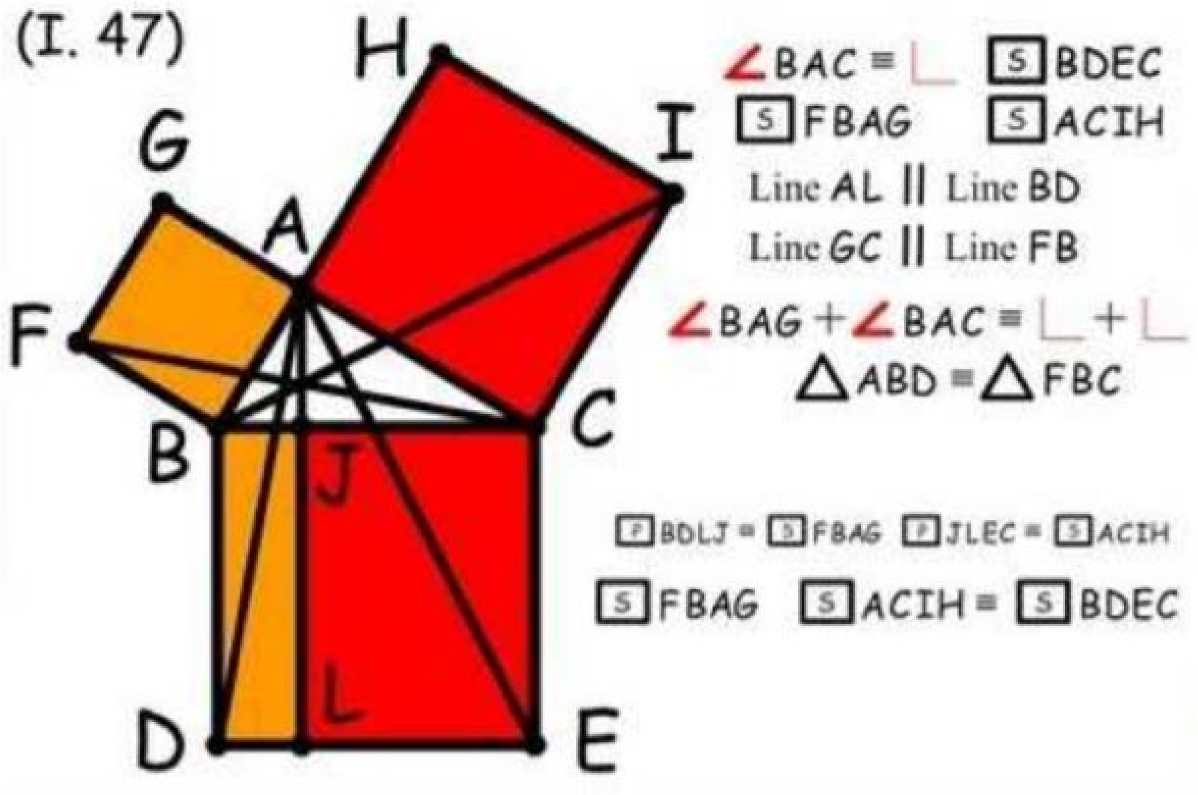


Figure: Euclid's Proposition 47, Book I. Why these auxiliary lines? Why are there equalities?

Geometric Proofs of Pythagoras

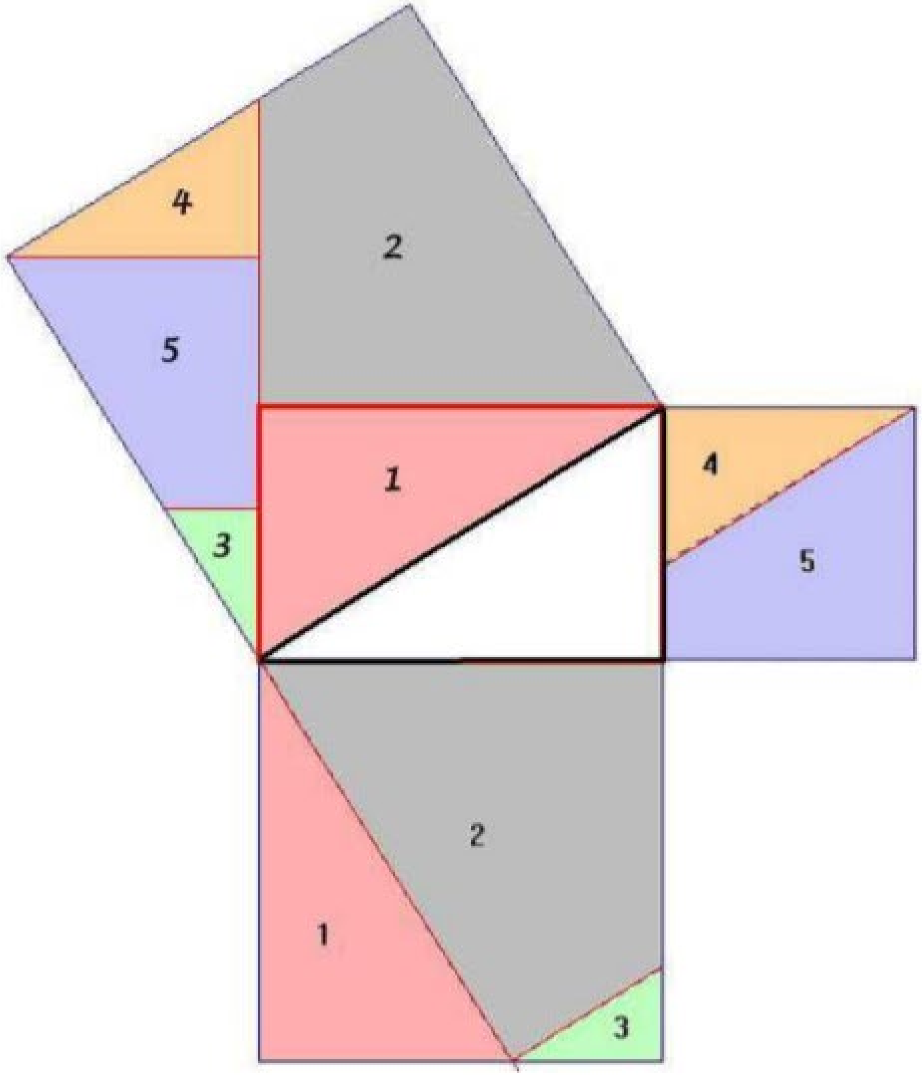
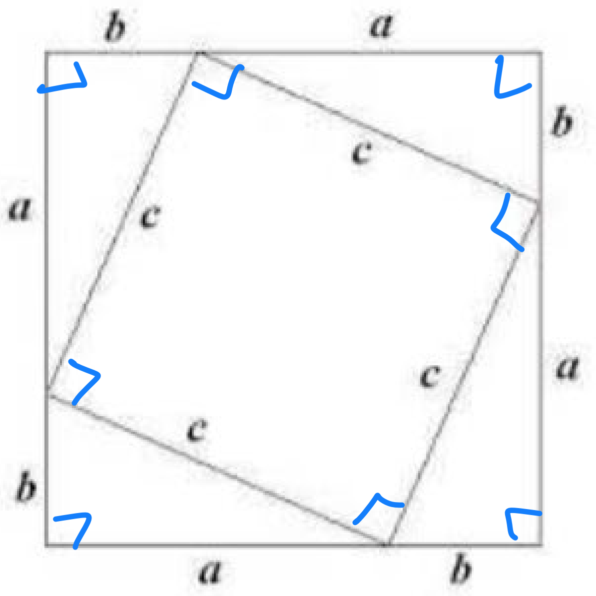


Figure: A nice matching proof, but how to find these slicings!



Big square: $(a+b)^2$

$= a^2 + 2ab + b^2$

Four triangles $= 2ab$

Little square $= c^2$

$a^2 + 2ab + b^2 = c^2 + 2ab$

$a^2 + b^2 = c^2$

Figure: Four triangles proof: I

Geometric Proofs of Pythagoras

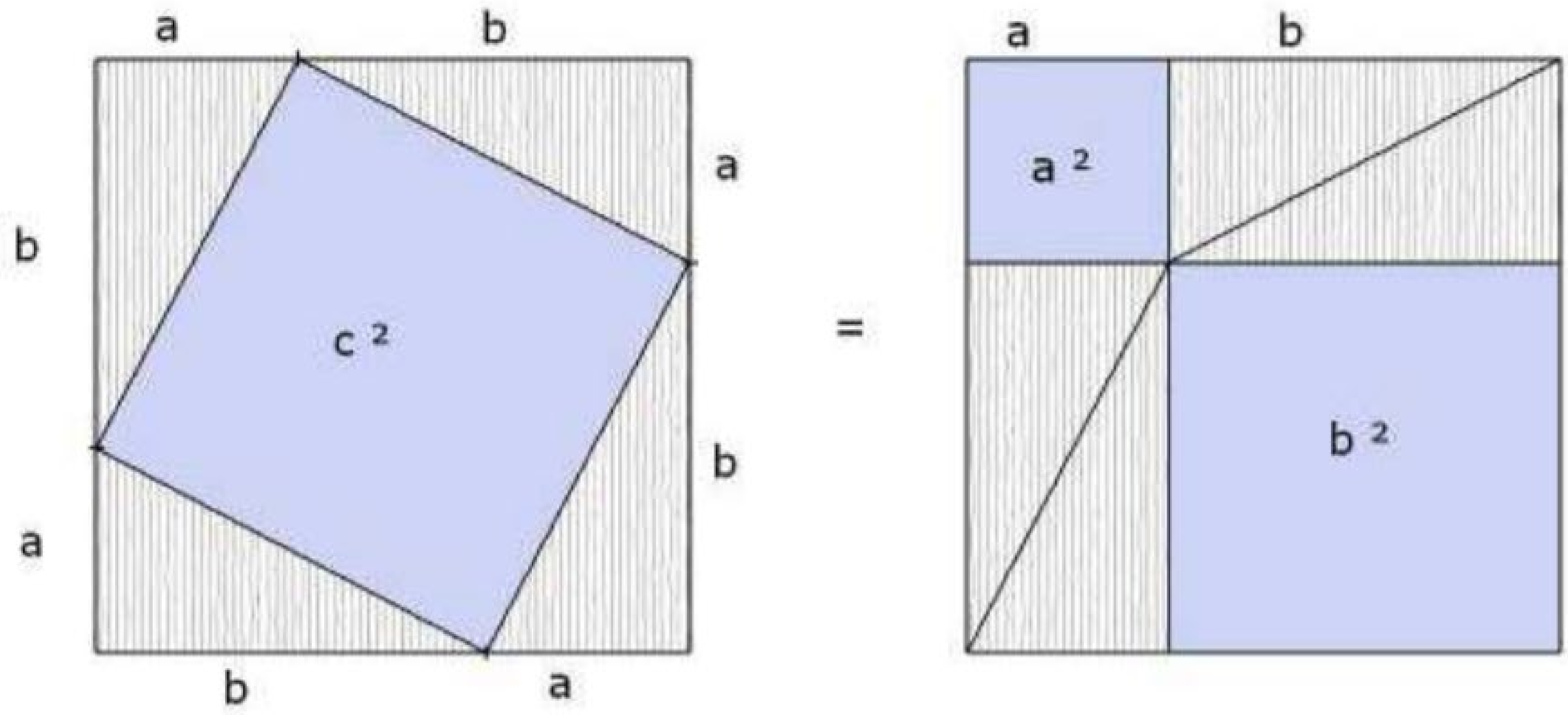


Figure: Four triangles proof: II

Geometric Proofs of Pythagoras

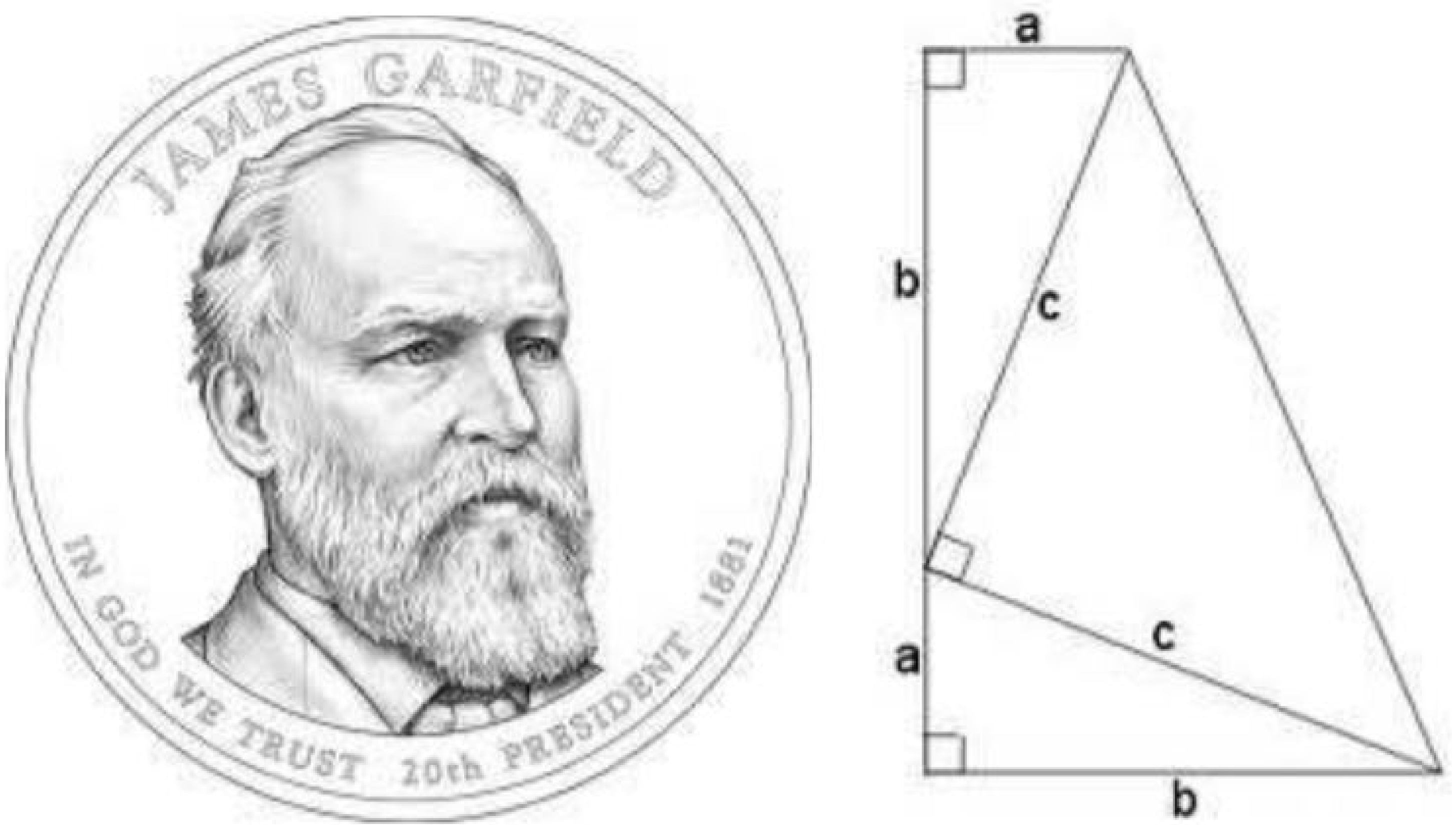


Figure: President James Garfield's (Williams 1856) Proof.

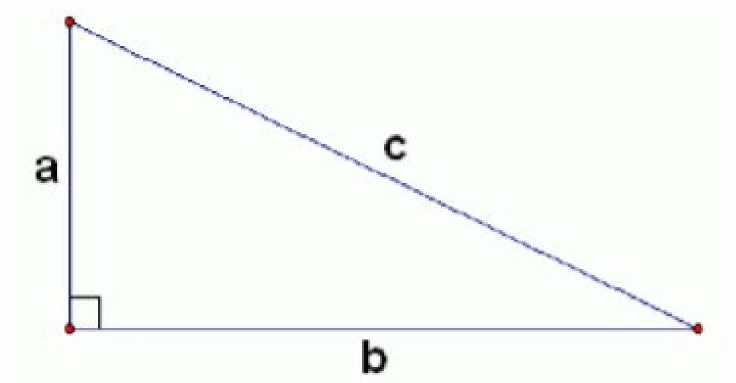
Geometric Proofs of Pythagoras

Lots of different proofs.

Difficulty: how to find these combinations?

At the end of the day, do you know *why* it's true?

Possible Pythagorean Theorems....

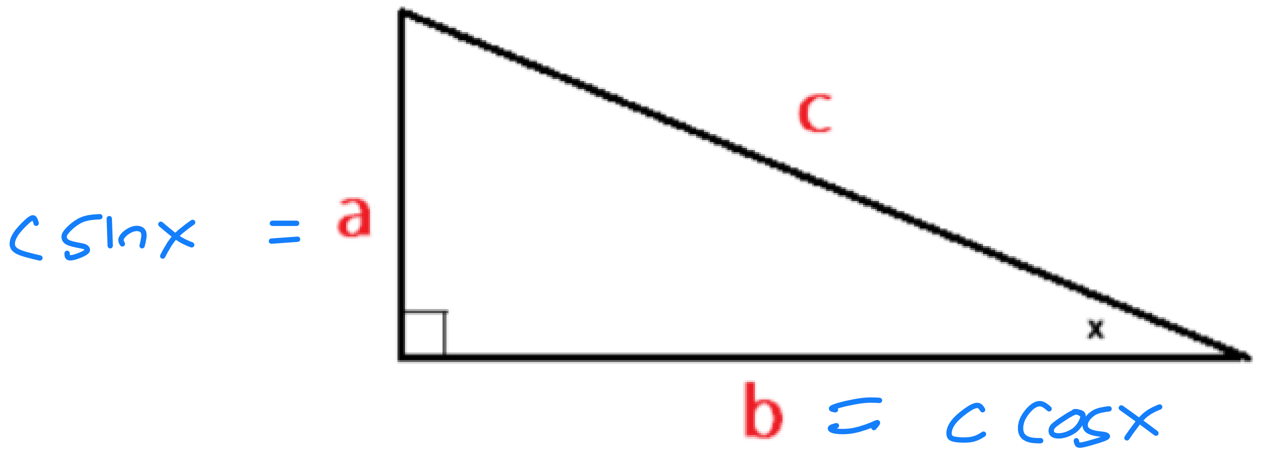


- $c^2 = a^3 + b^3$.
- $c^2 = a^2 + 2b^2$.
- $c^2 = a^2 - b^2$.
- $c^2 = a^2 + ab + b^2$.
- $c^2 = a^2 + 110ab + b^2$.

Possible Pythagorean Theorems....

- $c^2 = a^3 + b^3$. No: wrong dimensions.
- $c^2 = a^2 + 2b^2$. No: asymmetric in a, b.
- $c^2 = a^2 - b^2$. No: can be negative.
- $c^2 = a^2 + ab + b^2$. Maybe: passes all tests.
- $c^2 = a^2 + 110ab + b^2$. No: violates $a + b > c$.

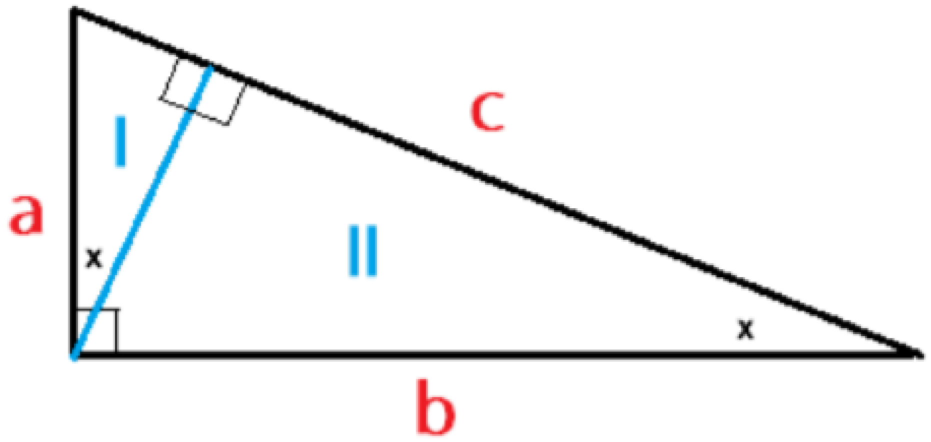
Dimensional Analysis Proof of the Pythagorean Theorem



$\text{area} = \frac{1}{2} c^2 \sin x \cos x$

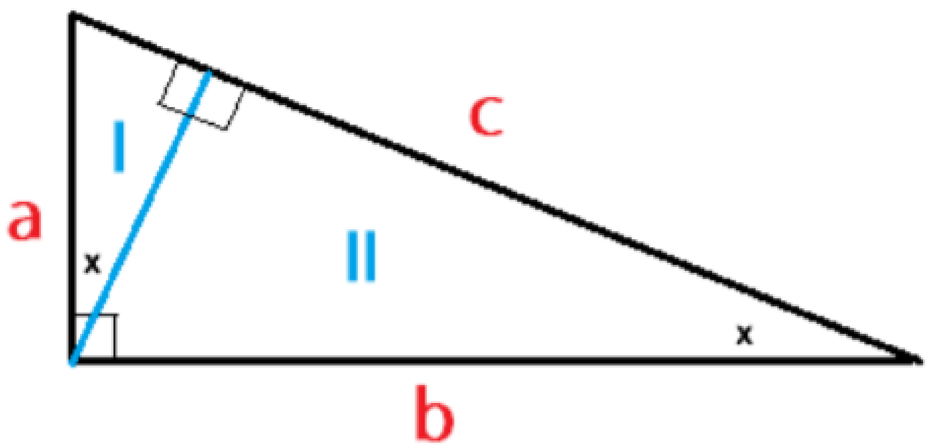
- Area is a function of hypotenuse c and angle x.
- $\mathrm{Area}(c, x) = f(x)c^2$ for some function f (similar triangles).
- Must draw an auxiliary line, but where? Need right angles!

Dimensional Analysis Proof of the Pythagorean Theorem



- ◇ Area is a function of hypotenuse c and angle x.
- ◇ $\mathrm{Area}(c, x) = f(x)c^2$ for some function f (CPCTC).
- ◇ Must draw an auxiliary line, but where? Need right angles!

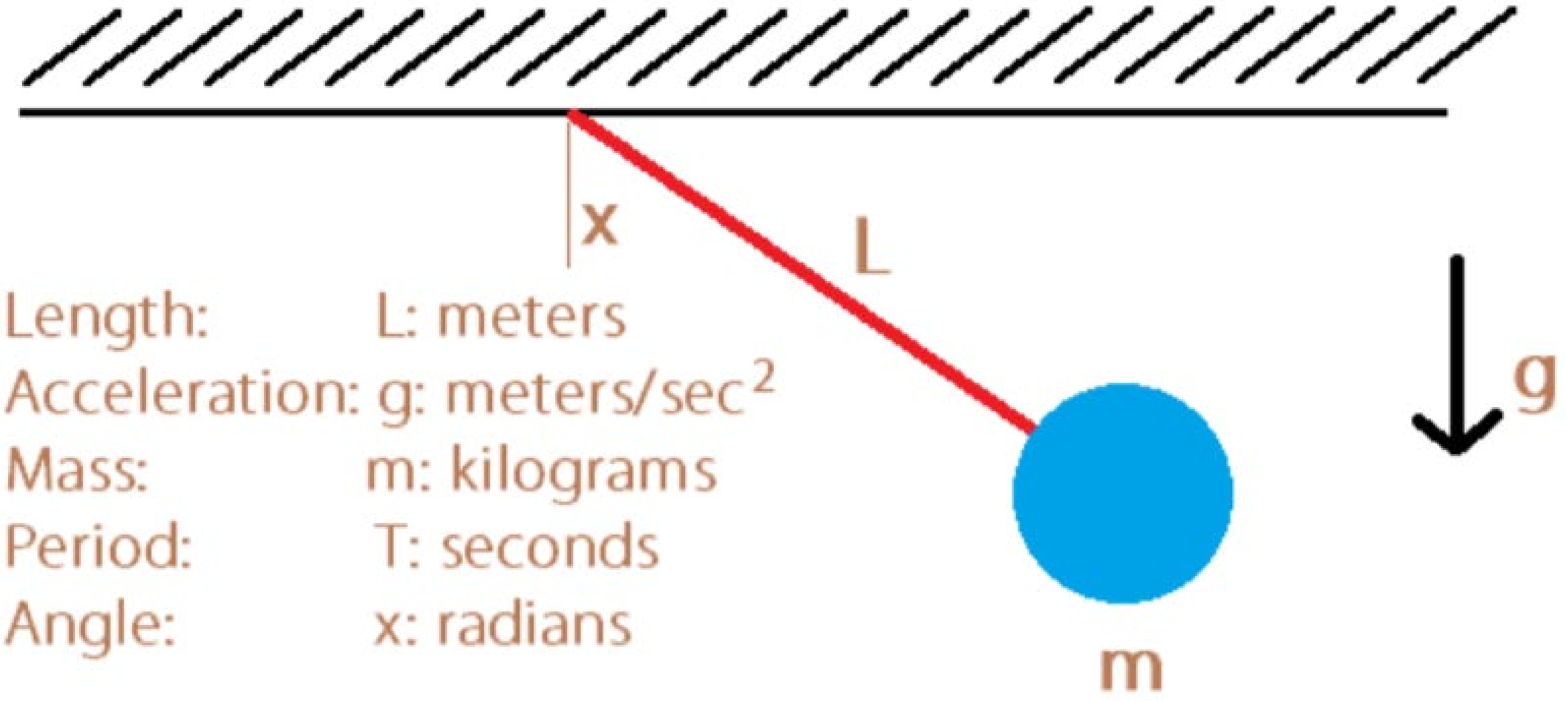
Dimensional Analysis Proof of the Pythagorean Theorem



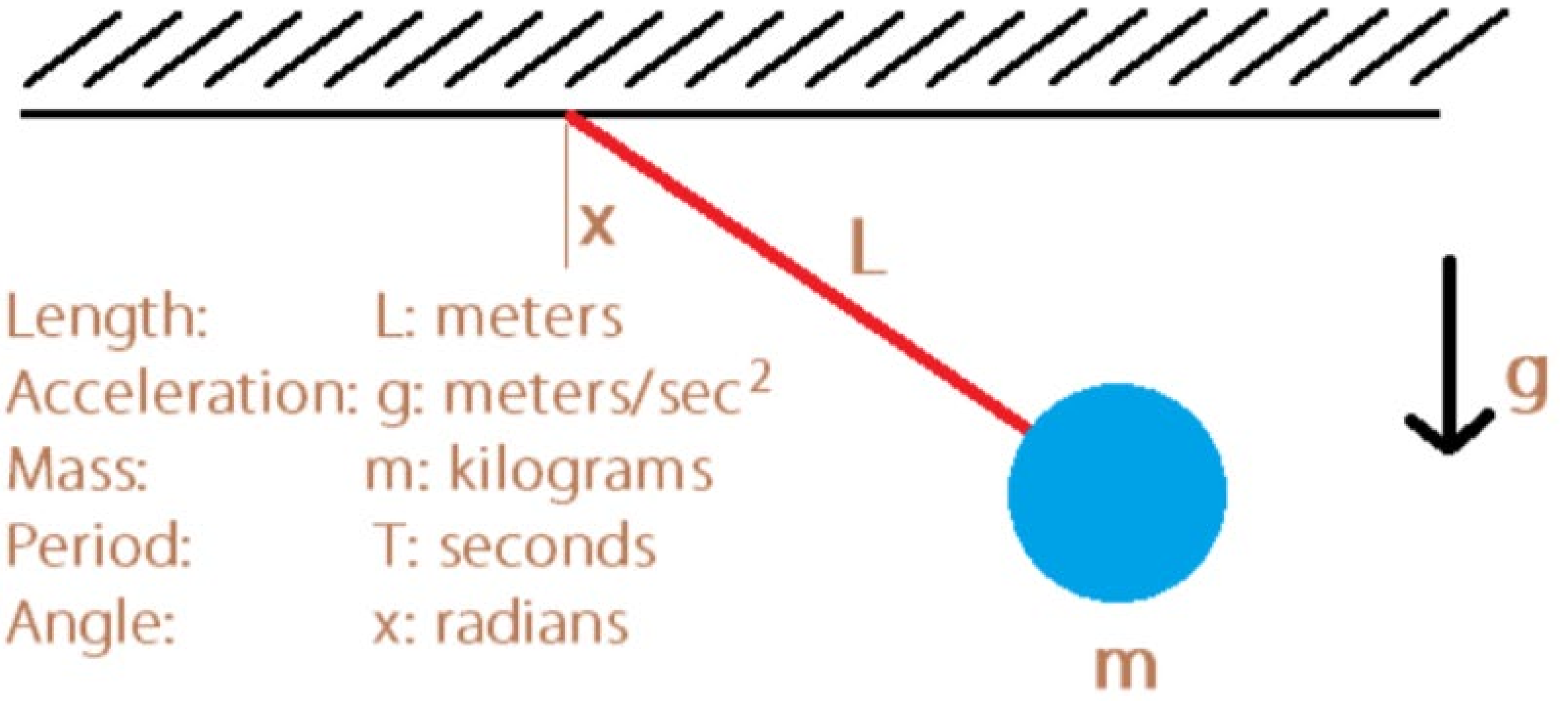
- ◇ Area is a function of hypotenuse c and angle x.
- ◇ $\mathrm{Area}(c, x) = f(x)c^2$ for some function f (CPCTC).
- ◇ Must draw an auxiliary line, but where? Need right angles!
- ◇ $f(x)a^2 + f(x)b^2 = f(x)c^2 \Rightarrow a^2 + b^2 = c^2$.

ass|u|me

Dimensional Analysis and the Pendulum



Dimensional Analysis and the Pendulum



Period: Need combination of quantities to get seconds.

$$T = f(x)\sqrt{L/g}.$$

For $s > 0$ (or actually $\Re(s) > 0$), the **Gamma function** $\Gamma(s)$ is

$$\Gamma(s) := \int_0^\infty e^{-x} x^{s-1} dx = \int_0^\infty e^{-x} x^s \frac{dx}{x}.$$

Existence of $\Gamma(s)$

- at ∞ ok: e^{-x} decays much faster than x^{s-1} grows

$$\frac{x^{s-1}}{e^x} \le \frac{x^{s-1}}{x^m/m!} \le \frac{1}{x^2} \quad \text{if } m > |s| + 100$$

- near 0, e^{-x} looks like $e^0 = 1$

need $\int_0^\varepsilon x^{s-1}\,dx$ to be finite $\quad \left|\int_0^\varepsilon x^{s-1}\,dx\right| \le \int_0^\varepsilon x^{Re(s)-1}\,dx$

if s is real: $\int_0^\varepsilon x^{s-1}\,dx = \begin{cases} x^s/s\,\Big|_0^\varepsilon & \text{if } s \ne 0 \\ \ln x\,\Big|_0^\varepsilon & \text{if } s = 0 \end{cases}$

ok if $Re(s) > 0$

not ok if $Re(s) \le 0$

Functional equation of $\Gamma(s)$: The Gamma function satisfies $(\mathrm{Re}(s)>0)$

$$\Gamma(s+1) \;=\; s\Gamma(s).$$

This allows us to extend the Gamma function to all s. We call the extension the Gamma function as well, and it's well-defined and finite for all s save the negative integers and zero.

If $\Gamma(1)=1$

$$\left.\begin{aligned}\Gamma(1+1) &= \Gamma(2) = 1\cdot\Gamma(1) = 1 = 1!\\ \Gamma(2+1) &= \Gamma(3) = 2\cdot\Gamma(2) = 2 = 2!\\ \Gamma(3+1) &= \Gamma(4) = 3\,\Gamma(3) = 3\cdot 2 = 3!\end{aligned}\right\}\ \Gamma(n+1) = n!$$

Proof: $\Gamma(s+1) = \int_0^\infty e^{-x} x^{s+1-1}\,dx = \int_0^\infty e^{-x} x^{s}\,dx$

$u = x^s \qquad dv = e^{-x}dx \qquad \int_0^\infty u\,dv = uv\Big|_0^\infty - \int_0^\infty v\,du$

$du = sx^{s-1} \qquad v = -e^{-x}$

$\Gamma(s+1) = -x^s e^{-x}\Big|_0^\infty + \int_0^\infty s\,e^{-x}x^{s-1}dx = s\int_0^\infty e^{-x}x^{s-1}dx = s\Gamma(s)$

Note $\Gamma(1) = \int_0^\infty e^{-x}x^0\,dx = \int_0^\infty e^{-x}dx = 1$

$\Gamma(s)$ and the Factorial Function. If n is a non-negative integer, then $\Gamma(n+1) = n!$. Thus the Gamma function is an extension of the factorial function.

$$\Gamma(4) = \int_0^\infty e^{-x} x^{4-1}\,dx$$

$$= \int_0^\infty e^{-x} x^3\,dx$$

$$\Gamma(1) = \text{\_\_\_\_\_\_\_\_}$$

The cosecant identity. If s is not an integer, then

$$\Gamma(s)\Gamma(1-s) \; = \; \pi \csc(\pi s) \; = \; \frac{\pi}{\sin(\pi s)}.$$

$$\Gamma(1/2) = \sqrt{\pi}.$$ view as $(-1/2)!$

Fundamental Relation of the Beta Function: For $a, b > 0$ we have

$$B(a,b) \; := \; \int_0^1 t^{a-1}(1-t)^{b-1}dt \; = \; \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Beta distribution: Let $a, b > 0$. If X is a random variable with the **Beta distribution** with parameters a and b, then its density is

$$f_{a,b} \; = \; \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1}dt & \text{if } 0 \leq t \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

We write $X \sim B(a,b)$.

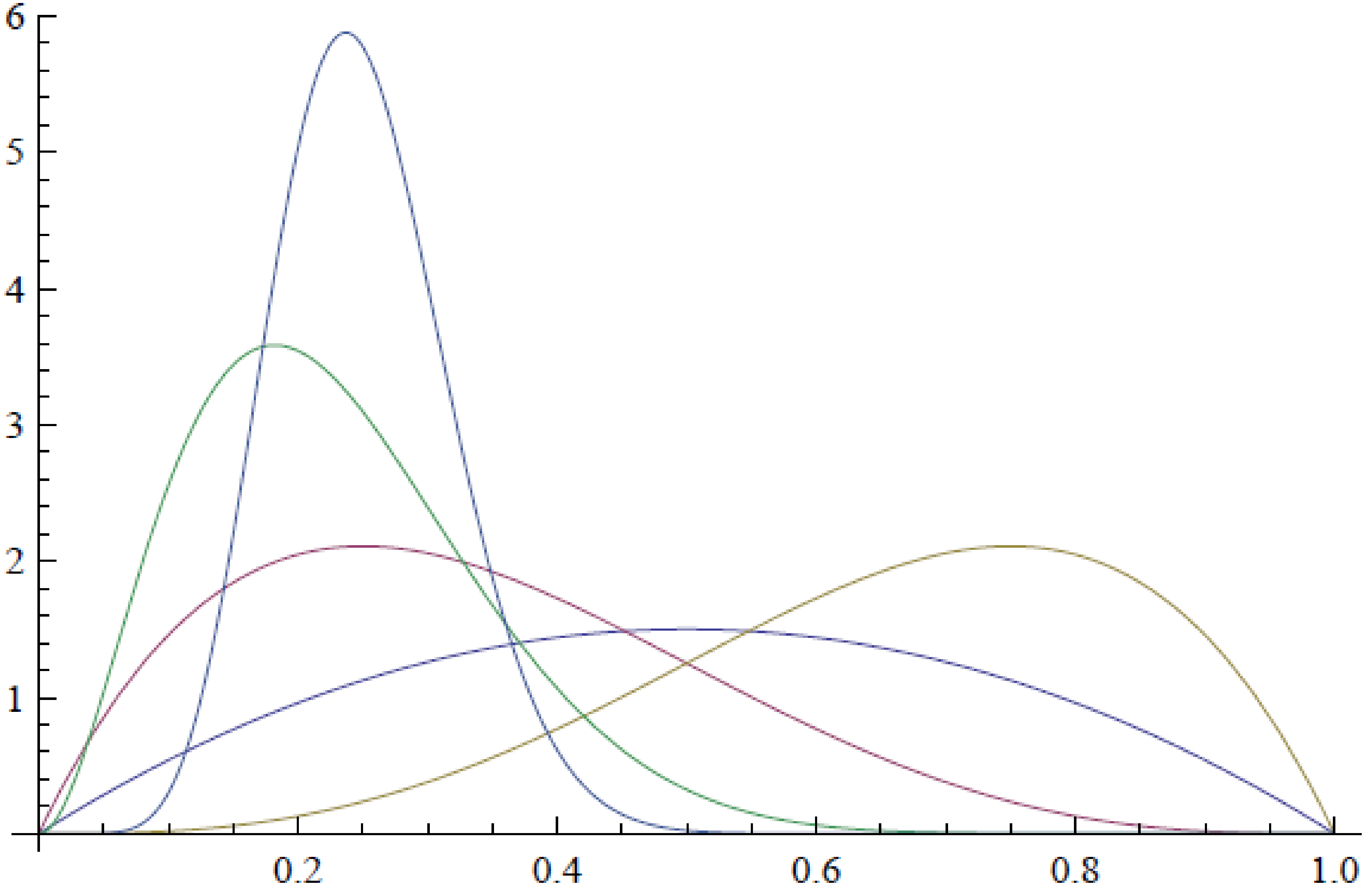


Figure 15.1: Plots of Beta densities for (a, b) equal to (2, 2), (2, 4), (4, 2), (3, 10), and (10, 30).

The Normal Distribution and the Gamma Function

$$\mu_{2m} = \int_{-\infty}^{\infty} x^{2m} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = (2m-1)!!$$

$$\mu_{2m} = \frac{2^m}{\sqrt{\pi}} \Gamma\left(m + \frac{1}{2}\right)$$

$$\mu_0 = \frac{2}{\sqrt{2\pi}} \int_{x=0}^{\infty} e^{-x^2/2} dx$$

$$= \frac{2}{\sqrt{2\pi}} \int_{u=0}^{\infty} e^{-u} 2^{-1/2} u^{-1/2} du$$

$$= \frac{1}{\sqrt{\pi}} \underbrace{\int_{u=0}^{\infty} e^{-u} u^{\frac{1}{2}-1} du}$$

$$= \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right)$$

Build intuition: take $m=0$

$$\mu_0 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

$u = x^2/2$

$x = (2u)^{1/2}$

$du = x\,dx$

$dx = x^{-1} du$

$dx = (2u)^{-1/2} du$

$$\Gamma(s) := \int_0^{\infty} e^{-x} x^{s-1} dx$$

Generalize: have x^{2m} incorporated

$x^{2m} = (2u)^m = 2^m u^m$

Yield $\mu_{2m} = \frac{2^m}{\sqrt{\pi}} \Gamma(m+\frac{1}{2})$

The Gamma and Weibull Distributions. A random variable X has the **Gamma distribution** with (positive) parameters k and σ if its density is

$$f_{k,\sigma}(x) = \begin{cases} \frac{1}{\Gamma(k)\sigma^k} x^{k-1} e^{-x/\sigma} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

We call k the **shape** parameter and σ the **scale** parameter, and write $X \sim \Gamma(k, \sigma)$ or $X \sim \text{Gamma}(k, \sigma)$.

A random variable X has the **Weibull** distribution with (positive) parameters k and σ if its density is

$u = (x/\sigma)^k$

$$f_{k,\sigma}(x) = \begin{cases} (k/\sigma)(x/\sigma)^{k-1} e^{-(x/\sigma)^k} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

$du = \frac{1}{\sigma}\left(\frac{x}{\sigma}\right)^{k-1} dx$

We call k the **shape** parameter and σ the **scale** parameter, and write $X \sim W(k, \sigma)$.

Chi-square distribution: If X is a chi-square distribution with $\nu \geq \cancel{0}\,1$ degrees of freedom, then X has density

$$f(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2-1)} e^{-x/2} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

We write $X \sim \chi^2(\nu)$ to denote this.

$$\Gamma(s) := \int_0^\infty e^{-x} x^{s-1} dx$$

$$\int_0^\infty x^{\frac{\nu}{2}-1} e^{-x/2} dx \qquad \frac{x}{2} = u \quad \text{so} \quad dx = 2du$$

after alg got closed normalization constant

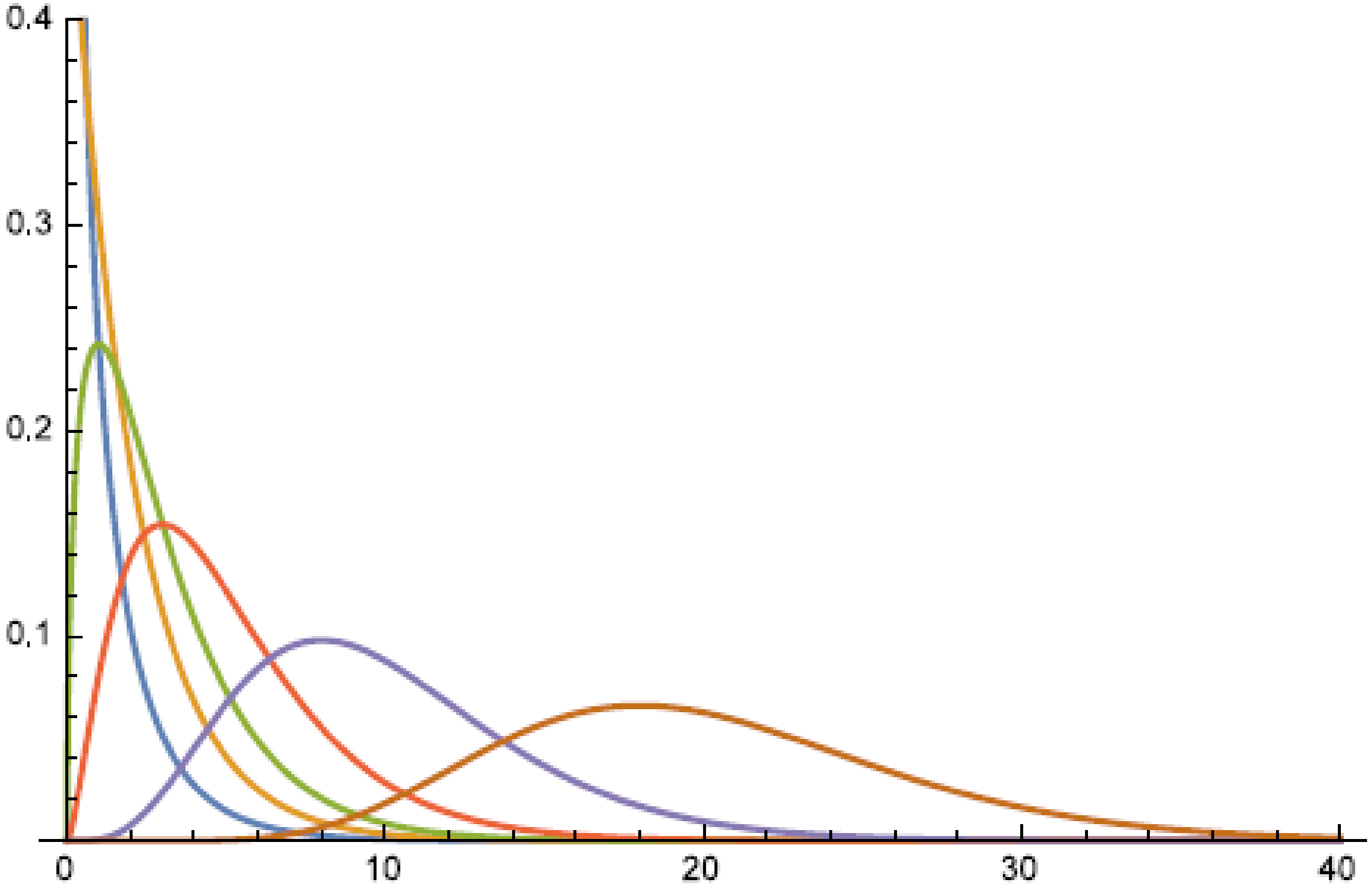


Figure 16.1: Plot of chi-square distributions with $\nu \in \{1, 2, 3, 5, 10, 20\}$; as the degree of freedom increases, the location of the bump moves rightward.

Relation between Chi-square and Normal Random Variables. If $X \sim N(0,1)$ then $X^2 \sim \chi^2(1)$.

$$f(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2-1)} e^{-x/2} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

CDF technique

$Y = X^2$, X is $N(0,1)$

$$\text{Prob}(Y \le y) = \text{Prob}(X^2 \le y) = \text{Prob}(-\sqrt{y} \le X \le \sqrt{y})$$

$$= 2\,\text{Prob}(0 \le X \le \sqrt{y})$$

$$F_Y(y) = 2\int_0^{\sqrt{y}} f_X(x)\,dx = 2F_X(\sqrt{y}) - 2F_X(0)$$

$$f_Y(y) = 2F_X'(\sqrt{y})(\sqrt{y})' - 0$$

$$= 2 f_X(\sqrt{y}) \tfrac{1}{2} y^{-1/2}$$

$$= y^{-1/2} \frac{1}{\sqrt{2\pi}} e^{-y/2}$$

$$= \frac{1}{2^{1/2}\Gamma(\frac{1}{2})} y^{\frac{1}{2}-1} e^{-y/2} \text{ is } \chi^2(1\text{ d.f.}$$

Mean and Variance of $X \sim \chi^2(1)$

$$\mathbb{E}[X] = \int_0^\infty x f(x) dx = \int_0^\infty \frac{x}{2^{(1/2)}\Gamma(1/2)} x^{\frac{1}{2}-1} e^{-x/2} dx.$$

$$= \frac{1}{2^{1/2}\Gamma(1/2)} \int_0^\infty x^{\frac{3}{2}-1} e^{-x/2} dx$$

would be $\Gamma(3/2)$ if had e^{-x} not $e^{-x/2}$

so let $u = x/2$ change variables

Chi-square distribution and sums of normal random variables: Let k be a positive integer, and $X_1, \ldots, X_k$ independent standard normal random variables; this means each $X_i \sim N(0,1)$. Then if $Y_k = X_1^2 + \cdots + X_k^2$, $Y_k \sim \chi^2(k)$. More generally, let $Y_{\nu_1}, \ldots, Y_{\nu_m}$ be m independent chi-square random variables, where $Y_{\nu_i} \sim \chi^2(\nu_i)$. Then $Y = Y_{\nu_1} + \cdots + Y_{\nu_m}$ is a chi-square random variable with $\nu_1 + \cdots + \nu_m$ degrees of freedom.

If $Y_{\nu_1} \sim \chi^2(\nu_1)$ and $Y_{\nu_2} \sim \chi^2(\nu_2)$ are two independent, chi-square random variables, then $Y_{\nu_1} + Y_{\nu_2} \sim \chi^2(\nu_1 + \nu_2)$.

$$((Y_1 + Y_2) + Y_3) + Y_4$$

$$\underbrace{\chi^2(\nu_1+\nu_2) + \chi^2(\nu_3)}_{\chi^2(\nu_1+\nu_2+\nu_3)}$$

(induction)

Grouping ( )

Change of Variables Theorem: Let V and W be bounded open sets in $\mathbb{R}^k$. Let $h : V \to W$ be a 1-1 and onto map, given by

$$h(u_1, \ldots, u_k) = (h_1(u_1, \ldots, u_k), \ldots, h_k(u_1, \ldots, u_k)).$$

Let $f : W \to \mathbb{R}$ be a continuous, bounded function. Then

$$\begin{aligned} &\int \cdots \int_W f(x_1, \ldots, x_k) dx_1 \cdots dx_k \\ &\quad = \int \cdots \int_V f(h(u_1, \ldots, u_k)) J(u_1, \ldots, u_v) du_1 \cdots du_k, \end{aligned}$$

where J is the Jacobian

$$J = \begin{vmatrix} \frac{\partial h_1}{\partial u_1} & \cdots & \frac{\partial h_1}{\partial u_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_k}{\partial u_1} & \cdots & \frac{\partial h_k}{\partial u_k} \end{vmatrix}.$$

Letting the densities be f_{ν_1} and f_{ν_2}, the density of $Y = Y_{\nu_1} + Y_{\nu_2}$ is

$$\begin{aligned} f_Y(y) = (f_{\nu_1} * \cdots * f_{\nu_2})(y) &= \int_{-\infty}^{\infty} f_{\nu_1}(t) f_{\nu_2}(y-t)\,dt \\ &= \int_0^y c_{\nu_1} t^{\frac{\nu_1}{2}-1} e^{-t/2} \cdot c_{\nu_2}(y-t)^{\frac{\nu_2}{2}-1} e^{-(y-t)/2}\,dt. \end{aligned}$$

Chi-square distribution: If X is a chi-square distribution with $\nu \geq 0$ degrees of freedom, then X has density

$$f(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2-1)} e^{-x/2} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

We write $X \sim \chi^2(\nu)$ to denote this.

$$= c_{\nu_1} c_{\nu_2} e^{-\frac{y}{2}} \int_0^y t^{\frac{\nu_1}{2}-1} (y-t)^{\frac{\nu_2}{2}-1}\,dt$$

$t = yu \quad t: 0 \to y$

$dt = y\,du \quad u: 0 \to 1$

$$= c_{\nu_1} c_{\nu_2} e^{-y/2} \int_0^1 y^{\frac{\nu_1}{2}-1} u^{\frac{\nu_1}{2}-1} y^{\frac{\nu_2}{2}-1} (1-u)^{\frac{\nu_2}{2}-1} y\,du$$

$$= c_{\nu_1} c_{\nu_2} e^{-y/2} y^{\frac{\nu_1+\nu_2}{2}-1} \left(\int_0^1 u^{\frac{\nu_1}{2}-1} (1-u)^{\frac{\nu_2}{2}-1}\,du \right)$$

Something ! $\gamma(\nu_1,\nu_2)$

indep of y

$$= \underbrace{c_{\nu_1} c_{\nu_2} \gamma(\nu_1,\nu_2)}_{c_{\nu_1+\nu_2}} \underbrace{e^{-y/2} y^{\frac{\nu_1+\nu_2}{2}-1}}_{\text{functional form of a } \chi^2 \text{ random var with } \nu_1+\nu_2 \text{ d.f.}}$$

Sums of squares by the Change of Variables Theorem

We now return to our problem. Let $Y = X_1^2 + \cdots + X_k^2$. We again use the cumulative distribution function technique and find

$$\begin{aligned} F_Y(y) &= \text{Prob}(X_1^2 + \cdots + X_k^2 \le y) \\ &= \int \cdots \int_{x_1^2+\cdots+x_k^2 \le y} \frac{1}{\sqrt{2\pi}}\, e^{-x_1^2/2} \cdots \frac{1}{\sqrt{2\pi}}\, e^{-x_k^2/2} dx_1 \cdots dx_k \\ &= \int \cdots \int_{x_1^2+\cdots+x_k^2 \le y} \frac{1}{(2\pi)^{k/2}}\, e^{-(x_1^2+\cdots+x_k^2)/2} dx_1 \cdots dx_k. \end{aligned}$$

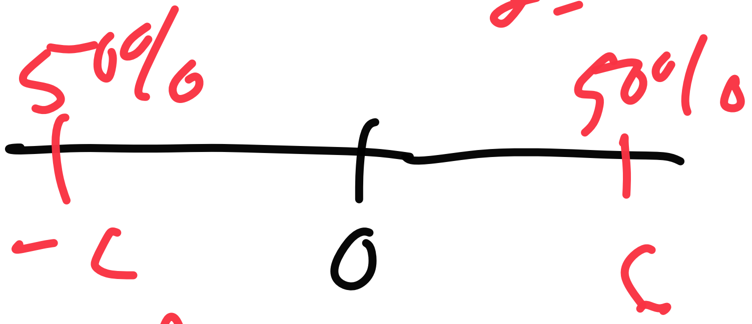
Markov's inequality. Let X be a non-negative random variable with finite mean $\mathbb{E}[X]$ (this means $\text{Prob}(X < 0) = 0$). Then for any positive a we have

$$\text{Prob}(X \geq a) \;\leq\; \frac{\mathbb{E}[X]}{a}.$$

Some authors write μ_X for $\mathbb{E}[X]$. An alternative formulation is

$$\text{Prob}(X < a) \;\geq\; 1 - \frac{\mathbb{E}[X]}{a}.$$

Need non-neg!

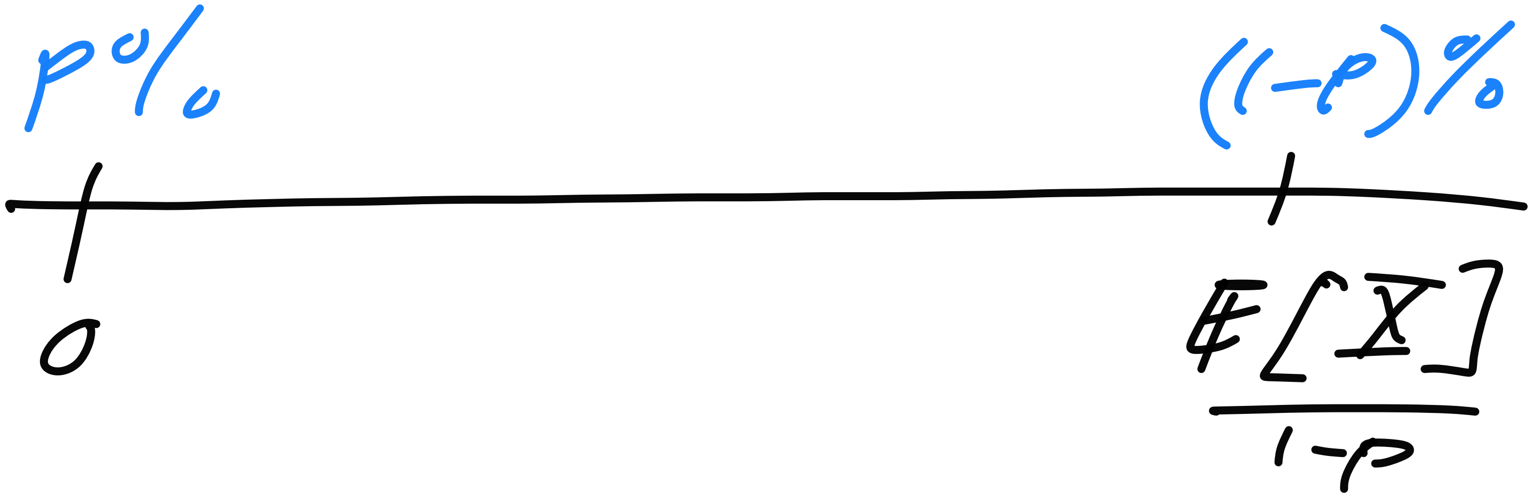


$E[X] = 0$

$\text{Prob}(X \geq a) \leq 0$

Markov's inequality: Sanity Checks:

- Units
- Choices of a
- Special cases / Extreme cases



This has expected value $E[X]$

Markov's inequality. Let X be a non-negative random variable with finite mean $\mathbb{E}[X]$ (this means $\text{Prob}(X < 0) = 0$). Then for any positive a we have

$$\text{Prob}(X \geq a) \;\leq\; \frac{\mathbb{E}[X]}{a}.$$

Some authors write μ_X for $\mathbb{E}[X]$. An alternative formulation is

$$\text{Prob}(X < a) \;\geq\; 1 - \frac{\mathbb{E}[X]}{a}.$$

If $X \geq a$

then $X/a \geq 1$

use nonneg. when throw away integral over some region

Proof:

$$\text{Prob}(X \geq a) = \int_{x=a}^{\infty} f_X(x)\,dx$$

$$\frac{\mathbb{E}[X]}{a} = \int_0^{\infty} \frac{x}{a} f_X(x)\,dx \geq \int_{x=a}^{\infty} \frac{x}{a} f_X(x)\,dx$$

in this region $\frac{x}{a} \geq 1$

$$\geq \int_{x=a}^{\infty} 1\, f_X(x)\,dx$$

$$= \text{Prob}(X \geq a)$$

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 17: 4/17/25: Markov and Chebyshev's inequalities, Divide and Conquer vs Newton's Method, Exponential Function, Stirling's Formula, Dyadic Decomposition, Poisson Random Variables, CLT to Stirling:
Video: https://youtu.be/e17FtuD-qJk

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 16: April 15, 2021:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- Markov’s Inequality
- Chebyshev’s Inequality
- Divide and Conquer
- Newton’s Method
- Exponential Function
- Poisson Random Variables, Stirling’s Formula

General items.

- The more you assume, the more you can deduce…
- Estimation

Markov's inequality. Let X be a non-negative random variable with finite mean $\mathbb{E}[X]$ (this means $\text{Prob}(X < 0) = 0$). Then for any positive a we have

$$\text{Prob}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Some authors write μ_X for $\mathbb{E}[X]$. An alternative formulation is

$$\text{Prob}(X < a) \geq 1 - \frac{\mathbb{E}[X]}{a}.$$

If $X \geq a$

then $X/a \geq 1$

use remaining when

throw away integral

over some region

Proof:

look at $\text{Prob}(X \geq a) = \int_a^\infty f_X(x)\,dx$

use $\frac{x}{a} \geq 1$

$$\frac{\mathbb{E}[X]}{a} = \int_0^\infty \frac{x}{a} f_X(x)\,dx \geq \int_a^\infty \frac{x}{a} f_X(x)\,dx \geq \int_a^\infty f_X(x)\,dx$$

Now that we've seen a proof, let's do an example. *Imagine the mean US income is \$60,000. What's the probability a household chosen at random has an income of at least \$120,000? Of at least \$1,000,000?*

As stated, we don't have enough information to solve this problem. Maybe there's a few very rich people and everyone else earns essentially nothing. Or, the opposite extreme, maybe everyone makes close to the average. Without knowing more about how incomes are distributed, we can't get an exact answer. We can, however, get some bounds on the answer by using Markov's inequality. To use this, we need a non-negative random variable with finite mean. If we assume that no household has a negative income then we're fine, as the other condition is met (the mean is \$60,000, which is finite).

Thus the probability of an income of at least \$120,000 is at most $60000/120000 = 1/2$; or, at most half the population makes twice the mean. What about the millionaire's club? The probability of being a millionaire is at most $60000/1000000 = .06$, or at most 6% of the households.

Let X be a non-negative random variable with finite mean $\mathbb{E}[X]$. Then the probability of being at least ℓ times the mean is at most $1/\ell$:

$$\text{Prob}(X \geq \ell\mathbb{E}[X]) \ \leq \ \frac{1}{\ell}.$$

Unfortunately this is the best we can do with our limited information. So long as our random variable has finite mean and is non-negative, the probability of being 100 or more times the mean is at most $1/100$ or 1%. Of course, in many problems the true probability is *magnitudes* less than this. This is an excessively high over-estimate at times. This suggests, of course, the next step: incorporate more information and get a better bound! We do this in the next section.

Theorem 17.3.1 (Chebyshev's Inequality) *Let X be a random variable with finite mean μ_X and finite variance σ_X^2. Then for any $k > 0$ we have*

$$\text{Prob}(|X - \mu_X| \geq k\sigma_X) \ \leq \ \frac{1}{k^2}.$$

Some authors write $\mathbb{E}[X]$ for μ_X. This means that the probability of obtaining a value at least k standard deviations from the mean is at most $1/k^2$. A useful, alternative formulation is

$$\text{Prob}(|X - \mu_X| < k\sigma_X) \ > \ 1 - \frac{1}{k^2}.$$

Useless if $k \leq 1$

Uses more info than Markov. Use stdev

Chebyshev's inequality: Sanity Checks:

- Units
- Choices of k
- Special cases / Extreme cases
- Better than Markov for large deviations (reciprocal of quadratic vs linear)

Theorem 17.3.1 (Chebyshev's Inequality) *Let X be a random variable with finite mean μ_X and finite variance σ_X^2. Then for any $k > 0$ we have*

$$\mathrm{Prob}(|X - \mu_X| \geq k\sigma_X) \;\leq\; \frac{1}{k^2}.$$

Some authors write $\mathbb{E}[X]$ for μ_X. This means that the probability of obtaining a value at least k standard deviations from the mean is at most $1/k^2$. A useful, alternative formulation is

$$\mathrm{Prob}(|X - \mu_X| < k\sigma_X) \;>\; 1 - \frac{1}{k^2}.$$

$Y = |X - E[X]|$

non-neg

$Var(X) = E[X^2] - E[X]^2$

$= \int_{-\infty}^{\infty} \underbrace{(x - E[X])^2 f_X(x)}_{\text{non-neg}} dx$

Chebyshev's Inequality: Proof from Markov:

$W = (X - E[X])^2 \geq 0$

$E[W] = Var(X) = \sigma_X^2$

Markov: $\mathrm{Prob}(W \geq a) \leq \frac{E[W]}{a}$

$\mathrm{Prob}(|X - \mu_X| \geq a^{\frac{1}{2}}) \leq \frac{\sigma_X^2}{a}$

take square-root

want $a^{1/2} = k\sigma_X$

so $\sigma_X^2/a = 1/k^2$

substitute!

Theorem 17.3.1 (Chebyshev's Inequality) *Let X be a random variable with finite mean μ_X and finite variance σ_X^2. Then for any $k > 0$ we have*

$$\text{Prob}(|X - \mu_X| \geq k\sigma_X) \;\leq\; \frac{1}{k^2}.$$

Direct proof of Chebyshev's inequality. Let f_X be the probability density function of X. We assume X is a continuous random variable, though a similar proof holds in the discrete case. We have

$$\begin{aligned}
\text{Prob}(|X - \mu_X| \geq k\sigma_X) &= \int_{x:|x-\mu_X| \geq k\sigma_X} 1 \; f_X(x)dx \\
&\leq \int_{x:|x-\mu_X| \geq k\sigma_X} \left(\frac{x - \mu_X}{k\sigma_X}\right)^2 \cdot f_X(x)dx \\
&= \frac{1}{k^2\sigma_X^2} \int_{x:|x-\mu_X| \geq k\sigma_X} (x - \mu_X)^2 f_X(x)dx \\
&\leq \frac{1}{k^2\sigma_X^2} \int_{x=-\infty}^{\infty} (x - \mu_X)^2 f_X(x)dx \\
&= \frac{1}{k^2\sigma_X^2} \cdot \sigma_X^2 \;=\; \frac{1}{k^2},
\end{aligned}$$

completing the proof. $\square$

From $\mathbb{C}$ to Shining Sea: $\mathbb{C}$omplex Dynamics from $\mathbb{C}$ombinatorics to $\mathbb{C}$oastlines

Steven J. Miller, Williams $\mathbb{C}$ollege
sjm1@williams.edu
http://web.williams.edu/Mathematics/sjmiller/public_html/

Michigan Math $\mathbb{C}$lub, April 30, 2020

https://web.williams.edu/Mathematics/sjmiller/public_html/math/talks/CToShiningSeaMichigan2020.pdf
https://youtu.be/TMILk79N_Bs

Finding roots

Much of math is about solving equations.

Example: polynomials:

- $ax + b = 0$, root $x = -b/a$.
- $ax^2 + bx + c = 0$, roots $(-b \pm \sqrt{b^2 - 4ac})/2a$.
- Cubic, quartic: formulas exist in terms of coefficients; not for quintic and higher.

In general cannot find exact solution, how to estimate?

Cubic: For fun, here's the solution to $ax^3 + bx^2 + cx + d = 0$

```
Solve[a x^3 + b x^2 + c x + d == 0, x]
```

$$\left\{\left\{x \to -\frac{b}{3a} - \frac{2^{1/3}\left(-b^2+3ac\right)}{3a\left(-2b^3+9abc-27a^2d+\sqrt{4\left(-b^2+3ac\right)^3+\left(-2b^3+9abc-27a^2d\right)^2}\right)^{1/3}} + \frac{\left(-2b^3+9abc-27a^2d+\sqrt{4\left(-b^2+3ac\right)^3+\left(-2b^3+9abc-27a^2d\right)^2}\right)^{1/3}}{3\times 2^{1/3}a}\right\},\right.$$

$$\left\{x \to -\frac{b}{3a} + \frac{\left(1+i\sqrt{3}\right)\left(-b^2+3ac\right)}{3\times 2^{2/3}a\left(-2b^3+9abc-27a^2d+\sqrt{4\left(-b^2+3ac\right)^3+\left(-2b^3+9abc-27a^2d\right)^2}\right)^{1/3}} - \frac{\left(1-i\sqrt{3}\right)\left(-2b^3+9abc-27a^2d+\sqrt{4\left(-b^2+3ac\right)^3+\left(-2b^3+9abc-27a^2d\right)^2}\right)^{1/3}}{6\times 2^{1/3}a}\right\},$$

$$\left.\left\{x \to -\frac{b}{3a} + \frac{\left(1-i\sqrt{3}\right)\left(-b^2+3ac\right)}{3\times 2^{2/3}a\left(-2b^3+9abc-27a^2d+\sqrt{4\left(-b^2+3ac\right)^3+\left(-2b^3+9abc-27a^2d\right)^2}\right)^{1/3}} - \frac{\left(1+i\sqrt{3}\right)\left(-2b^3+9abc-27a^2d+\sqrt{4\left(-b^2+3ac\right)^3+\left(-2b^3+9abc-27a^2d\right)^2}\right)^{1/3}}{6\times 2^{1/3}a}\right\}\right\}$$

wlog
$a=1$ if non-zero
wlog $b=0$
$x \Rightarrow x-r$

One of four solutions to quartic $ax^4 + bx^3 + cx^2 + dx + e = 0$

```
Solve[a x^4 + b x^3 + c x^2 + d x + e == 0, x]
```

$$
\begin{aligned}
\Biggl\{\Biggl\{x \to & -\frac{b}{4\,a} - \frac{1}{2}\sqrt{\left(\frac{b^2}{4\,a^2} - \frac{2\,c}{3\,a} + \right.}\\
& \left(2^{1/3}\left(c^2 - 3\,b\,d + 12\,a\,e\right)\right) \Big/ \left(3\,a\left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e + \sqrt{-4\left(c^2 - 3\,b\,d + 12\,a\,e\right)^3 + \left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e\right)^2}\right)^{1/3}\right) +\\
& \frac{1}{3 \times 2^{1/3}\,a}\left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e + \sqrt{-4\left(c^2 - 3\,b\,d + 12\,a\,e\right)^3 + \left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e\right)^2}\right)^{1/3}\Biggr) - \frac{1}{2}\sqrt{\left(\frac{b^2}{2\,a^2} - \frac{4\,c}{3\,a} - \right.}\\
& \left(2^{1/3}\left(c^2 - 3\,b\,d + 12\,a\,e\right)\right) \Big/ \left(3\,a\left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e + \sqrt{-4\left(c^2 - 3\,b\,d + 12\,a\,e\right)^3 + \left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e\right)^2}\right)^{1/3}\right) -\\
& \frac{1}{3 \times 2^{1/3}\,a}\left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e + \sqrt{-4\left(c^2 - 3\,b\,d + 12\,a\,e\right)^3 + \left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e\right)^2}\right)^{1/3} -\\
& \left(-\frac{b^3}{a^3} + \frac{4\,b\,c}{a^2} - \frac{8\,d}{a}\right) \Big/ \left(4\sqrt{\left(\frac{b^2}{4\,a^2} - \frac{2\,c}{3\,a} + \left(2^{1/3}\left(c^2 - 3\,b\,d + 12\,a\,e\right)\right)\right.} \Big/\right.\\
& \quad \left(3\,a\left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e + \sqrt{-4\left(c^2 - 3\,b\,d + 12\,a\,e\right)^3 + \left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e\right)^2}\right)^{1/3}\right) + \frac{1}{3 \times 2^{1/3}\,a}\\
& \quad \left.\left.\left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e + \sqrt{-4\left(c^2 - 3\,b\,d + 12\,a\,e\right)^3 + \left(2\,c^3 - 9\,b\,c\,d + 27\,a\,d^2 + 27\,b^2\,e - 72\,a\,c\,e\right)^2}\right)^{1/3}\right)\right)\Biggr)\Biggr\},
\end{aligned}
$$

Divide and Conquer

Divide and Conquer

Divide and Conquer

Assume f is continuous, $f(a) < 0 < f(b)$. Then f has a root between a and b. To find, look at the sign of f at the midpoint $f\left(\frac{a+b}{2}\right)$; if sign positive look in $[a, \frac{a+b}{2}]$ and if negative look in $[\frac{a+b}{2}, b]$. Lather, rinse, repeat.

Example:

- $f(0) = -1$, $f(1) = 3$, look at $f(.5)$.
- $f(.5) = 2$, so look at $f(.25)$.
- $f(.25) = -.4$, so look at $f(.375)$.

$2^{10} = 1024$

every 10 iterations $\approx \frac{1}{1000}$

or 3 digits

Divide and Conquer (continued)

How fast? Every 10 iterations uncertainty decreases by a factor of $2^{10} = 1024 \approx 1000$.

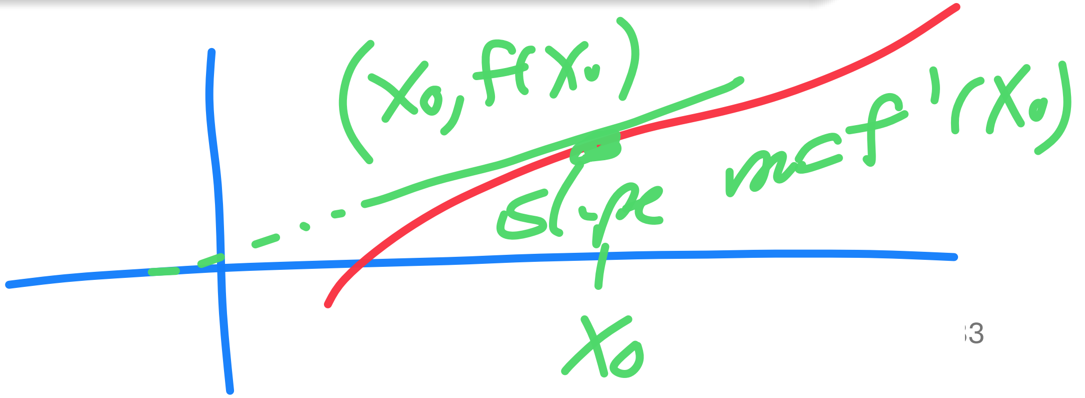
Thus 10 iterations essentially give three decimal digits.

		f(x) = x^2 - 3, sqrt(3)		1.732051		
n	left	right	f(left)	f(right)	left error	right error
1	1	2	-2	1	0.732051	-0.26795
2	1.5	2	-0.75	1	0.232051	-0.26795
3	1.5	1.75	-0.75	0.0625	0.232051	-0.01795
4	1.625	1.75	-0.35938	0.0625	0.107051	-0.01795
5	1.6875	1.75	-0.15234	0.0625	0.044551	-0.01795
6	1.71875	1.75	-0.0459	0.0625	0.013301	-0.01795
7	1.71875	1.734375	-0.0459	0.008057	0.013301	-0.00232
8	1.726563	1.734375	-0.01898	0.008057	0.005488	-0.00232
9	1.730469	1.734375	-0.00548	0.008057	0.001582	-0.00232
10	1.730469	1.732422	-0.00548	0.001286	0.001582	-0.00037

Figure: Approximating $\sqrt{3} \approx 1.7320508075688772935274463415$.

Newton's Method

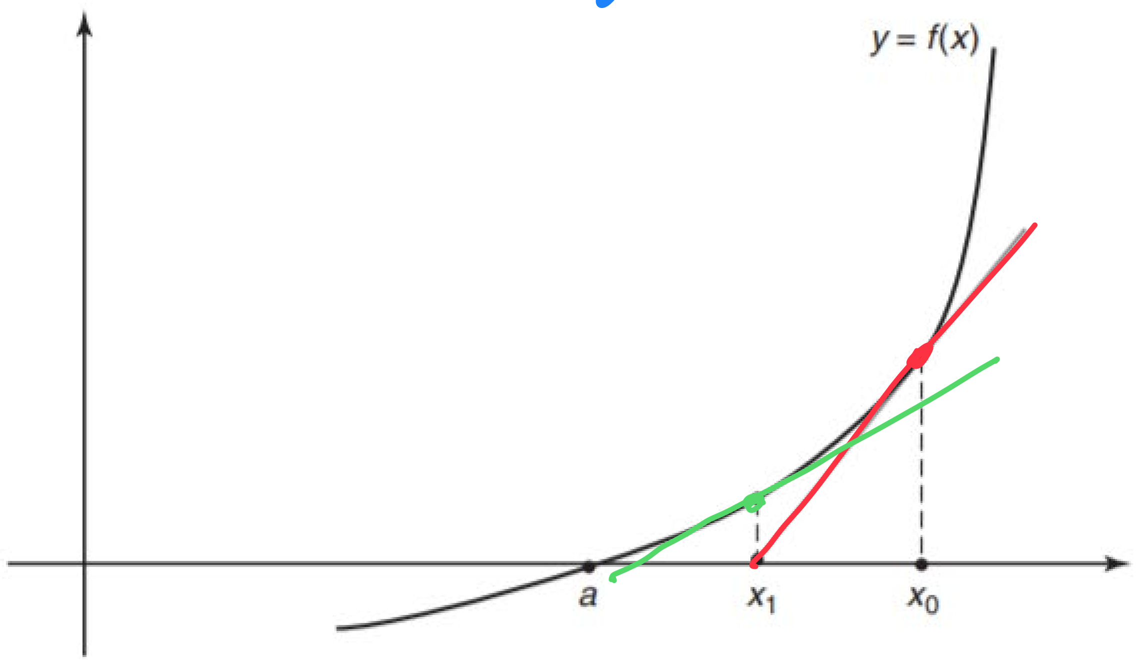
Assume f is continuous and differentiable. We generate a sequence hopefully converging to the root of $f(x) = 0$ as follows. Given x_n, look at the tangent line to the curve $y = f(x)$ at x_n; it has slope $f'(x_n)$ and goes through $(x_n, f(x_n))$ and gives line $y - f(x_n) = f'(x_n)(x - x_n)$. This hits the x-axis at $y = 0, x = x_{n+1}$, and yields $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$.



Newton's Method

$y - f(x_0) = f'(x_0)(x - x_0)$

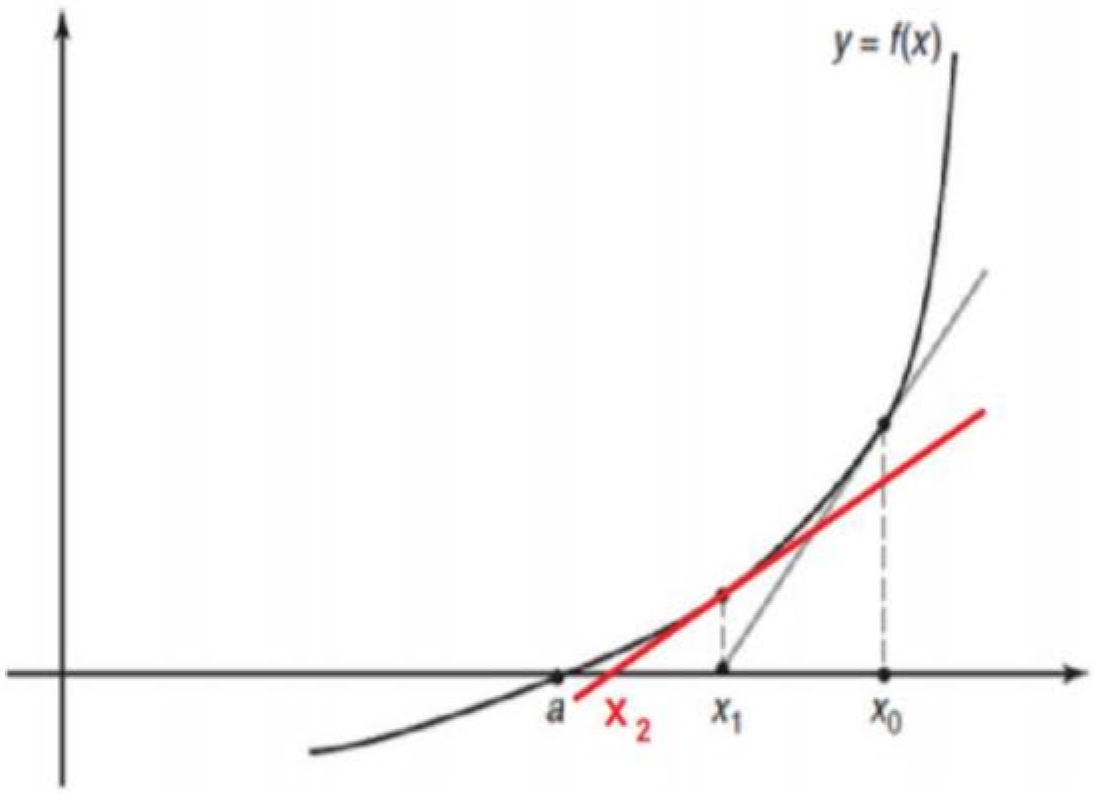
when $y = 0$ call that x_1



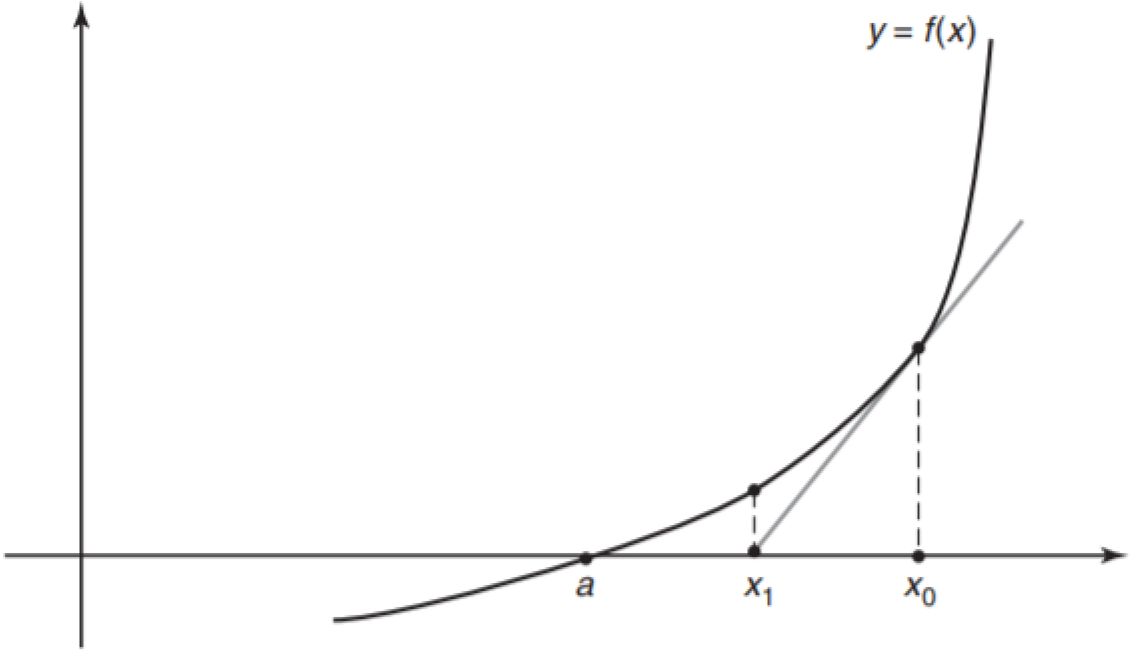
$0 = f(x_0) + f'(x_0)(x_1 - x_0)$

Solve for x_1

Newton's Method



Newton's Method



For example, $f(x) = x^2 - 3$ after algebra get
$x_{n+1} = \frac{1}{2}\left(x_n + \frac{3}{x_n}\right)$.

$x_1 = \sqrt{3}$

$x_{n+1} = \sqrt{3}$

$x_0 = 100$

$x_1 \approx 50$

$x_2 \approx 25$

$x_3 = 12.5$

$x_4 = 6.25$

$x_3 = \vdots$

Newton's Method

n	x[n]	1.0 x[n]	Sqrt[3] - x[n]
0	2	2.0000000000000000000000000000000000000000	-0.267949192431122706472553658494127633057
1	$\frac{7}{4}$	1.7500000000000000000000000000000000000000	-0.017949192431122706472553658494127633057
2	$\frac{97}{56}$	1.7321428571428572062984585500008945167065	-0.000092049573979849329696515636984775914
3	$\frac{18817}{10864}$	1.7320508100147276042690691610914655029774	$-2.4458502469732900355191644519 08 \times 10^{-9}$

Sqrt[3] = 1.7320508075688772935274463415058723669428

x[5] = 1.7320508075688772935274463415058723678037

x[4] = 1.7320508075688772952543539460721719142351

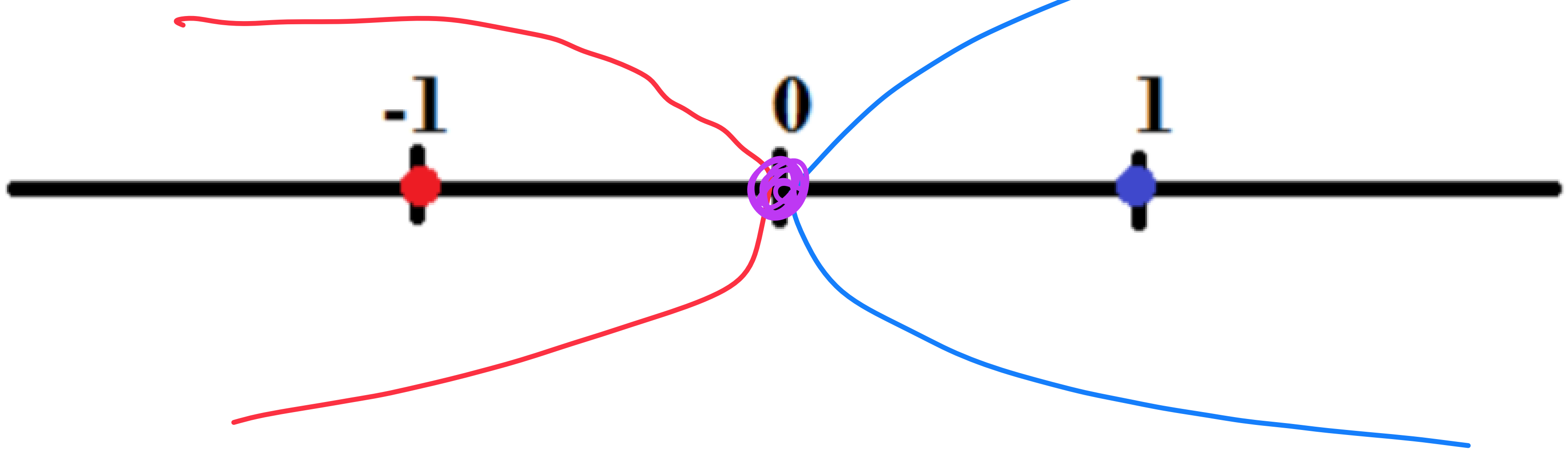
$$\sqrt{3} = 1.7320508075688772935274463415058723\mathbf{6}69428$$
$$x_5 = 1.7320508075688772935274463415058723\mathbf{6}78037$$
$$x_5 = \frac{10029782734113730 57}{5790697761454023 04}.$$

Newton Method: $x^2 - 3 = 0$

Consider $x^2 - 1 = (x-1)(x+1) = 0$.

Roots are 1, -1; if we start at a point x_0 do we approach a root? If so which?

Recall $x_{n+1} = \frac{1}{2}\left(x_n + \frac{1}{x_n}\right)$.

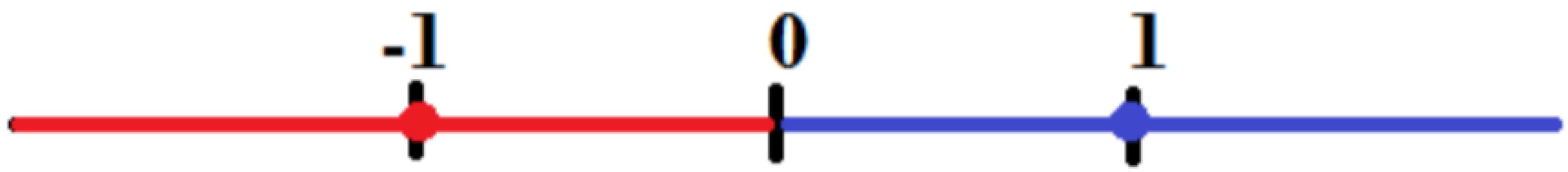


Newton Method: $x^2 - 3 = 0$

Consider $x^2 - 1 = (x-1)(x+1) = 0$.

Roots are 1, -1; if we start at a point x_0 do we approach a root? If so which?

Recall $x_{n+1} = \frac{1}{2}\left(x_n + \frac{1}{x_n}\right)$.



Newton Fractal: $x^3 - 1 = 0$:

https://www.youtube.com/watch?v=ZsFixqGFNRc

What are the roots to $x^3 - 1 = 0$?

Here comes Complex Numbers!

$\mathbb{C} = \{x + iy : x, y \in \mathbb{R}, i = \sqrt{-1}\}$.

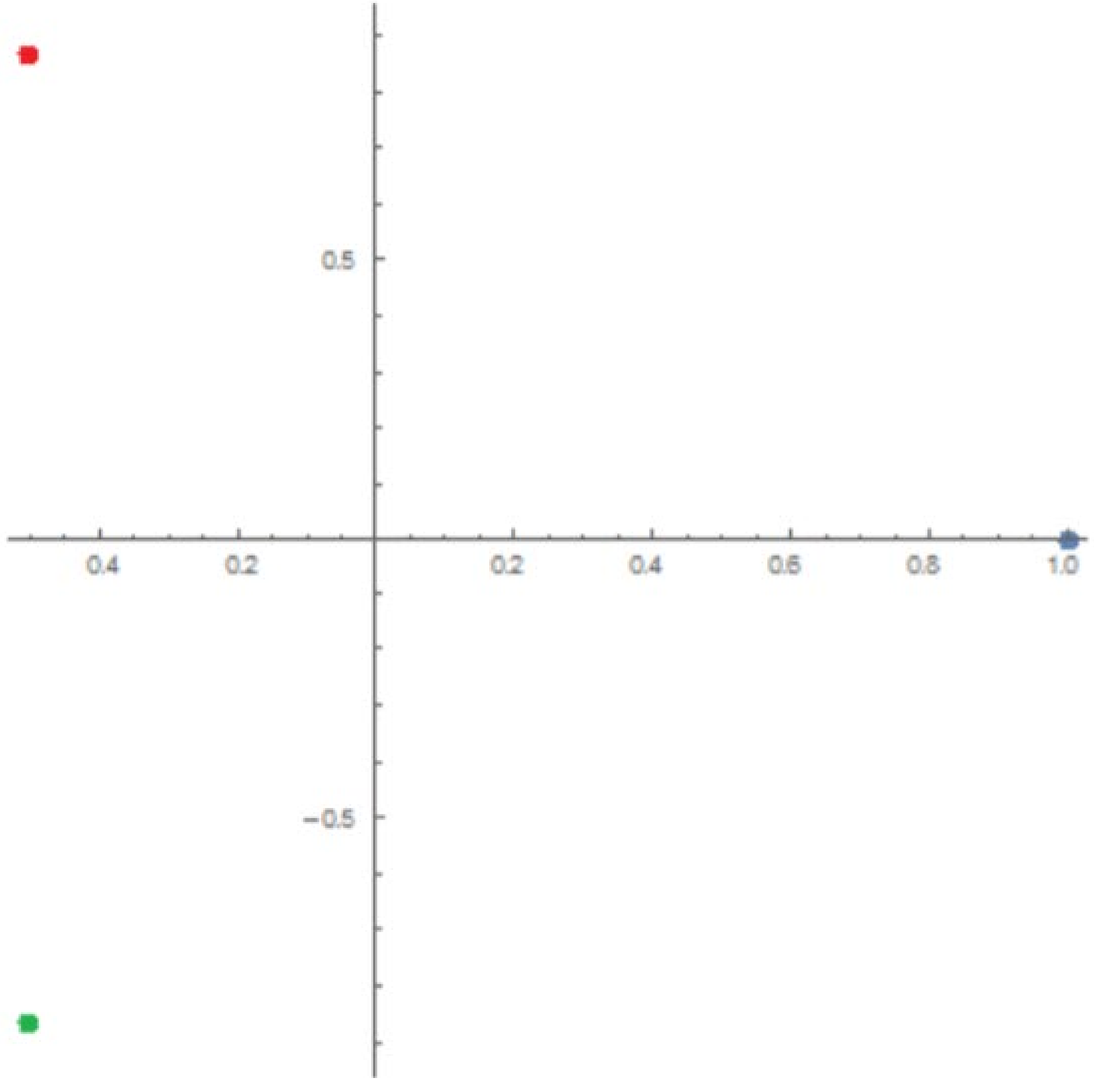
$$
\begin{aligned}
x^3 - 1 &= (x-1)(x^2 + x + 1) \\
&= (x-1) \cdot \left(x - \frac{-1 + \sqrt{1^2 - 4 \cdot 1 \cdot 1}}{2} \right) \cdot \left(x - \frac{-1 - \sqrt{1^2 - 4 \cdot 1 \cdot 1}}{2} \right) \\
&= (x-1) \cdot \left(x - \frac{-1 + \sqrt{-3}}{2} \right) \cdot \left(x - \frac{-1 - \sqrt{-3}}{2} \right) \\
&= (x-1) \cdot \left(x - \frac{-1 + i\sqrt{3}}{2} \right) \cdot \left(x - \frac{-1 - i\sqrt{3}}{2} \right).
\end{aligned}
$$

Roots are 1, $-1/2 + i\sqrt{3}/2$, $-1/2 - i\sqrt{3}/2$.

Newton Fractal: $x^3 - 1 = 0$:

https://www.youtube.com/watch?v=ZsFixqGFNRc

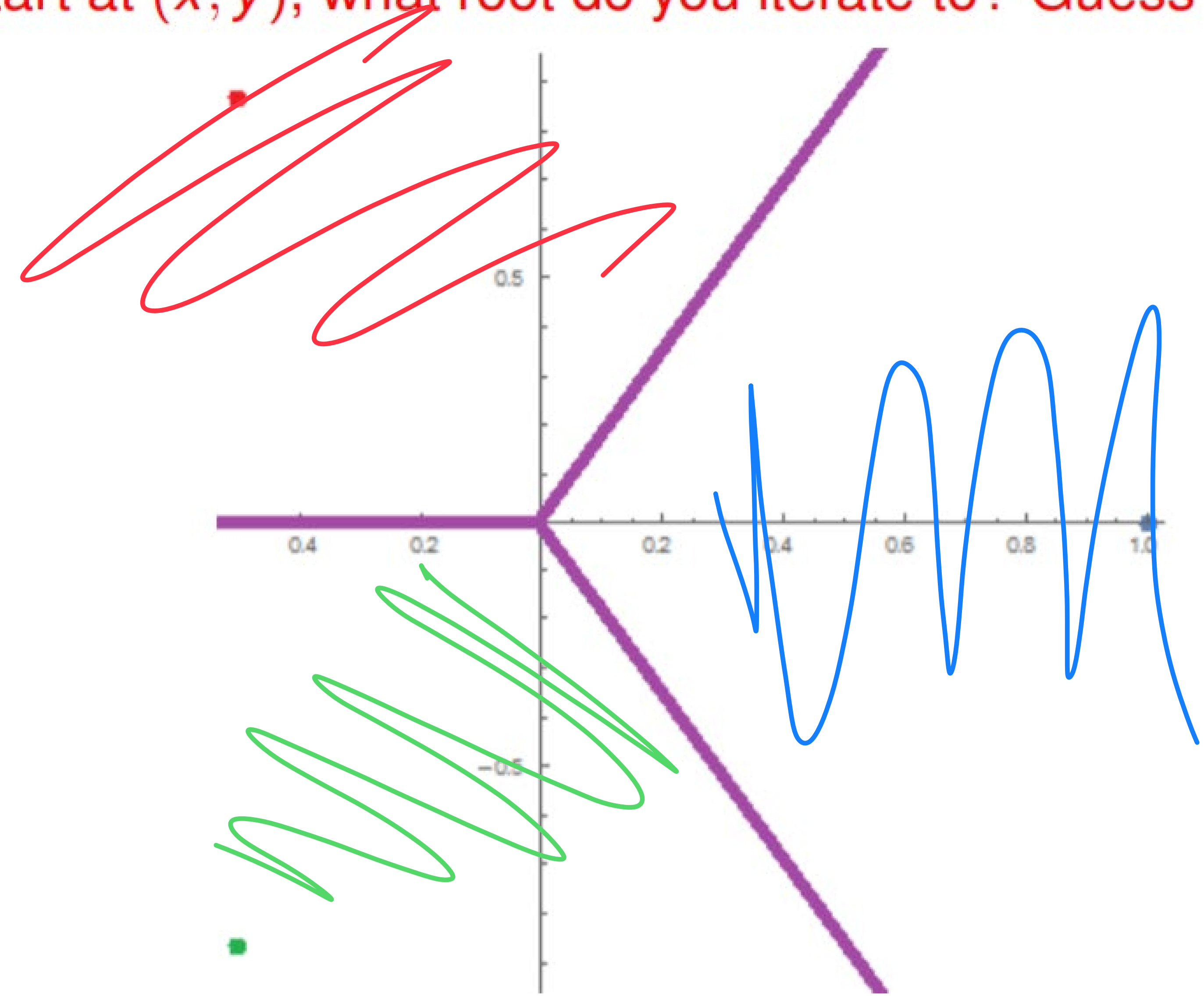
If start at (x, y), what root do you iterate to?



Newton Fractal: $x^3 - 1 = 0$:

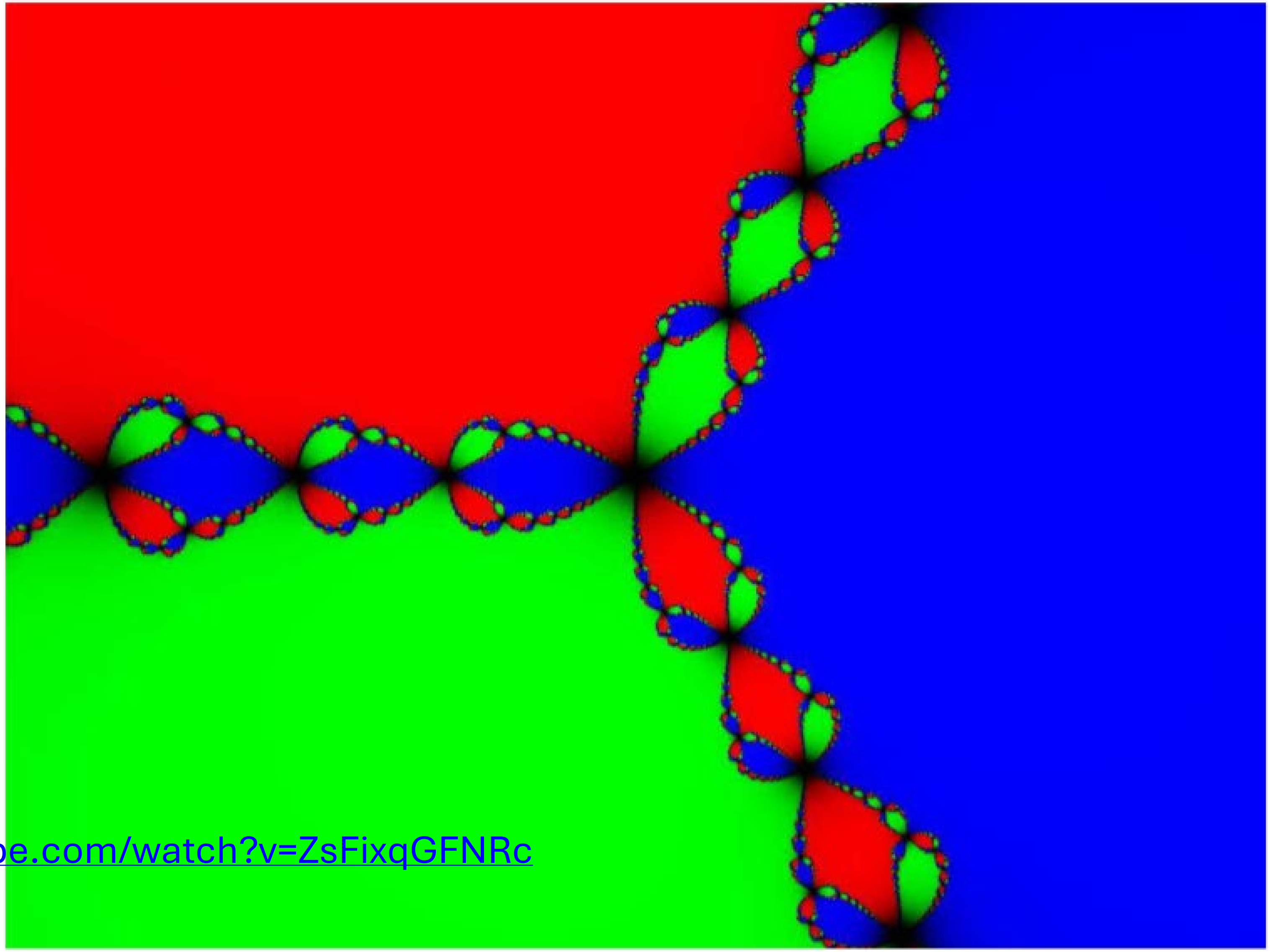
https://www.youtube.com/watch?v=ZsFixqGFNRc

If start at (x, y), what root do you iterate to? Guess



Newton Fractal: $x^3 - 1 = 0$:

https://www.youtube.com/watch?v=ZsFixqGFNRc



https://www.youtube.com/watch?v=ZsFixqGFNRc

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

$m = n-1$

$n: 1 \to \infty \quad m: 0 \to \infty$

$$(e^x)' = \frac{d}{dx} \sum_{n=0}^{\infty} \frac{x^n}{n!} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{d}{dx} x^n = \sum_{n=1}^{\infty} \frac{1}{n!} n x^{n-1} = \sum_{n=1}^{\infty} \frac{1}{(n-1)!} x^{n-1} = e^x$$

$$e^x e^y = \sum_{n=0}^{\infty} \frac{x^n}{n!} \sum_{m=0}^{\infty} \frac{y^m}{m!} = \sum_{k=0}^{\infty} \sum_{l=0}^{k} \frac{x^l y^{k-l}}{l!\,(k-l)!} \cdot \frac{k!}{k!}$$

$$e^{x+y} = \sum_{k=0}^{\infty} \frac{(x+y)^k}{k!}$$

$$= \sum_{k=0}^{\infty} \frac{1}{k!} \left[\sum_{l=0}^{k} \binom{k}{l} x^l y^{k-l} \right]$$

Binomial Thm

$$= \sum_{k=0}^{\infty} \frac{(x+y)^k}{k!}$$

The Gamma function. The Gamma function $\Gamma(s)$ is

$$\Gamma(s) = \int_0^{\infty} e^{-x} x^{s-1} dx, \quad \Re(s) > 0.$$

$\Gamma(n+1)$

$= n!$

n nonneg integer

Stirling's formula: As $n \to \infty$, we have

$$n! \approx n^n e^{-n} \sqrt{2\pi n};$$

by this we mean

$$\lim_{n\to\infty} \frac{n!}{n^n e^{-n}\sqrt{2\pi n}} = 1.$$

More precisely, we have the following series expansion:

$$n! = n^n e^{-n} \sqrt{2\pi n} \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} - \cdots\right).$$

Crude upper/lower bounds for n! $= n(n-1)\cdots 3\cdot 2\cdot 1$

$$0 \le 1 \le n \le n! \le n^n$$

$$\underbrace{1\cdot 2 \cdots \tfrac{n}{2}}_{\text{b/w } 1 \text{ and } \frac{n}{2}} \quad \underbrace{\tfrac{n}{2}+1 \cdots n}_{\text{b/w } \frac{n}{2} \text{ and } n}$$

$$1^{n/2}\cdot\left(\frac{n}{2}\right)^{\frac{n}{2}} \le n! \le \left(\frac{n}{2}\right)^{\frac{n}{2}} n^{\frac{n}{2}}$$

Dyadic Decompositions

$$n! \le \left(\frac{n}{4}\right)^{\frac{n}{4}}\left(\frac{2n}{4}\right)^{\frac{n}{4}}\left(\frac{3n}{4}\right)^{\frac{n}{4}}(n)^{\frac{n}{4}}$$

$$= \left(\frac{1}{4}\cdot\frac{2}{4}\cdot\frac{3}{4}\cdot 1\right)^{\frac{n}{4}}\cdot n^n$$

$$= \left(\frac{3}{16}\right)^n n^n = n^n\left(\frac{16}{3}\right)^{-n}$$

Note (n+1)!/n! = n+1; let's see what Stirling gives:

$$\frac{(n+1)!}{n!} = \frac{(n+1)^{n+1} e^{-(n+1)} \sqrt{2\pi(n+1)}}{n^n e^{-n} \sqrt{2\pi n}}$$

$$\left(\frac{n+1}{n}\right)^n = \left(1+\frac{1}{n}\right)^n$$

$$= (n+1)\left(1+\frac{1}{n}\right)^n e^{-1} \sqrt{\frac{n+1}{n}}$$

$$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$$

$$= (n+1) \underbrace{\underbrace{\left(1+\frac{1}{n}\right)^n}_{\substack{\text{as } n\to\infty \\ \text{goes } e^1}} e^{-1} \underbrace{\sqrt{1+\frac{1}{n}}}_{\substack{\text{as } n\to\infty \\ \text{goes to } 1}}}_{\text{looks like } 1}$$

Integral Test and the Poor Mathematician's Stirling

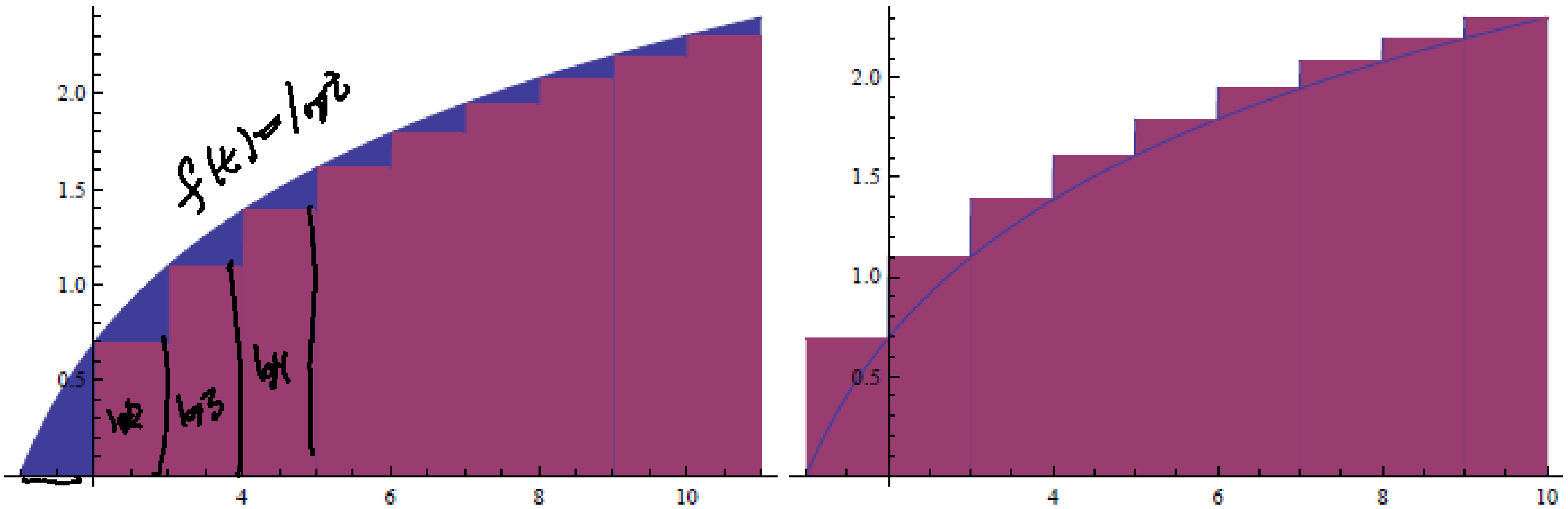
factors are 1 and n

$GM(1,n) = \sqrt{1\cdot n} = \sqrt{n}$

> **Poor man's Stirling.** Let $n \geq 3$ be a positive integer. Then
>
> $$n^n e^{-n} \cdot e \;\leq\; n! \;\leq\; n^n e^{-n} \cdot en.$$

$$\log P \;=\; \log n! \;=\; \log 1 + \log 2 + \cdots + \log n \;=\; \sum_{k=1}^{n} \log k.$$

$$\int_1^n \log t dt \;\leq\; \sum_{k=1}^{n} \log k \;\leq\; \int_2^{n+1} \log t dt.$$



Lower and upper bound for $\log n!$ when $n = 10$.

Stirling's Formula: Lower bound from Integral Test:

Claim $(t\log t - t)' = \log t$

$(t \log t)' = 1 \cdot \log t - t\frac{1}{t}$

$= \log t - 1$

$$
\begin{aligned}
(t\log t - t)\Big|_{t=1}^{n} \ &\le\ \log n! \ \le\ (t\log t - t)\Big|_{t=2}^{n+1} \\
n\log n - n + 1 \ &\le\ \log n! \ \le\ (n+1)\log(n+1) - (n+1) - (2\log 2 - 2).
\end{aligned}
$$

We'll study the lower bound first. From

$$n\log n - n + 1 \ \le\ \log n!,$$

we find after exponentiating that

$$e^{n\log n - n + 1} \ =\ n^n e^{-n}\cdot e \ \le\ n!.$$

Euler–Maclaurin formula

From Wikipedia, the free encyclopedia: https://en.wikipedia.org/wiki/Euler%E2%80%93Maclaurin_formula

If m and n are natural numbers and $f(x)$ is a real or complex valued continuous function for real numbers x in the interval $[m,n]$, then the integral

$$I = \int_m^n f(x)\,dx$$

can be approximated by the sum (or vice versa)

$$S = f(m+1) + \cdots + f(n-1) + f(n)$$

(see rectangle method). The Euler–Maclaurin formula provides expressions for the difference between the sum and the integral in terms of the higher derivatives $f^{(k)}(x)$ evaluated at the endpoints of the interval, that is to say $x = m$ and $x = n$.

Explicitly, for p a positive integer and a function $f(x)$ that is p times continuously differentiable on the interval $[m,n]$, we have

$$S - I = \sum_{k=1}^{p} \frac{B_k}{k!} \left(f^{(k-1)}(n) - f^{(k-1)}(m) \right) + R_p,$$

where B_k is the kth Bernoulli number (with $B_1 = \frac{1}{2}$) and R_p is an error term which depends on n, m, p, and f and is usually small for suitable values of p.

The formula is often written with the subscript taking only even values, since the odd Bernoulli numbers are zero except for B_1. In this case we have[1][2]

$$\sum_{i=m}^{n} f(i) = \int_m^n f(x)\,dx + \frac{f(n)+f(m)}{2} + \sum_{k=1}^{\lfloor \frac{p}{2} \rfloor} \frac{B_{2k}}{(2k)!} \left(f^{(2k-1)}(n) - f^{(2k-1)}(m) \right) + R_p,$$

or alternatively

$$\sum_{i=m+1}^{n} f(i) = \int_m^n f(x)\,dx + \frac{f(n)-f(m)}{2} + \sum_{k=1}^{\lfloor \frac{p}{2} \rfloor} \frac{B_{2k}}{(2k)!} \left(f^{(2k-1)}(n) - f^{(2k-1)}(m) \right) + R_p.$$

The Poisson distribution. Let $\lambda > 0$. Then X is a **Poisson random variable** with parameter λ if

$$\mathrm{Prob}(X = n) = \begin{cases} \lambda^n e^{-\lambda}/n! & \text{if } n \in \{0, 1, 2, \ldots\} \\ 0 & \text{otherwise.} \end{cases}$$

We write $X \sim \mathrm{Pois}(\lambda)$. The mean and the variance are both λ.

non-neg? yes

Sum to 1?

$$\sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}$$

$$= e^{-\lambda} e^{\lambda} = 1 \quad \blacksquare$$

note $\frac{\lambda^n}{n!} \le \frac{\lambda^n}{n^n e^{-n}} = \left(\frac{e\lambda}{n}\right)^n \to 0$ fast

dominated by geometric series

Poisson random variable: $\mu_X = \lambda$

$$e^{\lambda} = \sum_{n=0}^{\infty} \lambda^n / n!$$

apply $\lambda \frac{d}{d\lambda}$

$$\lambda e^{\lambda} = \sum_{n=1}^{\infty} \lambda \frac{d}{d\lambda}\left(\frac{\lambda^n}{n!}\right)$$

$$= \sum_{n=1}^{\infty} \frac{n\lambda^n}{n!}$$

$$\lambda = \sum_{n=0}^{\infty} n \frac{\lambda^n e^{-\lambda}}{n!}$$

$$= E[X]$$

can extend down to 0 as the 0 terms is 0

issue!

$$E[X] = \sum_{n=0}^{\infty} n \, \mathrm{Prob}(X = n)$$

$$= \sum_{n=0}^{\infty} n \cdot \frac{\lambda^n e^{-\lambda}}{n!}$$

$$= \sum_{n=1}^{\infty} \frac{n}{n!} \lambda^n e^{-\lambda} \qquad (\lambda^n = \lambda^{n-1}\lambda)$$

$$= \sum_{n=1}^{\infty} \frac{\lambda \cdot \lambda^{n-1} e^{-\lambda}}{(n-1)!}$$

$$= \lambda \sum_{m=0}^{\infty} \frac{\lambda^m e^{-\lambda}}{m!} \qquad \begin{array}{l} m = n-1 \\ n: 1 \to \infty \\ m: 0 \to \infty \end{array}$$

$$= \lambda \cdot 1 = \lambda$$

$\lambda(\lambda-1)$ vs λ^2

Sums of Poisson random variables. The sum of n independent Poisson random variables with parameters $\lambda_1, \ldots, \lambda_n$ is a Poisson random variable with parameter $\lambda_1 + \cdots + \lambda_n$.

$$\big(\underbrace{X_{\lambda_1} + X_{\lambda_2}}\big) + X_{\lambda_3} \ldots$$

$$\underbrace{X_{\lambda_1+\lambda_2} + X_{\lambda_3}}_{X_{\lambda_1+\lambda_2+\lambda_3}}$$

Proof: $Y = X_{\lambda_1} + X_{\lambda_2}$

$$\mathrm{Prob}(Y=n) = \sum_{k=0}^{n} \mathrm{Prob}(X_{\lambda_1}=k)\,\mathrm{Prob}(X_{\lambda_2}=n-k)$$

$$= \sum_{k=0}^{n} \frac{\lambda_1^k e^{-\lambda_1}}{k!} \frac{\lambda_2^{n-k} e^{-\lambda_2}}{(n-k)!} \frac{n!}{n!}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \Big[\underbrace{\sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k}}_{\text{Binomial Thm } (\lambda_1+\lambda_2)^n} \Big] = \frac{(\lambda_1+\lambda_2)^n e^{-(\lambda_1+\lambda_2)}}{n!}$$

Definition 20.2.1 (Normal distribution) *A random variable X is normally distributed (or has the normal distribution, or is a Gaussian random variable) with mean μ and variance σ^2 if the density of X is*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

We often write $X \sim N(\mu, \sigma^2)$ to denote this. If $\mu = 0$ and $\sigma^2 = 1$, we say X has the standard normal distribution.

Theorem 20.2.2 (Central Limit Theorem (CLT)) *Let $X_1, \dots, X_N$ be independent, identically distributed random variables whose moment generating functions converge for $|t| < \delta$ for some $\delta > 0$ (this implies all the moments exist and are finite). Denote the mean by μ and the variance by σ^2, let*

$$\overline{X}_N = \frac{X_1 + \cdots + X_N}{N}$$

and set

$$Z_N = \frac{\overline{X}_N - \mu}{\sigma/\sqrt{N}}.$$

Then as $N \to \infty$, the distribution of Z_N converges to the standard normal (see Definition 20.2.1 for a statement).

$$\mathrm{Var}(\overline{X}_N) = \sum \mathrm{Var}\left(\frac{X_k}{N}\right)$$
$$= \frac{1}{N^2} \sum \mathrm{Var}(X_k)$$
$$= \frac{1}{N^2} N \sigma^2$$
$$= \frac{\sigma}{N}$$

1. X has a Poisson distribution with parameter λ means

$$\text{Prob}(X = n) = \begin{cases} \frac{\lambda^n e^{-\lambda}}{n!} & \text{if } n \geq 0 \text{ is an integer} \\ 0 & \text{otherwise.} \end{cases}$$

2. If $X_1, \ldots, X_N$ are independent, identically distributed random variables with mean μ, variance σ^2 and a little more (such as the third moment is finite, or the moment generating function exists), then $X_1 + \cdots + X_N$ converges to being normally distributed with mean $n\mu$ and variance $n\sigma^2$.

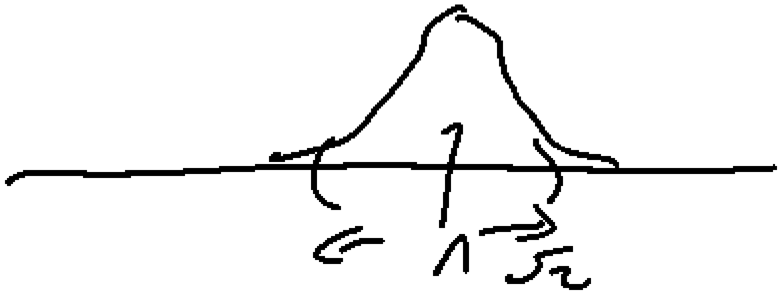
$X_k \sim \text{Poiss}(1)$

$E[X_k] = 1$

$Var(X_k) = 1$

$X = X_1 + \cdots + X_n \sim \text{Poiss}(n)$

$E[X] = n \qquad Var(X) = n$



X converges to $N(n, n) = N(\mu, \sigma^2)$

$$\text{Prob}(X = n) = \frac{\lambda^n e^{-\lambda}}{n!} = \frac{n^n e^{-n}}{n!}$$

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

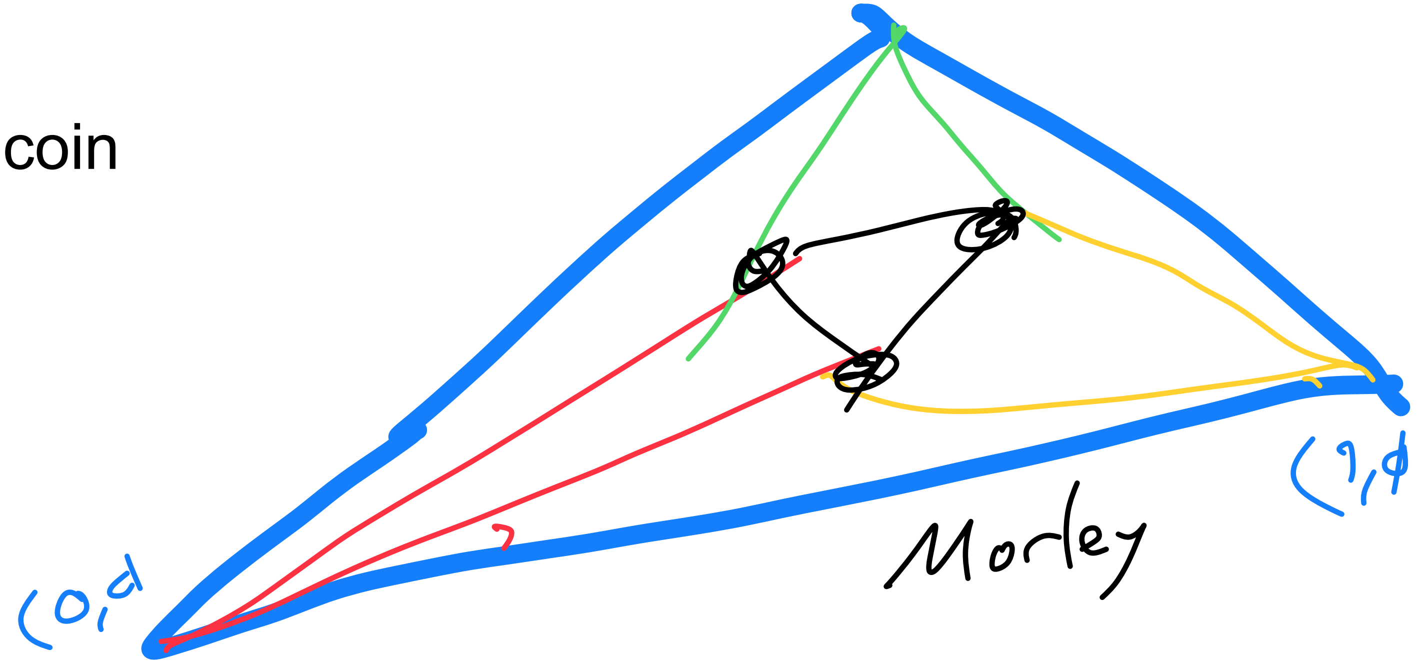
Lecture 18: 4/22/25: CLT to Stirling, CLT for random walk of fair coin tosses, intro to generating fns:
Video: https://youtu.be/PAgTajBx5ak

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 2: November , 2021:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- CLT to Stirling via Poisson
- Central Limit Theorem for fair coin
- Random Walks….
- Generating Functions
- Poisson Random Variables



General items.

- Power of Stirling’s Formula
- Intuition from Special Cases, but dangers…. (prime counting?)
- Finding good approach through algebra: Generating Functions

X approx $\mathcal{N}(n, n)$

$X_1, \ldots, X_n$ indep Poiss(1), $X = X_1 + \cdots + X_n \sim \text{Poiss}(n)$

$$\text{Prob}(X = n) = \frac{\lambda^n e^{-\lambda}}{n!} = \frac{n^n e^{-n}}{n!} \quad \text{as } \lambda = n \text{ for } X$$

By CLT, X is approx normal with mean n and variance n

$$E[X] = nE[X_1] \quad \text{and} \quad \text{Var}(X) = n\text{Var}(X_1)$$

$$\text{Prob}(X = n) \approx \int_{n-1/2}^{n+1/2} \frac{1}{\sqrt{2\pi n}} \underbrace{e^{-(x-n)^2/2n}}_{\text{essentially } 1, \text{ fluctuations of size } e^{\pm 1/8n}} dx$$

$$\frac{n^n e^{-n}}{n!} \approx 1 \cdot \frac{1}{\sqrt{2\pi n}} \cdot 1 \Rightarrow n! \approx n^n e^{-n} \sqrt{2\pi n}$$

estimating Prob$(X = n)$ by normal curve from $n - \frac{1}{2}$ to $n + \frac{1}{2}$

Definition 20.2.1 (Normal distribution) *A random variable X is normally distributed (or has the normal distribution, or is a Gaussian random variable) with mean μ and variance σ^2 if the density of X is*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

We often write $X \sim N(\mu, \sigma^2)$ to denote this. If $\mu = 0$ and $\sigma^2 = 1$, we say X has the standard normal distribution.

Theorem 20.2.2 (Central Limit Theorem (CLT)) *Let $X_1, \ldots, X_N$ be independent, identically distributed random variables whose moment generating functions converge for $|t| < \delta$ for some $\delta > 0$ (this implies all the moments exist and are finite). Denote the mean by μ and the variance by σ^2, let*

$$\overline{X}_N = \frac{X_1 + \cdots + X_N}{N}$$

and set

$$Z_N = \frac{\overline{X}_N - \mu}{\sigma/\sqrt{N}}.$$

Then as $N \to \infty$, the distribution of Z_N converges to the standard normal (see Definition 20.2.1 for a statement).

$\mathrm{Var}(\overline{X}_N) = \frac{1}{N^2} N \mathrm{Var}(X_1) = \sigma^2/N$

$\sqrt{\mathrm{Var}(\overline{X}_N)} = \sigma/\sqrt{N}$

The Gamma function. The Gamma function $\Gamma(s)$ is

$$\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx, \quad \Re(s) > 0.$$

n - non-neg integer

$\Gamma(n+1) = n!$

Stirling's formula: As $n \to \infty$, we have

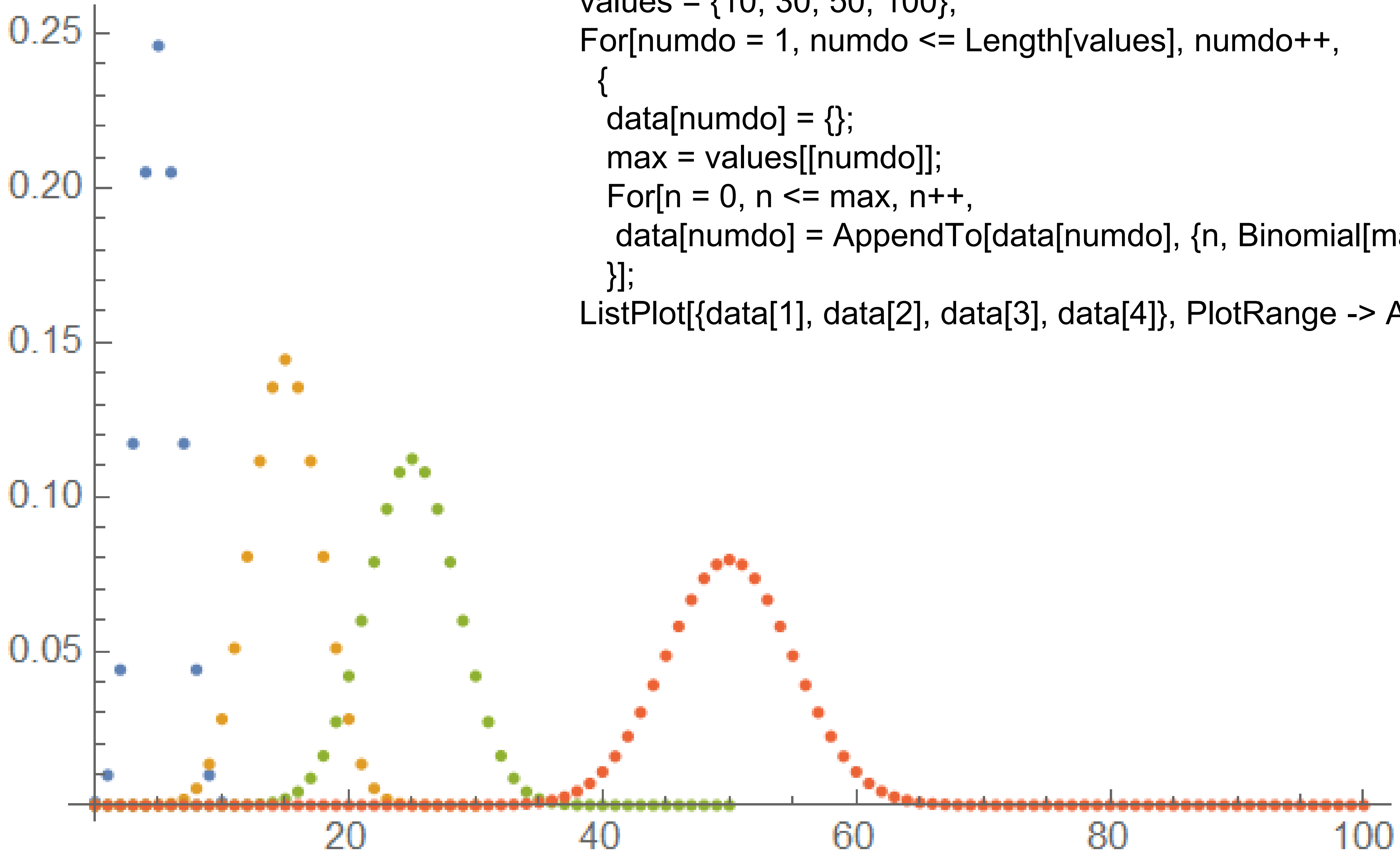
$$n! \approx n^n e^{-n} \sqrt{2\pi n};$$

by this we mean

$$\lim_{n\to\infty} \frac{n!}{n^n e^{-n}\sqrt{2\pi n}} = 1.$$

More precisely, we have the following series expansion:

$$n! = n^n e^{-n} \sqrt{2\pi n} \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} - \cdots \right).$$




```
values = {10, 30, 50, 100};
For[numdo = 1, numdo <= Length[values], numdo++,
  {
   data[numdo] = {};
   max = values[[numdo]];
   For[n = 0, n <= max, n++,
    data[numdo] = AppendTo[data[numdo], {n, Binomial[max, n]/2^max}]];
   }];
ListPlot[{data[1], data[2], data[3], data[4]}, PlotRange -> All]
```

$$\mathrm{Prob}(X_i = n) = \begin{cases} 1/2 & \text{if } n = 1 \\ 1/2 & \text{if } n = -1 \\ 0 & \text{otherwise.} \end{cases} \qquad S_{2N} = X_1 + \cdots + X_{2N}$$

$$Var(X_1) = (1-\mu)^2\, Prob(X_1 = 1) + (-1-\mu)^2\, Prob(X_1 = -1) = 1$$

$$E[S_{2N}] = 0 \qquad Var(S_{2N}) = 2N \cdot Var(X_1) = 2N$$

$$S_{2N} \to N(0, 2N) \qquad \text{or study} \quad \left(\frac{S_{2N}}{2N} - 0\right) \Big/ Var\left(\frac{S_{2N}}{2N}\right)$$

$$Prob(S_{2N} = 0) = \binom{2N}{N} \left(\frac{1}{2}\right)^N \left(\frac{1}{2}\right)^N$$

$$Prob(S_{2N} = 2k) = \binom{2N}{N+k} \frac{1}{2^{2N}}$$

$$\#H - \#Tails = 2k$$
$$\#H + \#T = 2N$$
$$2\#H = 2N + 2k$$
$$\#H = N + k$$
$$\#T = N - k$$

What k must we study? $Var(S_{2N}) = 2N$

$StDev(S_{2N}) = \sqrt{2N}$

Ave Value of k is 0, and StDev is of size $\sqrt{N}$

$$\text{Prob}\left(|k| \geq N^{3/4}\right) = \text{Prob}\left(|k| \geq N^{\frac{1}{4}} \cdot N^{\frac{1}{2}}\right) \leq \frac{\text{Const}}{(N^{1/4})^2}$$

Enough to study $|k| \leq \log N \cdot \sqrt{N}$ or $N^{\frac{1}{2}+\epsilon}$

$$\binom{2N}{N+k} \frac{1}{2^{2N}} = \frac{(2N)!}{(N+k)!\,(N-k)!} \frac{1}{2^{2N}}$$

Want large for Stirling

Need $2N$, $N+k$, $N-k$ large

Need k bounded away from $\pm N$

This is why should enough to study $|k|$ at most $\log N \cdot N^{1/2}$ or $N^{\frac{1}{2}+\epsilon}$

Stirling: $n! \approx n^n e^{-n} \sqrt{2\pi n}$

$$\frac{(2N)!}{(N+k)!\,(N-k)!}\,\frac{1}{2^{2N}} = \frac{(2N)^{2N}\, e^{-2N} \sqrt{2\pi\cdot 2N} \,/\, 2^{2N}}{(N+k)^{N+k}\, e^{-(N+k)} \sqrt{2\pi(N+k)}\,(N-k)^{N-k}\, e^{-(N-k)} \sqrt{2\pi(N-k)}}$$

$$= \frac{\cancel{2^{2N}}\, N^{2N}\, \cancel{e^{-2N}} \sqrt{2\cancel{\pi}\cdot 2N}\;\cancel{2^{-2N}}}{(N+k)^{N+k}\,(N-k)^{N-k}\, \cancel{e^{-2N}} \sqrt{2\pi\cdot 2\cancel{\pi}\cdot (N-k)(N+k)}}$$

$$= \sqrt{\frac{2N}{2\pi(N-k)(N+k)}}\;\frac{N^{2N}}{(N+k)^{N+k}\,(N-k)^{N-k}}$$

$$\sqrt{\frac{2}{2\pi N}} = \sqrt{\frac{2\cdot 2}{2\pi\cdot 2N}} = \frac{2}{\sqrt{2\pi\cdot 2N}}$$

$$\frac{N^{2N}}{(N+k)^{N+k}(N-k)^{N-k}} = \frac{1}{\left(1+\frac{k}{N}\right)^{N+k}\left(1-\frac{k}{N}\right)^{N-k}}$$

$|k| \le \log N \cdot N^{1/2}$ or $N^{\frac{1}{2}+\varepsilon}$ with HIGH probability

as $N \to \infty$

$$\left.\begin{aligned} \left(1+\frac{k}{N}\right)^{N+k} &\to e^{k} \\ \left(1-\frac{k}{N}\right)^{N-k} &\to e^{-k} \end{aligned}\right\} \text{product} \to 1$$

too crude

PAVLOVIAN RESPONSE!

Take log of the product

$$P = \left(1+\frac{k}{N}\right)^{N+k}\left(1-\frac{k}{N}\right)^{N-k}$$

$$\log P = (N+k)\log\left(1+\frac{k}{N}\right) + (N-k)\log\left(1-\frac{k}{N}\right)$$

$$\log(1-u) = -u - u^2/2 - u^3/3 \cdots$$

$$\log(1+u) = u - u^2/2 + u^3/3 \cdots$$

If u small and pos, $\log(1-u)$ is neg and $\log(1+u)$ is pos

as $|k| \le \log N \cdot N^{1/2}$ or $N^{1/2+\epsilon}$, only need quadratic

$$\log P = (N+k)\left[\frac{k}{N} - \frac{k^2}{2N^2} + \cdots\right] + (N-k)\left[-\frac{k}{N} - \frac{k^2}{2N^2} + \cdots\right]$$

$$= -\frac{k^2}{2N} + \frac{k^2}{N} = -\frac{k^2}{N} = -\frac{(2k/2)^2}{N} = -\frac{(2k)^2}{4N} = -\frac{(2k)^2}{2\cdot 2N}$$

$$= -\frac{(2k-c)^2}{2\cdot 2N}$$

Thus $p \simeq e^{-(2k-0)^2/2\cdot 2N}$

$$\text{Prob}(S_{2N} = 2k) = \frac{2}{\sqrt{2\pi \cdot 2N}}\, e^{-(2k-0)^2/2\cdot 2N}$$

$$\int_{2k-1}^{2k+1} \underset{(\text{mean } 0,\ \text{Var } 2N)}{\text{Normal Density}}(x)\,dx$$

Prob p of heads:

$$\frac{1}{2^{2N}} \longrightarrow p^{N+k}\,(1-p)^{N-k}$$

Definition 19.2.1 (Generating Function) *Given a sequence $\{a_n\}_{n=0}^{\infty}$, we define its generating function by*

$$G_a(s) = \sum_{n=0}^{\infty} a_n s^n$$

for all s where the sum converges.

Simplest? $a_n = 0$ yields $G_{\vec{0}}(s) = 0$ converges for all s

$a_n = 1$ yields $G_{\vec{1}}(s) = \sum_{n=0}^{\infty} s^n = \frac{1}{1-s}$ if $|s|<1$

$a_n = 1/n!$ Get $G_{\frac{1}{!}}(s) = e^s$ converges for all s

Hope: $G_a(s)\, G_b(s) = G_{\underline{f(a,b)}}(s)$

Generating Function (Example: Binet's Formula)

Binet's Formula

$F_0=0$

$$F_1 = F_2 = 1;\ F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{-1+\sqrt{5}}{2}\right)^n\right].$$

$0, 1, 1, 2, 3, 5, 8, \ldots$

- Recurrence relation: $F_{n+1} = F_n + F_{n-1}$ (1)
- Generating function: $g(x) = \sum_{n>0} F_n x^n$.

$$(1) \Rightarrow \sum_{n\geq 2} F_{n+1}x^{n+1} = \sum_{n\geq 2} F_n x^{n+1} + \sum_{n\geq 2} F_{n-1}x^{n+1}$$

$$\Rightarrow \sum_{n\geq 3} F_n x^n = \sum_{n\geq 2} F_n x^{n+1} + \sum_{n\geq 1} F_n x^{n+2}$$

$$\Rightarrow \sum_{n\geq 3} F_n x^n = x\sum_{n\geq 2} F_n x^n + x^2\sum_{n\geq 1} F_n x^n$$

$$\Rightarrow g(x) - F_1 x - F_2 x^2 = x(g(x) - F_1 x) + x^2 g(x)$$

$$\Rightarrow g(x) = x/(1-x-x^2).$$

Partial Fraction Expansion (Example: Binet's Formula)

- Generating function: $g(x) = \sum_{n>0} \boldsymbol{F}_n x^n = \frac{x}{1-x-x^2}$.

$r = (x+x^2)$

- Partial fraction expansion:

$$\Rightarrow \quad g(x) = \frac{x}{1-x-x^2} = \frac{1}{\sqrt{5}}\left(\frac{\frac{1+\sqrt{5}}{2}x}{1-\frac{1+\sqrt{5}}{2}x} - \frac{\frac{-1+\sqrt{5}}{2}x}{1-\frac{-1+\sqrt{5}}{2}x}\right).$$

Coefficient of x^n (power series expansion):

$$\boldsymbol{F}_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{-1+\sqrt{5}}{2}\right)^n\right]$$ - Binet's Formula!

(using geometric series: $\frac{1}{1-r} = 1 + r + r^2 + r^3 + \cdots$).

Math / Stat 341: Probability (Spring 2025)

Steven J Miller
Williams College
sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/

Lecture 19: 4/24/25: Generating Functions and Moment Generating Functions, Properties of MGF, Poisson and Normal Example, Poisson to CLT:
Video: https://youtu.be/x-RNxpyi_4I

https://web.williams.edu/Mathematics/sjmiller/public_html/341Sp25/Math341Sp25LecNotes.pdf

Plan for the day: Lecture 25: November 15, 2021:

https://web.williams.edu/Mathematics/sjmiller/public_html/341Fa21/handouts/341Notes_Chap1.pdf

- Generating Functions
- Moment Generating Functions
- Characteristic Functions
- Change of Base Formula

General items.

- Find the path through the algebra…. (telescoping, cubes)

Generating Function (Example: Binet's Formula)

Binet's Formula

$F_0 = 0$

$$F_1 = F_2 = 1; \; F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{-1+\sqrt{5}}{2}\right)^n\right].$$

0, 1, 1, 2, 3, 5, 8, …

- Recurrence relation: $F_{n+1} = F_n + F_{n-1}$ (1)
- Generating function: $g(x) = \sum_{n>0} F_n x^n$.

$$(1) \Rightarrow \sum_{n\geq 2} F_{n+1}x^{n+1} = \sum_{n\geq 2} F_n x^{n+1} + \sum_{n\geq 2} F_{n-1}x^{n+1}$$

$$\Rightarrow \sum_{n\geq 3} F_n x^n = \sum_{n\geq 2} F_n x^{n+1} + \sum_{n\geq 1} F_n x^{n+2}$$

$$\Rightarrow \sum_{n\geq 3} F_n x^n = x\sum_{n\geq 2} F_n x^n + x^2 \sum_{n\geq 1} F_n x^n$$

$$\Rightarrow g(x) - F_1 x - F_2 x^2 = x(g(x) - F_1 x) + x^2 g(x)$$

$$\Rightarrow g(x) = x/(1 - x - x^2).$$

Partial Fraction Expansion (Example: Binet's Formula)

- Generating function: $g(x) = \sum_{n>0} F_n x^n = \frac{x}{1-x-x^2}$. $= x \frac{1}{1-r}$

$r = (x+x^2)$

- Partial fraction expansion:

$$\Rightarrow \; g(x) \; = \; \frac{x}{1-x-x^2} = \frac{1}{\sqrt{5}}\left(\frac{\frac{1+\sqrt{5}}{2}x}{1-\frac{1+\sqrt{5}}{2}x} - \frac{\frac{-1+\sqrt{5}}{2}x}{1-\frac{-1+\sqrt{5}}{2}x}\right).$$

Coefficient of x^n (power series expansion):

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{-1+\sqrt{5}}{2}\right)^n\right]$$ - Binet's Formula!

(using geometric series: $\frac{1}{1-r} = 1 + r + r^2 + r^3 + \cdots$).

$$\frac{1}{1-(x+x^2)} = 1 + (x+x^2) + (x+x^2)^2 + (x+x^2)^3 + \cdots$$

$$= 1 + x + x^2$$

$$x^2 + 2x^3 + x^4$$

$$x^3 + 3x^4 + 3x^5 + x^6$$

x and x^2 different degrees

$$\frac{1}{1-x-x^2} = \frac{1}{(1-\alpha x)(1-\beta x)} = \frac{A}{1-\alpha x} + \frac{B}{1-\beta x}$$

$$= A(1+\alpha x + \alpha^2 x^2 + \cdots) + B(1+\beta x + \beta^2 x^2 + \cdots)$$

$$= \underbrace{(A+B)}_{F_0} + \underbrace{(A\alpha + B\beta)}_{F_1} x + \underbrace{(A\alpha^2 + B\beta^2)}_{F_2} x^2 + \cdots$$

X has a Poisson distribution with parameter λ means

$$\text{Prob}(X = n) = \begin{cases} \frac{\lambda^n e^{-\lambda}}{n!} & \text{if } n \geq 0 \text{ is an integer} \\ 0 & \text{otherwise.} \end{cases}$$

$\text{Poiss}(\lambda_1) + \text{Poiss}(\lambda_2) = \text{Poiss}(\lambda_1 + \lambda_2)$

algebra: $\sum_{m=0}^{n} \text{Prob}(X_1 = m)\,\text{Prob}(X_2 = n-m)$

$= \text{Prob}(X_1 + X_2 = n)$

Definition 19.2.1 (Generating Function) *Given a sequence $\{a_n\}_{n=0}^{\infty}$, we define its generating function by*

$$G_a(s) = \sum_{n=0}^{\infty} a_n s^n$$

for all s where the sum converges.

Warning: $\begin{cases} 0 & \text{if } x=0 \\ \exp(-1/x^2) & \text{if } x\neq 0 \end{cases}$

$$(Tf)(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$$

Use L'Hopital, find $f^{(n)}(0)=0$

Thus $(Tf)(x) \equiv 0$

$\sin x$ and $\sin x + 1701 f(x)$ have same Taylor series

Theorem 19.3.1 (Uniqueness of generating functions of sequences) *Let* $\{a_n\}_{n=0}^{\infty}$ *and* $\{b_n\}_{n=0}^{\infty}$ *be two sequences of numbers with generating functions* $G_a(s)$ *and* $G_b(s)$ *which converge for* $|s| < \delta$. *Then the two sequences are equal (i.e.,* $a_i = b_i$ *for all* i*) if and only if* $G_a(s) = G_b(s)$ *for all* $|s| < \delta$. *We may recover the sequence from the generating function by differentiating:* $a_n = \frac{1}{n!}\frac{d^n G_a(s)}{ds^n}.\Big|_{s=0}$

Proof: $a_n = b_n$ trivially implies $G_a(s) = G_b(s)$

Assume $G_a(s) = G_b(s)$

so $a_0 + a_1 s + \cdots = b_0 + b_1 s + \cdots \quad \forall |s| < \delta$

at $s=0$ get $a_0 = b_0$

study $\frac{G_a(s) - a_0}{s} = \frac{G_b(s) - a_0}{s} \Rightarrow a_1 + a_2 s + \cdots = b_1 + b_2 s + \cdots \quad \forall |s| < \delta$

take $s=0$ and get $a_1 = b_1$, continue.... $\blacksquare$

Definition 19.4.1 (Convolution of sequences) *If we have two sequences* $\{a_m\}_{m=0}^{\infty}$ *and* $\{b_n\}_{n=0}^{\infty}$, *we define their convolution to be the new sequence* $\{c_k\}_{k=0}^{\infty}$ *given by*

$$c_k = a_0 b_k + a_1 b_{k-1} + \cdots + a_{k-1} b_1 + a_k b_0 = \sum_{\ell=0}^{k} a_\ell b_{k-\ell}.$$

We frequently write this as $c = a * b$.

$G_a(s) = a_0 + a_1 s + a_2 s^2 + \cdots \qquad G_b(s) = b_0 + b_1 s + b_2 s^2 + \cdots$

$G_a(s) G_b(s) = (a_0 + a_1 s + a_2 s^2 + \cdots)(b_0 + b_1 s + b_2 s^2 + \cdots)$

$= \underbrace{a_0 b_0}_{c_0} + (\underbrace{a_0 b_1 + a_1 b_0}_{c_1}) s + (\underbrace{a_0 b_2 + a_1 b_1 + a_2 b_0}_{c_2}) s^2 + \cdots$

$= G_c(s) = G_{a*b}(s)$

Lemma 19.4.2 *Let $G_a(s)$ be the generating function for $\{a_m\}_{m=0}^{\infty}$ and $G_b(s)$ the generating function for $\{b_n\}_{n=0}^{\infty}$. Then the generating function of $c = a * b$ is $G_c(s) = G_a(s)G_b(s)$.*

$a * b * c \;=\; G_a(s)\, G_b(s)\, G_c(s) = G_{a*b*c}(s)$

Grouping: $(a*b)*c$

$$G_{(a*b)*c}(s) = G_{a*b}(s)\, G_c(s)$$

$$= G_a(s)\, G_b(s)\, G_c(s)$$

Commutativity is trivial:

$G_a(s)\, G_b(s) = G_b(s)\, G_a(s)$ so $G_{a*b}(s) = G_{b*a}(s)$

$$c_n = \sum_{\ell=0}^{n} a_\ell a_{n-\ell} = \sum_{\ell=0}^{n} \binom{n}{\ell}\binom{n}{n-\ell} = \sum_{\ell=0}^{n} \binom{n}{\ell}^2 \qquad \sum_{k=0}^{2n} c_k s^k = G_c(s) = G_a(s)^2 = (1+s)^n \cdot (1+s)^n = (1+s)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k} s^k$$

Think Binomial : $c = a * a$

$$G_c(s) = G_{a*a}(s) = G_a(s)\,G_a(s)$$

$$G_a(s) = \sum_{\ell=0}^{n} \binom{n}{\ell} 1^{n-\ell} s^\ell = (1+s)^n$$ from $a_\ell = \binom{n}{\ell}$ for $0 \le \ell \le n$

$$G_c(s) = (1+s)^n (1+s)^n = (1+s)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k} s^k$$

$$\overset{\|}{\sum_{k=0}^{2n}} c_k s^k$$

$$c_n = \sum_{k=0}^{2n} \binom{n}{k}^2 = \sum_{k=0}^{2n} \binom{n}{k}\binom{n}{n-k} = \binom{2n}{n}$$

Definition 19.4.3 (Probability generating function) *Let X be a discrete random variable taking on values in the integers. Let $G_X(s)$ be the generating function to $\{a_m\}_{m=-\infty}^{\infty}$ with $a_m = \text{Prob}(X = m)$. Then $G_X(s)$ is called the probability generating function. If X is only non-zero at the integers, a very useful way of computing $G_X(s)$ is to note that*

$$G_X(s) = \mathbb{E}[s^X] = \sum_{m=-\infty}^{\infty} s^m \text{Prob}(X = m).$$

More generally, if the probabilities are non-zero on an at most countable set $\{x_m\}$, then

$$G_X(s) = E[s^X] = \sum_m s^{x_m} \text{Prob}(X = x_m).$$

Poisson λ

$$\text{Prob}(X = m) = \frac{\lambda^m e^{-\lambda}}{m!}$$

$$G_X(s) = \mathbb{E}[s^X] = \sum_{m=0}^{\infty} s^m \frac{\lambda^m e^{-\lambda}}{m!} = \left(\sum_{m=0}^{\infty} \frac{(s\lambda)^m}{m!}\right) e^{-\lambda}$$

$$= e^{s\lambda} e^{-\lambda}$$

$$= e^{(s-1)\lambda}$$

$X_i \sim \text{Poiss}(\lambda_i)$

$$G_{X_1+X_2}(s) = G_{X_1}(s) G_{X_2}(s) = e^{(s-1)\lambda_1} e^{(s-1)\lambda_2} = e^{(s-1)(\lambda_1+\lambda_2)}$$

This is PGF of Poiss$(\lambda_1+\lambda_2)$

Theorem 19.4.4 *Let $X_1, \ldots, X_n$ be independent discrete random variables taking on non-negative integer values, with corresponding probability generating functions $G_{X_1}(s), \ldots, G_{X_n}(s)$. Then*

$$G_{X_1+\cdots+X_n}(s) = G_{X_1}(s) \cdots G_{X_n}(s).$$

Proof: $G_{X_1+\cdots X_n}(s) = E[s^{X_1+\cdots X_n}]$

$= E[s^{X_1} \cdots s^{X_n}]$

$= E[s^{X_1}] \cdots E[s^{X_n}]$

$= G_{X_1}(s) \cdots G_{X_n}(s)$

Shorter proof: adding is convolving use Lemma 19.4.2

Lemma 19.4.2 *Let $G_a(s)$ be the generating function for $\{a_m\}_{m=0}^{\infty}$ and $G_b(s)$ the generating function for $\{b_n\}_{n=0}^{\infty}$. Then the generating function of $c = a * b$ is $G_c(s) = G_a(s)G_b(s)$.*

The density of the sum of independent discrete random variables is the convolution of their probabilities!

Review, did this before

Since study sums all the time, shows the

value of convolutions . . . .

Definition 19.5.1 (Probability generating function) *Let X be a continuous random variable with density f. Then*

$$G_X(s) = \int_{-\infty}^{\infty} s^x f(x)dx$$

is the probability generating function of X.

Replace Σ with $\int$

$G_X(s) = \mathbb{E}[s^X]$

Maybe the correct general definition of generating function is the expectation

Definition 19.5.2 (Convolution of functions) *The convolution of two functions f_1 and f_2, denoted $f_1 * f_2$, is*

$$(f_1 * f_2)(x) = \int_{-\infty}^{\infty} f_1(t) f_2(x-t) dt.$$

If the f_i's are densities then the integral converges.

Proof: If f_i is density of X_i, and X_1, X_2 are indep. then $f_1 * f_2$ is the density of $X_1 + X_2$.

$$\int_{x=-\infty}^{\infty} (f_1 * f_2)(x) dx = \int_{x=-\infty}^{\infty} \int_{t=-\infty}^{\infty} f_1(t) f_2(x-t) dt\, dx$$

Fubini: if integral of abs value is finite can change order, ok here

$$= \int_{t=-\infty}^{\infty} f_1(t) \left[\int_{x=-\infty}^{\infty} f_2(x-t) dx \right] dt = \int_{t=-\infty}^{\infty} f_1(t) dt \int_{x=-\infty}^{\infty} f_2(u) du$$

$$= 1 \cdot 1 = 1$$

Theorem 19.5.3 (Sums of continuous random variables) *The probability density function of the sum of independent continuous random variables is the convolution of their probability density functions. In particular, if $X_1, \ldots, X_n$ have densities $f_1, \ldots, f_n$, then the density of $X_1 + \cdots + X_n$ is $f_1 * f_2 * \cdots * f_n$.*

Proved when n=2, general case by grouping

Theorem 19.5.4 (Commutativity of convolution) *The convolution of two sequences or functions is* ***commutative****; in other words, $a*b = b*a$ or $f_1 * f_2 = f_2 * f_1$.*

Trivial if f_1, f_2 are densities as have an interpretation

Then $X_1 + X_2 = X_2 + X_1$

in general: do some algebra/calculus

Definition 19.6.1 (Moments) *Let X be a random variable with density f. Its k^{th}* ***moment****, denoted μ'_k, is defined by*

$$\mu'_k \ := \ \sum_{m=0}^{\infty} x_m^k f(x_m)$$

if X is discrete, taking non-zero values only at the x_m's, and for continuous X by

$$\mu'_k \ := \ \int_{-\infty}^{\infty} x^k f(x) dx.$$

In both cases we denote this as $\mu'_k = \mathbb{E}[X^k]$. We define the k^{th} ***centered moment****, μ_k, by $\mu_k := \mathbb{E}[(X - \mu'_1)^k]$. We frequently write μ for μ'_1 and σ^2 for μ_2.*

Note $\mu_1 = 0$, $\mu_1' = \mu$

$X \to aX + b$ $\left[\begin{array}{l} \mu = 0 \\ \sigma = 1 \end{array} \right.$

Definition 19.6.2 (Moment generating function) *Let X be a random variable with density f. The moment generating function of X, denoted $M_X(t)$, is given by $M_X(t) = \mathbb{E}[e^{tX}]$. Explicitly, if X is discrete then*

$$M_X(t) = \sum_{m=-\infty}^{\infty} e^{tx_m} f(x_m),$$

while if X is continuous then

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx. \quad = \mathbb{E}[e^{tX}]$$

Note $M_X(t) = G_X(e^t)$, or equivalently $G_X(s) = M_X(\log s)$.

When see terminology such as this, need to justify the name….

$$G_X(s) = \mathbb{E}[s^X]$$

$$M_X(s) = \mathbb{E}[e^{tX}]$$

$$s \leftrightarrow e^t$$

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

$$= \int_{x=-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{t^n x^n}{n!} f(x)\,dx = \sum_{n=0}^{\infty} \frac{1}{n!}\left[\int_{x=-\infty}^{\infty} x^n f(x)dx\right] t^n$$

$$= \sum_{n=0}^{\infty} \frac{\mu_n'}{n!} t^n$$

$$\boxed{\frac{d^n}{dt^n} M_X(t)\Big|_{t=0} = \mu_n'}$$

Theorem 19.6.3 *Let X be a random variable with moments μ'_k.*

1. *We have*

$$M_X(t) \;=\; 1 + \mu'_1 t + \frac{\mu'_2 t^2}{2!} + \frac{\mu'_3 t^3}{3!} + \cdots;$$

   in particular, $\mu'_k = d^k M_X(t)/dt^k \Big|_{t=0}$.

2. *Let α and β be constants. Then*

$$M_{\alpha X+\beta}(t) \;=\; e^{\beta t} M_X(\alpha t).$$

   Useful special cases are $M_{X+\beta}(t) = e^{\beta t} M_X(t)$ and $M_{\alpha X}(t) = M_X(\alpha t)$; when proving the central limit theorem, it's also useful to have $M_{(X+\beta)/\alpha}(t) = e^{\beta t/\alpha} M_X(t/\alpha)$.

3. *Let X_1 and X_2 be independent random variables with moment generating functions $M_{X_1}(t)$ and $M_{X_2}(t)$ which converge for $|t| < \delta$. Then*

$$M_{X_1+X_2}(t) \;=\; M_{X_1}(t) M_{X_2}(t).$$

   More generally, if $X_1, \dots, X_N$ are independent random variables with moment generating functions $M_{X_i}(t)$ which converge for $|t| < \delta$, then

$$M_{X_1+\cdots+X_N}(t) = M_{X_1}(t) M_{X_2}(t) \cdots M_{X_N}(t).$$

   If the random variables all have the same moment generating function $M_X(t)$, then the right hand side becomes $M_X(t)^N$.

$M_X(t) = E[e^{tX}]$

$Y = \alpha X + \beta$

$M_Y(t) = E[e^{t(\alpha X+\beta)}]$

$= E[e^{t\alpha X} \cdot e^{t\beta}]$

$= e^{\beta t} E[e^{(\alpha t) X}]$

$= e^{\beta t} M_X(\alpha t)$ ∎

Theorem 19.6.5 ***(Uniqueness of moment generating functions for discrete random variables.)*** *Let X and Y be discrete random variables taking on non-negative integer values (i.e., they're non-zero only in $\{0, 1, 2, \ldots\}$) with moment generating functions $M_X(t)$ and $M_Y(t)$, each of which converges for $|t| < \delta$. Then X and Y have the same distribution if and only if there is an $r > 0$ such that $M_X(t) = M_Y(t)$ for $|t| < r$.*

There exist distinct probability distributions which have the same moments. In other words, knowing all the moments doesn't always uniquely determine the probability distribution.

Example 19.6.6 *The standard examples given are the following two densities, defined for $x \geq 0$ by*

$$\begin{aligned} f_1(x) &= \frac{1}{\sqrt{2\pi x^2}}\, e^{-(\log^2 x)/2} \\ f_2(x) &= f_1(x)\left[1 + \sin(2\pi \log x)\right]. \end{aligned} \tag{19.2}$$

It's a nice calculation to show that these two densities have the same moments; they're clearly different (see Figure 19.1).

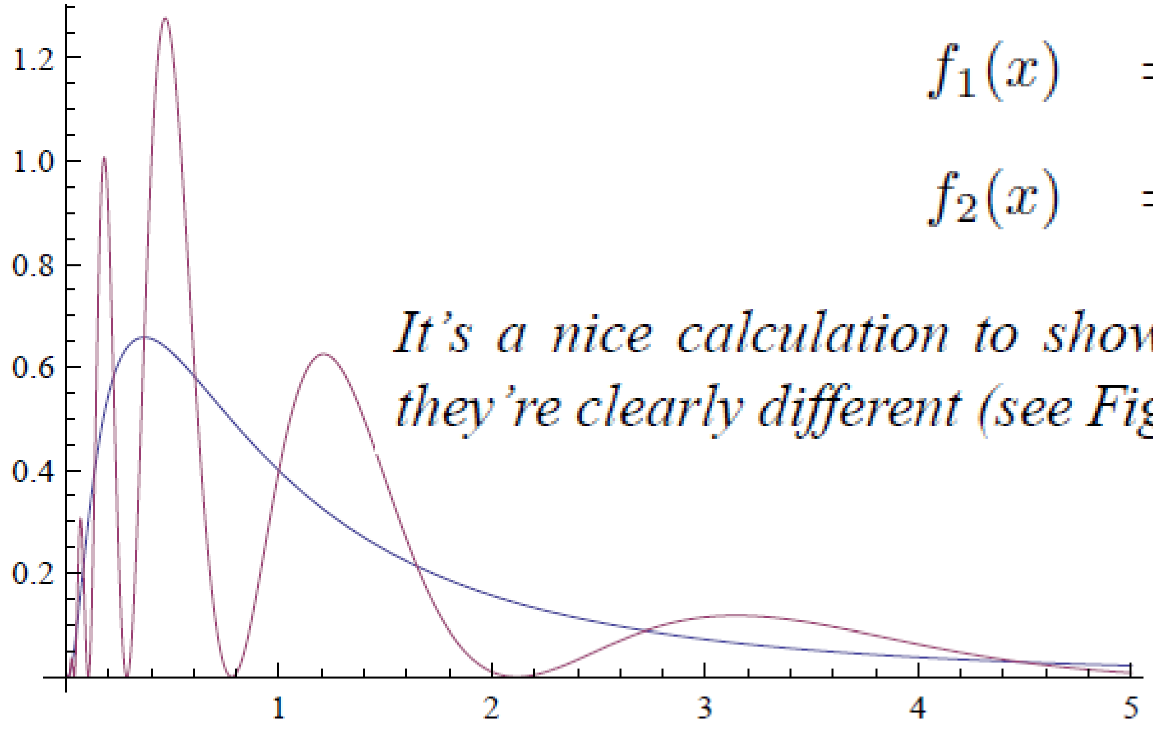


Figure 19.1: Plot of $f_1(x)$ and $f_2(x)$ from (19.2).

$$g(x) = \begin{cases} \exp(-1/x^2) & \text{if } x \neq 0 \\ 0 & \text{otherwise.} \end{cases} \qquad (19.3)$$

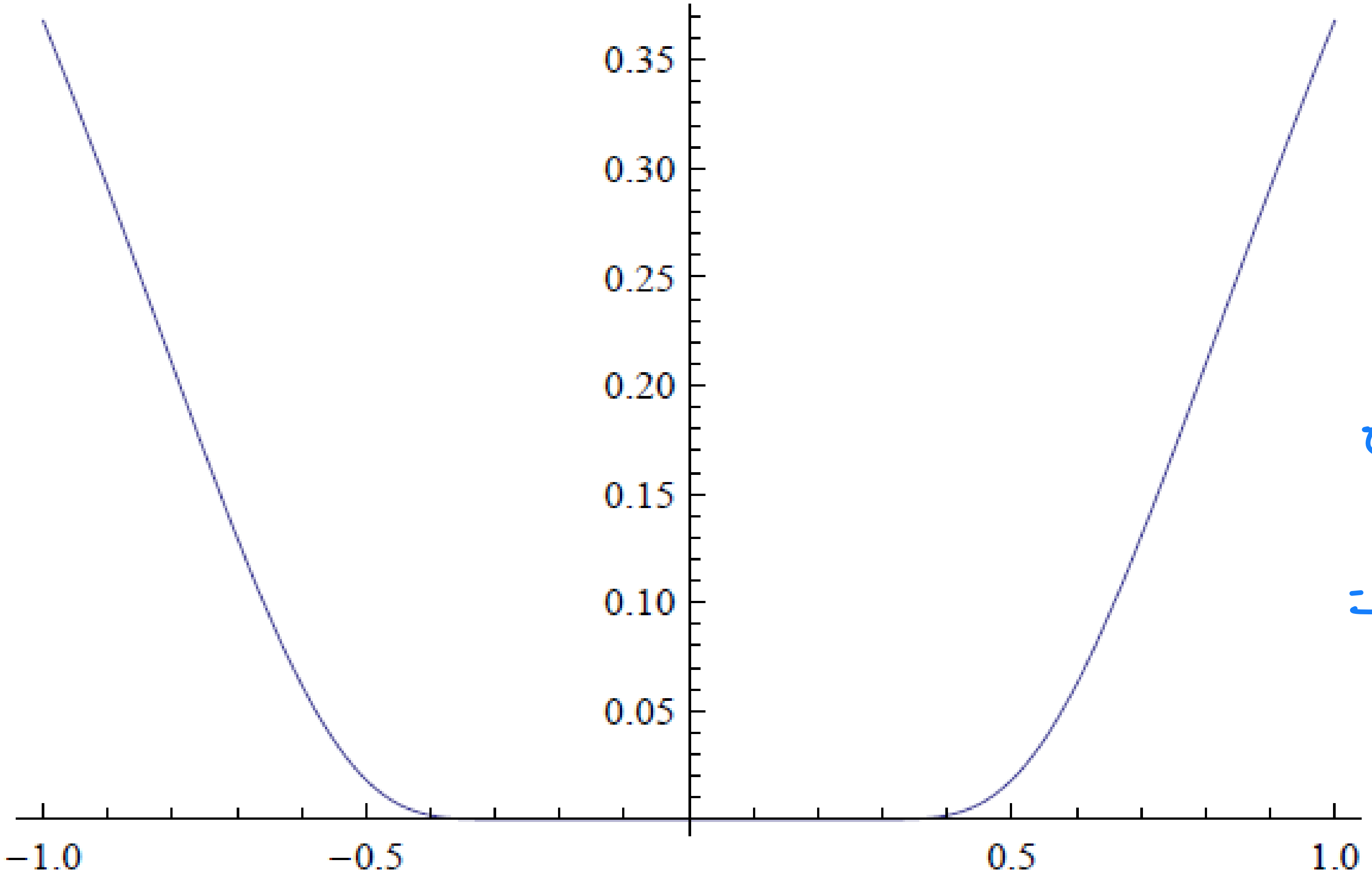


Figure 19.2: Plot of $g(x)$ from (19.3).

e^{-1/z^2}

with

$z = x + iy$

$i^2 = -1$

let $x = 0$

$y = 1/n$

$e^{-1/(i\frac{1}{n})^2}$

$= e^{1/\,1/n^2}$

$= e^{n^2} \to \infty$