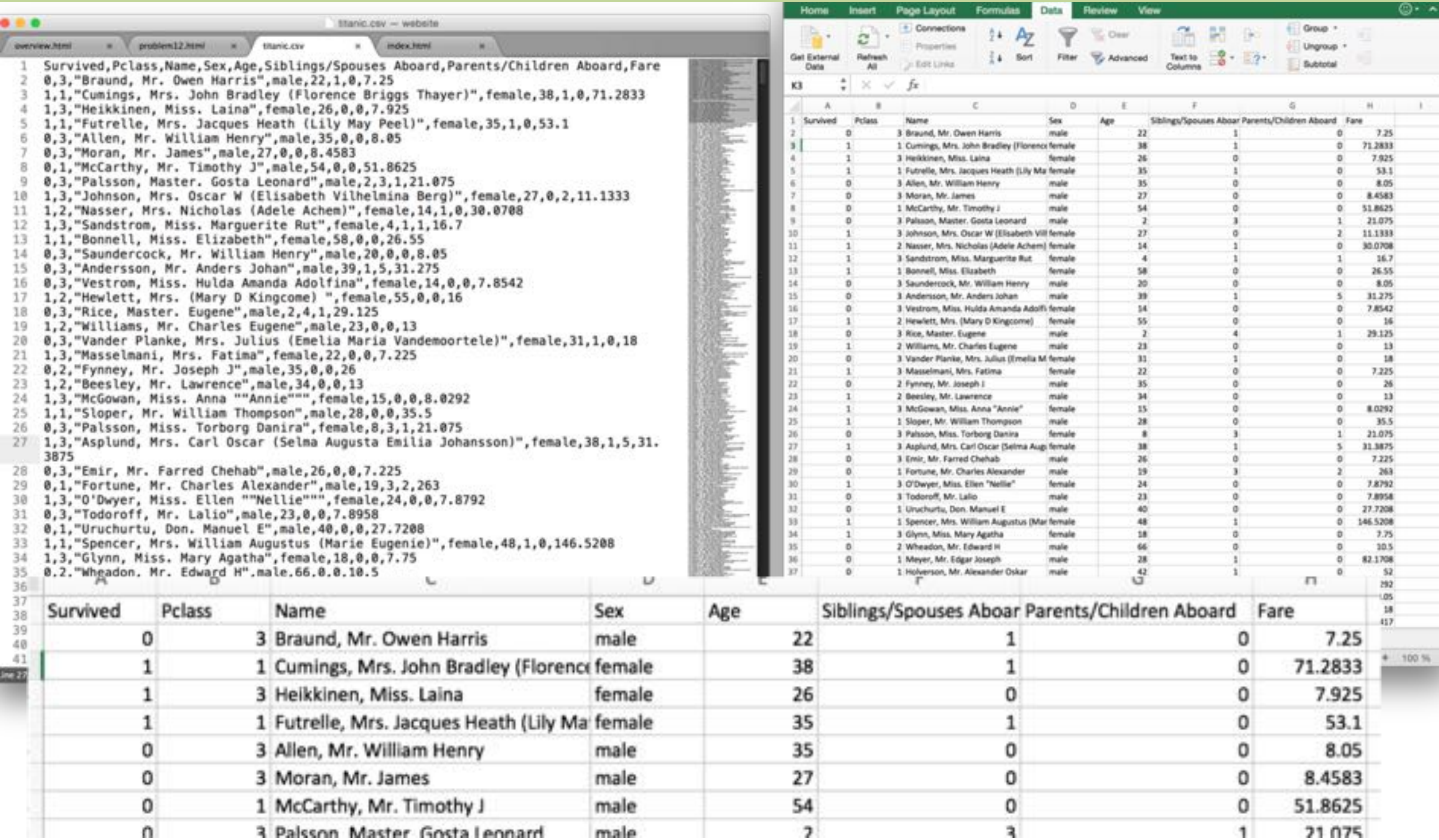
Properties of Joints

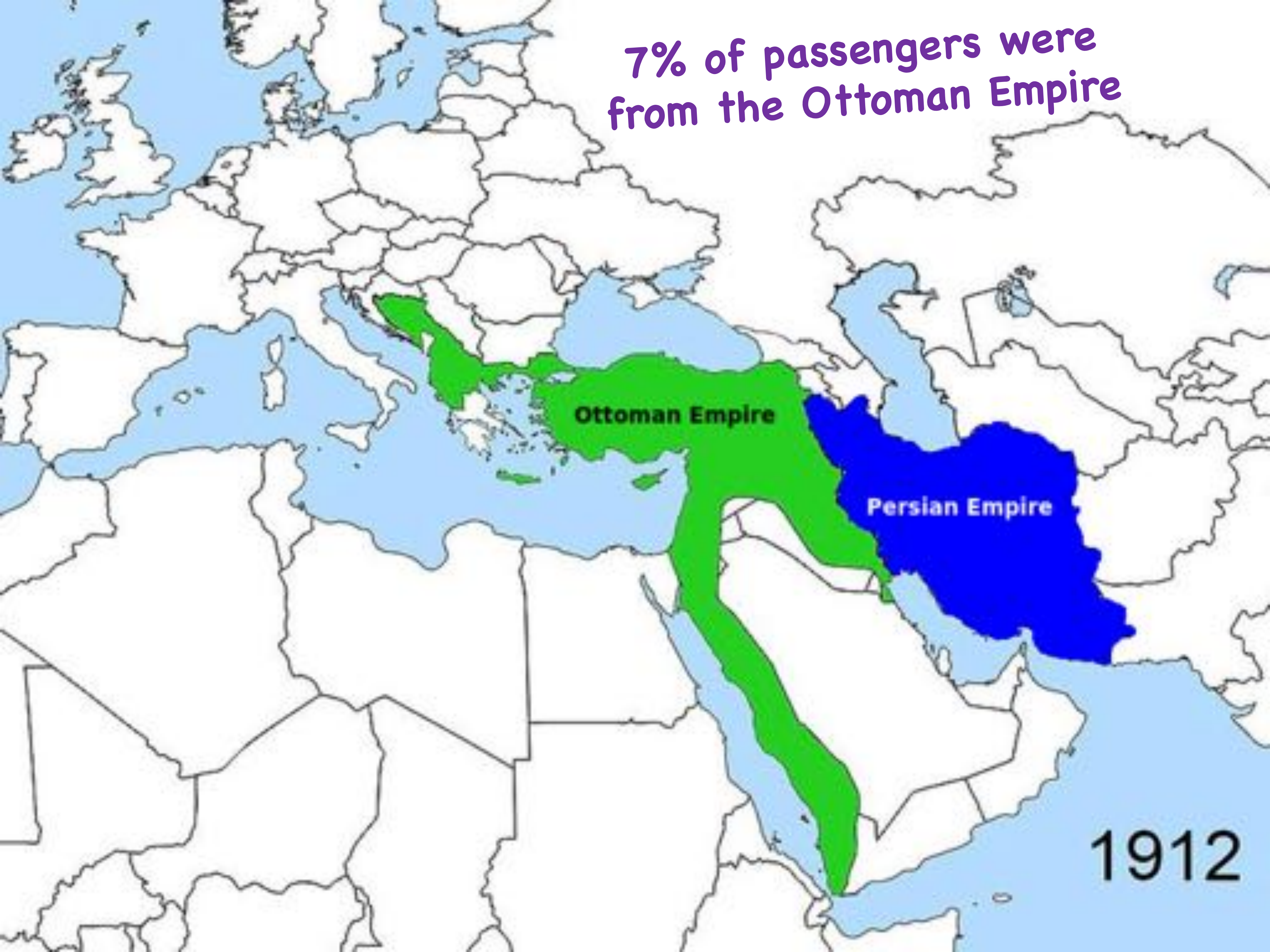
Chris Piech
CS109, Stanford University

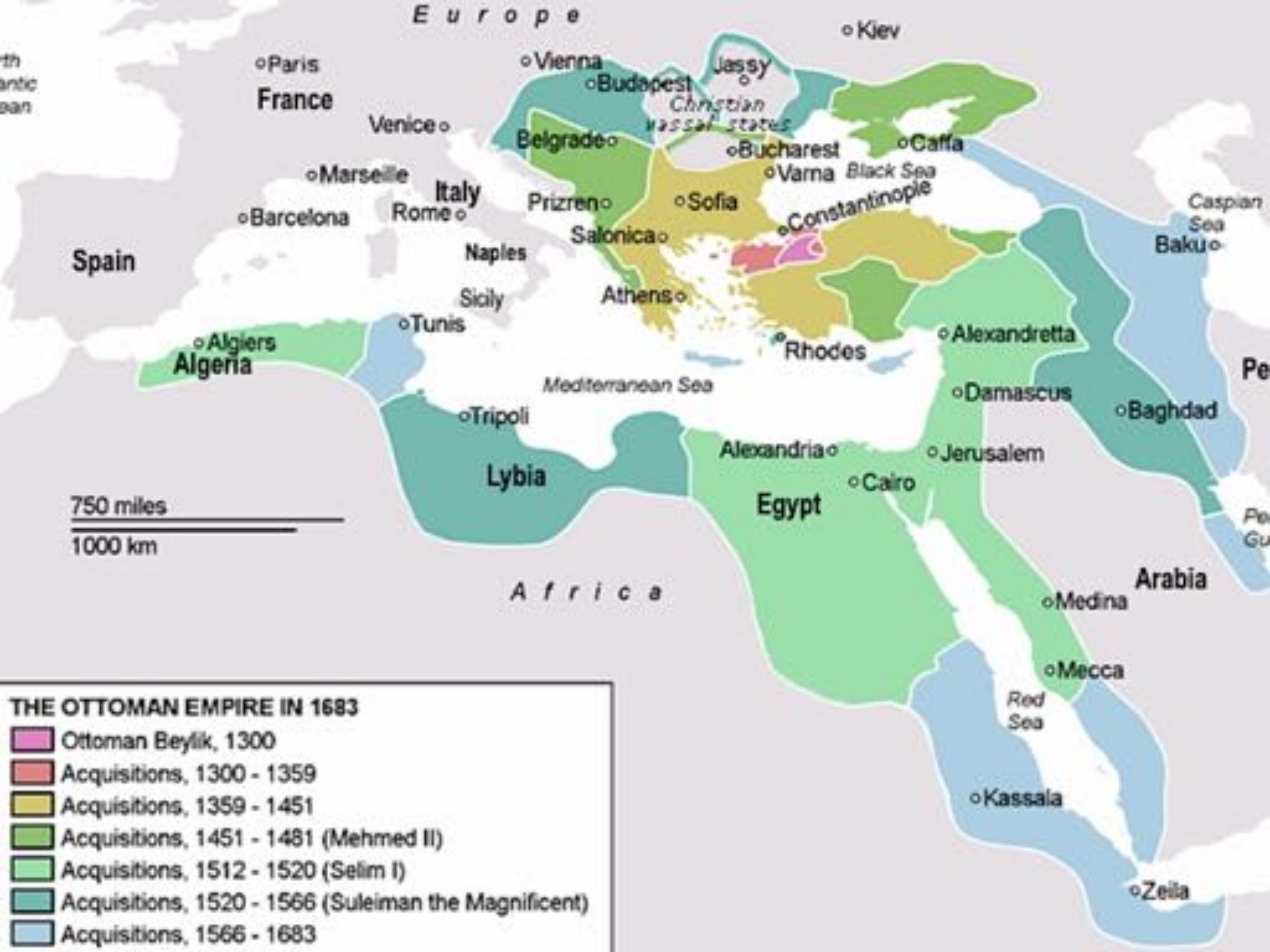
Titanic Probability



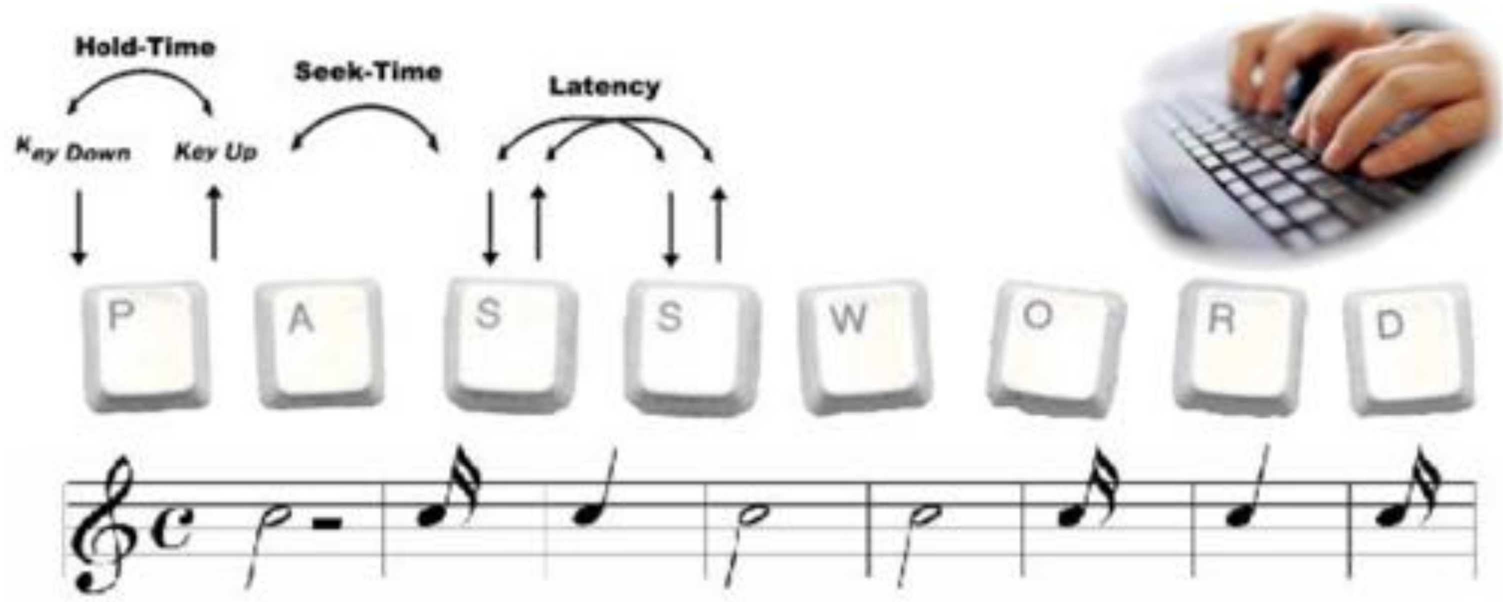

k
kaggle
problem
data
crowd
tools
models

7% of passengers were from the Ottoman Empire




Europe
Kiev
Paris
Vienna
Budapest
Jassy
Christian vassal states
France
Venice
Belgrade
Bucharest
Caffa
Marseille
Italy
Varna
Black Sea
Prizren
Sofia
Caspian Sea
Barcelona
Rome
Constantinople
Salonica
Baku
Spain
Naples
Athens
Sicily
Tunis
Algiers
Algeria
Alexandretta
Rhodes
Mediterranean Sea
Damascus
Baghdad
Tripoli
Alexandria
Jerusalem
Lybia
Cairo
Egypt
750 miles
1000 km
Arabia
Africa
Medina
Mecca
Red Sea
Kassala
Zeila
THE OTTOMAN EMPIRE IN 1683
Ottoman Beylik, 1300
Acquisitions, 1300 - 1359
Acquisitions, 1359 - 1451
Acquisitions, 1451 - 1481 (Mehmed II)
Acquisitions, 1512 - 1520 (Selim I)
Acquisitions, 1520 - 1566 (Suleiman the Magnificent)
Acquisitions, 1566 - 1683

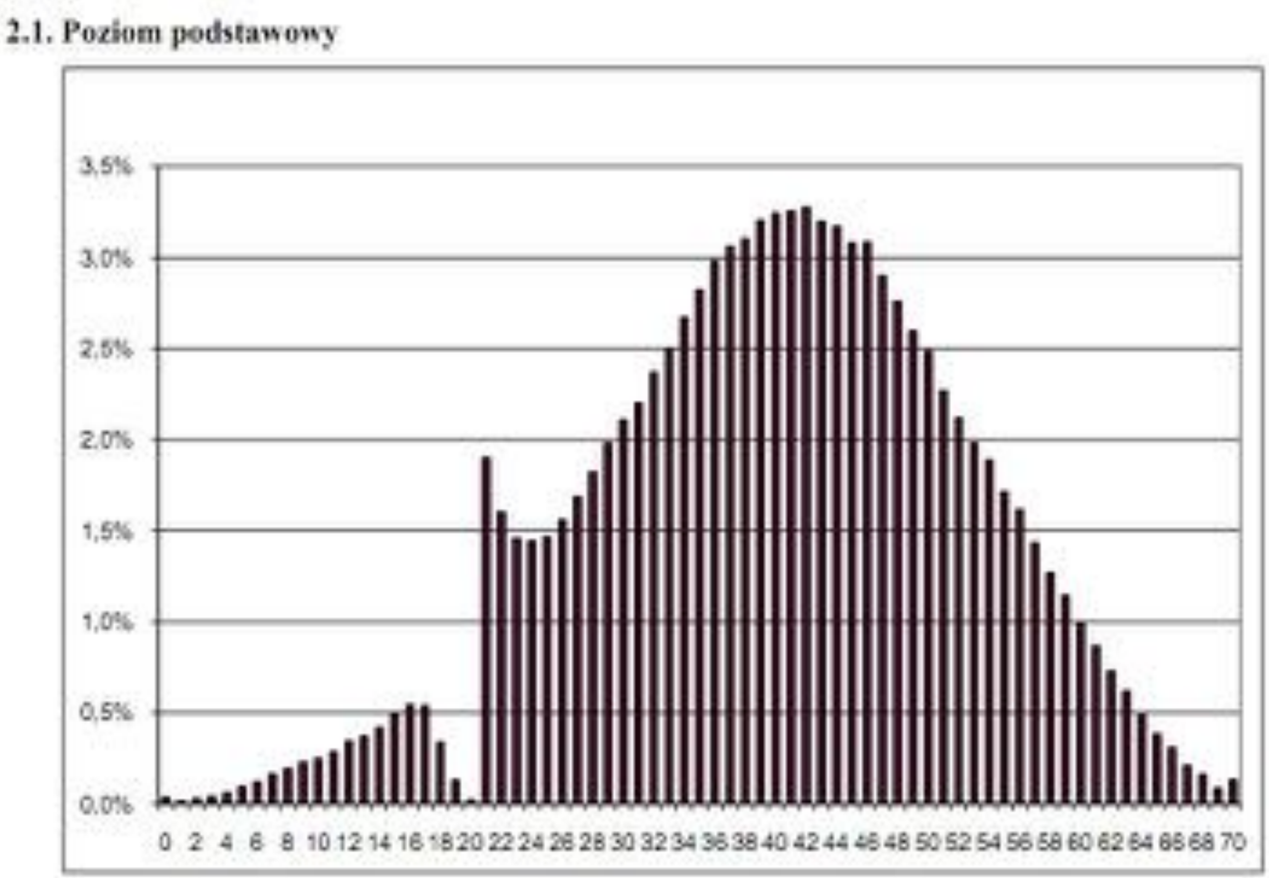
Biometric Keystrokes



Altruism?

Scores for a standardized test that students in Poland are required to pass before moving on in school

See if you can guess the minimum score to pass the test.



Wykres 1. Rozkład wyników na poziomie podstawowym

http://freakonomics.com/2011/07/07/another-case-of-teacher-cheating-or-is-it-just-altruism/comment-page-2/

Joint Random Variables

- Use a joint table, density function or CDF to solve probability question
- Think about **conditional** probabilities with joint variables (which might be continuous)
- Use and find **expectation** of random variables
- Use and find **independence** of random variables
- What happens when you **add** random variables?

Zero Sum Games

What is the probability that the Warriors win?

How do you model zero sum games?

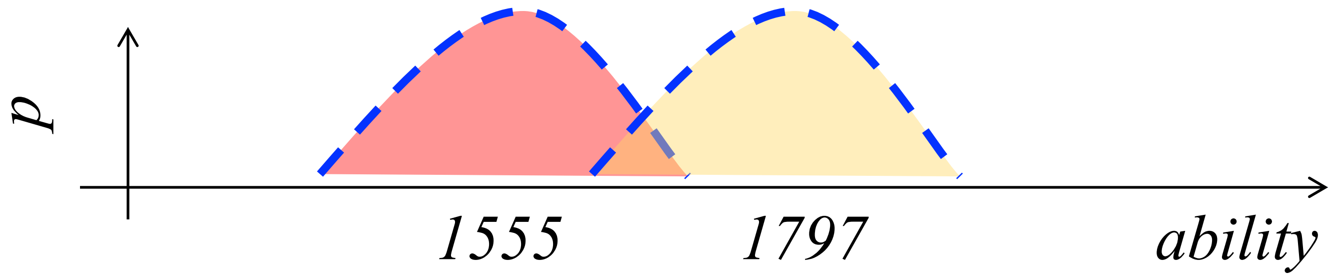
Motivating Idea: Zero Sum Games

How it works:

- Each team has an "ELO" score *S*, calculated based on their past performance.
- Each game, the team has ability $A \sim N(S, 200^2)$
- The team with the higher sampled ability wins.

Arpad Elo

$$A_B \sim \mathcal{N}(1555, 200^2) \qquad A_W \sim \mathcal{N}(1797, 200^2)$$



$$P(\text{Warriors win}) = P(A_W > A_B)$$

Motivating Idea: Zero Sum Games

$$A_W \sim \mathcal{N}(1797, 200^2)$$

$$A_B \sim \mathcal{N}(1555, 200^2)$$

$$P(\text{Warriors win}) = P(A_W > A_B)$$

$$P(\text{Warriors win}) = P(A_W - A_B > 0)$$

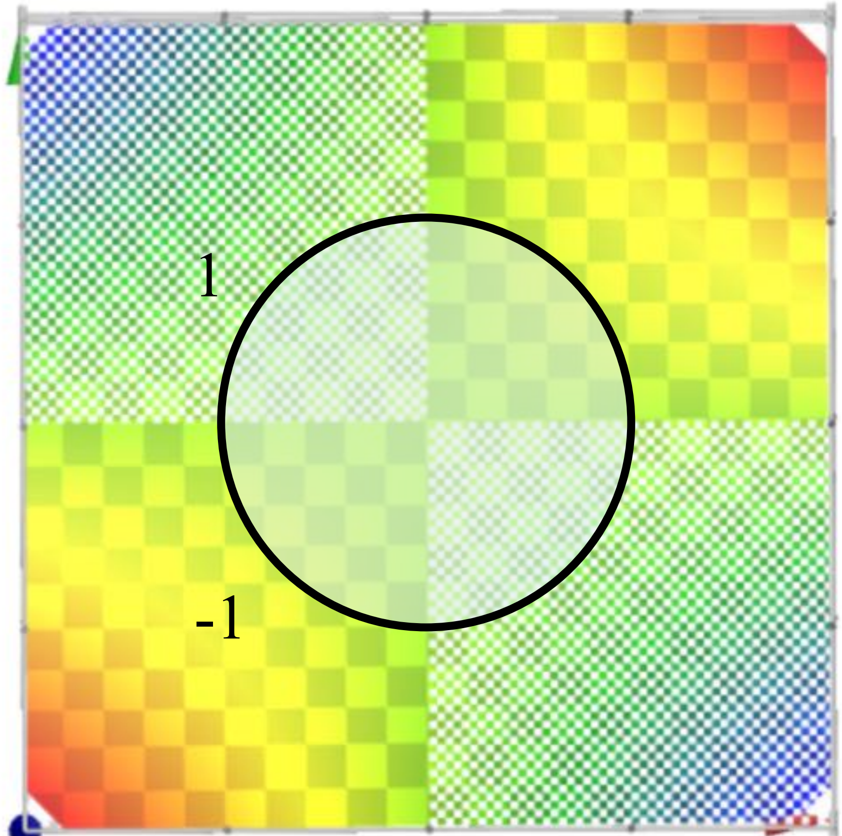
In class we solved this by sampling. But that is a bit of a "cheat" and is computationally expensive.

Sums (or subtractions) of random variables show up all the time. But we have no explicit tools for dealing with them!

Challenge: try and come up with the way to solve this by the end of class

You may have not seen this…

$$\iint\limits_{x^2+y^2<1} f_{x,y}\ dy\ dx = \int\limits_{-1}^{1} \int\limits_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} f_{x,y}\ dy\ dx$$



Expectation of Multiple RVs

Joint Expectation

$$E[X] = \sum_x x p(x)$$

- Expectation over a joint isn't nicely defined because it is not clear how to compose the multiple variables:
  - Add them? Multiply them?

- Lemma: For a function g(X,Y) we can calculate the expectation of that function:

$$E[g(X,Y)] = \sum_{x,y} g(x,y)p(x,y)$$

- Recall, this also holds for single random variables:

$$E[g(X)] = \sum_x g(x)p(x)$$

Expected Values of Sums

Big deal lemma: first stated without proof

$$E[X + Y] = E[X] + E[Y]$$

Generalized: $$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Holds regardless of dependency between X_i's

Skeptical Chris Wants a Proof!

Let g(X,Y) = [X + Y]

$$E[X+Y] = E[g(X,Y)] = \sum_{x,y} g(x,y)p(x,y)$$

What a useful lemma

By the definition of g(x,y)

$$= \sum_{x,y} [x+y]p(x,y)$$

Break that sum into parts!

$$= \sum_{x,y} xp(x,y) + \sum_{x,y} yp(x,y)$$

Change the sum of (x,y) into separate sums

$$= \sum_{x} x \sum_{y} p(x,y) + \sum_{y} y \sum_{x} p(x,y)$$

That is the definition of marginal probability

$$= \sum_{x} xp(x) + \sum_{y} yp(y)$$

That is the definition of expectation

$$= E[X] + E[Y]$$

Independence and Random Variables

Independent Discrete Variables

- Two discrete random variables X and Y are called **independent** if:

$$p(x,y) = p_X(x)\,p_Y(y) \quad \text{for all } x, y$$

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

- Intuitively: knowing the value of X tells us nothing about the distribution of Y (and vice versa)
  - If two variables are **not** independent, they are called ***dependent***
- Similar conceptually to independent *events*, but we are dealing with multiple ***variables***
  - Keep your events and variables distinct (and clear)!

Is Year Independent of Lunch?

Joint Probability Table					
	Dining Hall	Eating Club	Cafe	Self-made	**Marginal Year**
Freshman	0.03	0.00	0.02	0.00	0.05
Sophomore	0.50	0.15	0.03	0.03	0.68
Junior	0.08	0.02	0.02	0.02	0.12
Senior	0.02	0.05	0.01	0.01	0.08
5+	0.02	0.01	0.05	0.05	0.07
Marginal Status	0.65	0.22	0.12	0.11	

For all values of Year, Status:

$$\underbrace{P(\text{Year} = y, \text{Lunch} = s)}_{0.50} = \underbrace{P(\text{Year} = y)}_{0.68}\underbrace{P(\text{Lunch} = s)}_{0.65}$$

Yes!

Is Year Independent of Lunch?

Joint Probability Table					
	Dining Hall	Eating Club	Cafe	Self-made	**Marginal Year**
Freshman	0.03	0.00	0.02	0.00	0.05
Sophomore	0.50	0.15	0.03	0.03	0.68
Junior	0.08	0.02	0.02	0.02	0.12
Senior	0.02	0.05	0.01	0.01	0.08
5+	0.02	0.01	0.05	0.05	0.07
Marginal Status	0.65	0.22	0.12	0.11	

For all values of Year, Status:

$$\underbrace{P(\text{Year} = y, \text{Lunch} = s)}_{0.03} = \underbrace{\underbrace{P(\text{Year} = y)}_{0.68}\underbrace{P(\text{Lunch} = s)}_{0.12}}_{0.08}$$

No ☹

Aside: Butterfly Effect

Coin Flips

- Flip coin with probability *p* of "heads"
  - Flip coin a total of *n* + *m* times
  - Let X = number of heads in first *n* flips
  - Let Y = number of heads in next *m* flips

$$P(X = x, Y = y) = \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y}$$

$$= P(X = x)P(Y = y)$$

  - X and Y are independent
  - Let Z = number of total heads in *n* + *m* flips
  - Are X and Z independent?
    - What if you are told Z = 0?

Independent Continuous Variables

- Two continuous random variables X and Y are called **independent** if:

  P(X ≤ *a*, Y ≤ *b*) = P(X ≤ *a*) P(Y ≤ *b*) for any *a*, *b*

- Equivalently:

$$F_{X,Y}(a,b) = F_X(a)F_Y(b) \quad \text{for all } a,b$$

$$f_{X,Y}(a,b) = f_X(a)f_Y(b) \quad \text{for all } a,b$$

- More generally, joint density factors separately:

$$f_{X,Y}(x,y) = h(x)g(y) \quad \text{where } -\infty < x, y < \infty$$

Is the Blur Distribution Independent?

In image processing, a Gaussian blur is the result of blurring an image by a Gaussian function. It is a widely used effect in graphics software, typically to reduce image noise.

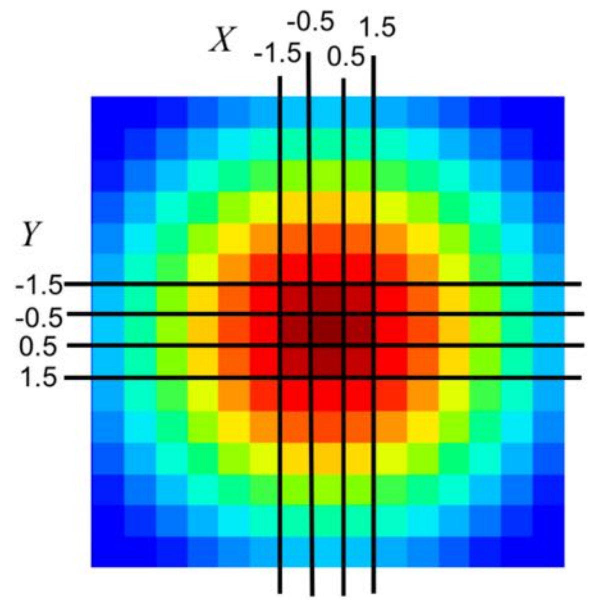
Gaussian blurring with StDev = 3, is based on a joint probability distribution:

Joint PDF

$$f_{X,Y}(x,y) = \frac{1}{2\pi \cdot 3^2} e^{-\frac{x^2+y^2}{2\cdot 3^2}}$$

Joint CDF

$$F_{X,Y}(x,y) = \Phi\left(\frac{x}{3}\right) \cdot \Phi\left(\frac{y}{3}\right)$$



Used to generate this weight matrix

Pop Quiz (just kidding)

- Consider joint density function of X and Y:

$$f_{X,Y}(x,y) = 6e^{-3x}e^{-2y} \quad \text{for } 0 < x, y < \infty$$

  - Are X and Y independent? Yes!

Let $h(x) = 3e^{-3x}$ and $g(y) = 2e^{-2y}$, so $f_{X,Y}(x,y) = h(x)g(y)$

- Consider joint density function of X and Y:

$$f_{X,Y}(x,y) = 4xy \quad \text{for } 0 < x, y < 1$$

  - Are X and Y independent? Yes!

Let $h(x) = 2x$ and $g(y) = 2y$, so $f_{X,Y}(x,y) = h(x)g(y)$

  - Now add constraint that: $0 < (x + y) < 1$
  - Are X and Y independent? No!
    - Cannot capture constraint on x + y in factorization!

What happens when you add random variables?

Sum of Independent Binomials

- Let X and Y be independent binomials with the same value for *p*:
  - $X \sim Bin(n_1, p)$ and $Y \sim Bin(n_2, p)$
  - $X + Y \sim Bin(n_1 + n_2, p)$

- Intuition:
  - X has n_1 trials and Y has n_2 trials
    - Each trial has same "success" probability p
  - Define Z to be $n_1 + n_2$ trials, each with success prob. p
  - $Z \sim Bin(n_1 + n_2, p)$, and also $Z = X + Y$

If only it were always that simple

The Insight to Convolution

Imagine a game
where each player *independently* scores between 0 and 100 points:

Let X be the amount of points you score.
Let Y be the amount of points your opponent scores.
Let's say you know P(X = x) and P(Y = y).

What is the probability of a tie?

$$P(\text{tie}) = \sum_{i=0}^{100} P(X = i, Y = i)$$

$$= \sum_{I=0}^{100} P(X = i)P(Y = i)$$

The Insight to Convolution Proofs

$$P(X + Y = n)?$$

What is the probability that X + Y = n?

X	Y	k	
0	n	0	$P(X = 0, Y = n)$
1	n - 1	1	$P(X = 1, Y = n-1)$
2	n - 2	2	$P(X = 2, Y = n-2)$
	• • •		
n	0	n	$P(X = n, Y = 0)$

The Insight to Convolution Proofs

$$P(X + Y = n)?$$

What is the probability that X + Y = n?

Since this is the OR or mutually exclusive events

$$P(X+Y=n) = \sum_{k=0}^{n} P(X=k, Y=n-k)$$

If the random variables are independent

$$= \sum_{k=0}^{n} P(X=k)P(Y=n-k)$$

Sum of Independent Poissons

Recall the Binomial Theorem

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Sum of Independent Poissons

- Let X and Y be independent random variables
  - X ~ Poi(λ_1) and Y ~ Poi(λ_2)
  - X + Y ~ Poi($\lambda_1 + \lambda_2$)
- Proof: (just for reference)
  - Rewrite (X + Y = *n*) as (X = *k*, Y = *n* – *k*) where $0 \le k \le n$

$$P(X+Y=n)=\sum_{k=0}^{n}P(X=k,Y=n-k)=\sum_{k=0}^{n}P(X=k)P(Y=n-k)$$

$$=\sum_{k=0}^{n}e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}=e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!}=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}$$

  - Noting Binomial theorem: $(\lambda_1+\lambda_2)^n=\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}$
  - $P(X+Y=n)=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\left(\lambda_1+\lambda_2\right)^n$ so, X + Y = n ~ Poi($\lambda_1 + \lambda_2$)

Reference: Sum of Independent RVs

- Let X and Y be independent Binomial RVs
  - $X \sim \text{Bin}(n_1, p)$ and $Y \sim \text{Bin}(n_2, p)$
  - $X + Y \sim \text{Bin}(n_1 + n_2, p)$
  - More generally, let $X_i \sim \text{Bin}(n_i, p)$ for $1 \le i \le N$, then

$$\left(\sum_{i=1}^{N} X_i\right) \sim \text{Bin}\left(\sum_{i=1}^{N} n_i,\ p\right)$$

- Let X and Y be independent Poisson RVs
  - $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$
  - $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$
  - More generally, let $X_i \sim \text{Poi}(\lambda_i)$ for $1 \le i \le N$, then

$$\left(\sum_{i=1}^{N} X_i\right) \sim \text{Poi}\left(\sum_{i=1}^{N} \lambda_i\right)$$

Convolution of Probability Distributions

We talked about sum of Binomial and Poisson…who's missing from this party?

Uniform.

Summation: not just for the 1%

Dance, Dance Convolution

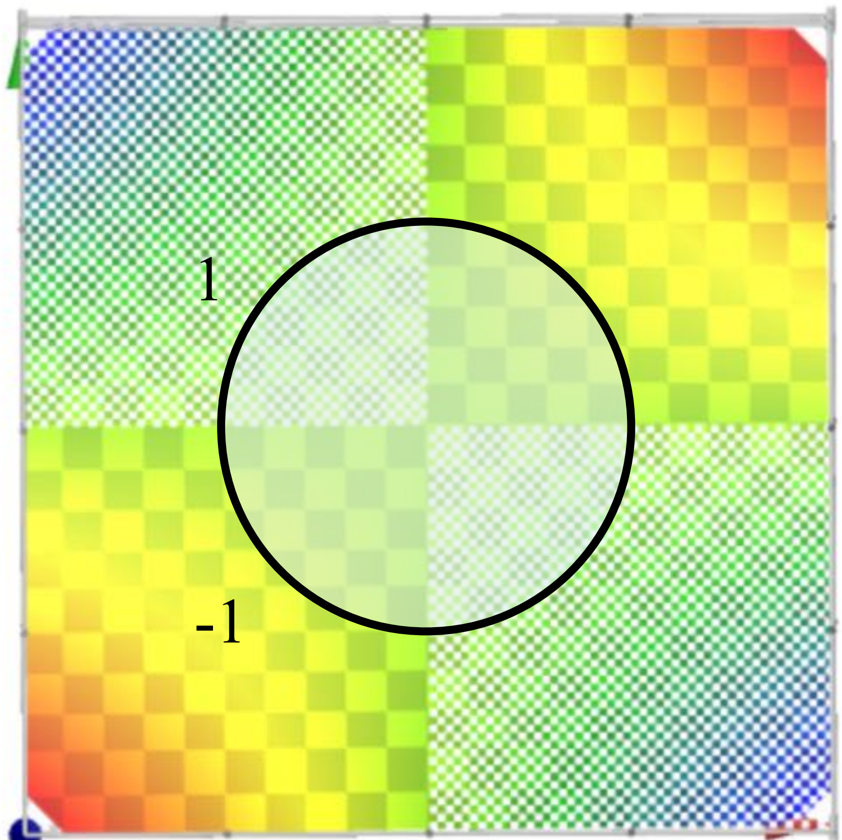
- Let X and Y be independent random variables
  - Probability Density Function (PDF) of X + Y:

$$f_{X+Y}(a) = \int_{y=-\infty}^{\infty} f_X(a-y)\, f_Y(y)\, dy$$

  - In discrete case, replace $\int_{y=-\infty}^{\infty}$ with $\sum_y$, and $f(y)$ with $p(y)$

Integration with Constraint

$$\iint\limits_{x^2+y^2<1} f_{x,y}\; dy\; dx = \int\limits_{-1}^{1} \int\limits_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} f_{x,y}\; dy\; dx$$



Dance, Dance Convolution

- Let X and Y be independent random variables
  - Cumulative Distribution Function (CDF) of X + Y:

$$
\begin{aligned}
F_{X+Y}(a) &= P(X+Y \le a) \\
&= \iint_{x+y\le a} f_X(x) f_Y(y)\, dx\, dy = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{a-y} f_X(x)\, dx\, f_Y(y)\, dy \\
&= \int_{y=-\infty}^{\infty} F_X(a-y)\, f_Y(y)\, dy
\end{aligned}
$$

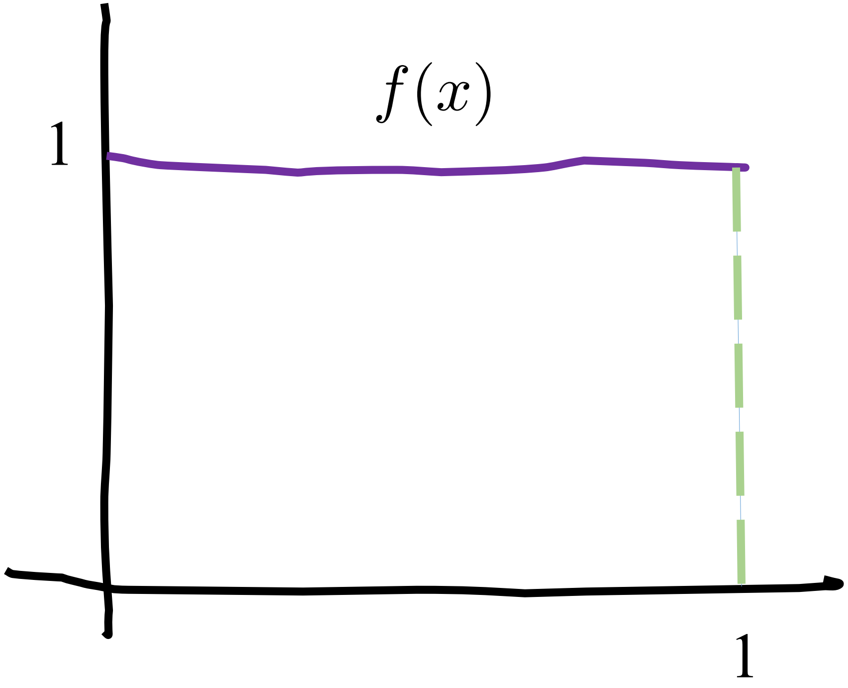
CDF of X

PDF of Y

  - In discrete case, replace $\int_{y=-\infty}^{\infty}$ with $\sum_{y}$, and $f(y)$ with $p(y)$

Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \leq x \leq 1$



For both X and Y

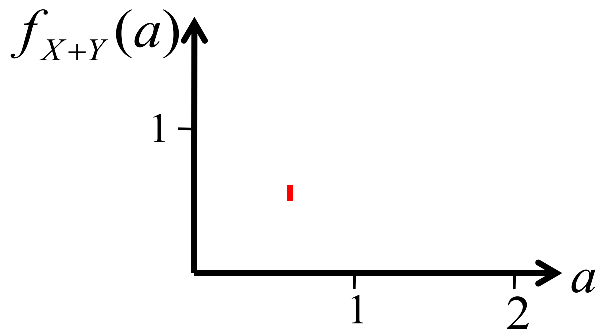
Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \leq x \leq 1$
  - What is PDF of X + Y?

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

When $a = 0.5$:

$$\begin{aligned} f_{X+Y}(0.5) &= \int_{y=?}^{y=?} f_X(0.5-y)dy \\ &= \int_{0}^{0.5} f_X(0.5-y)dy \\ &= \int_{0}^{0.5} 1dy \\ &= 0.5 \end{aligned}$$



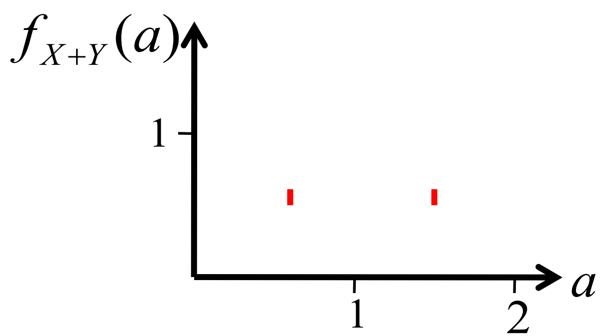
Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$
  - What is PDF of X + Y?

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

When $a = 1.5$:

$$\begin{aligned} f_{X+Y}(1.5) &= \int_{y=?}^{y=?} f_X(1.5-y)dy \\ &= \int_{0.5}^{1} f_X(1.5-y)dy \\ &= \int_{0.5}^{1} 1dy \\ &= 0.5 \end{aligned}$$



Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$
  - What is PDF of X + Y?

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

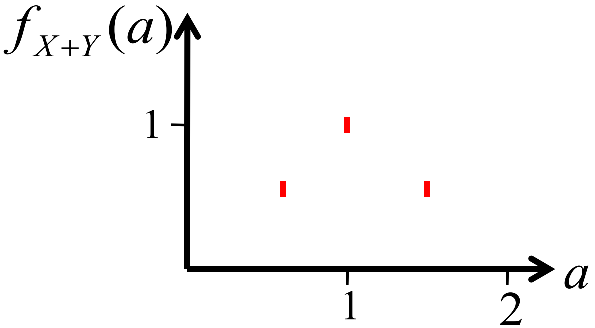
When $a = 1$:

$$f_{X+Y}(1) = \int_{y=?}^{y=?} f_X(1-y)dy$$

$$= \int_0^1 f_X(1-y)dy$$

$$= \int_0^1 1dy$$

$$= 1$$



Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$
  - What is PDF of X + Y?

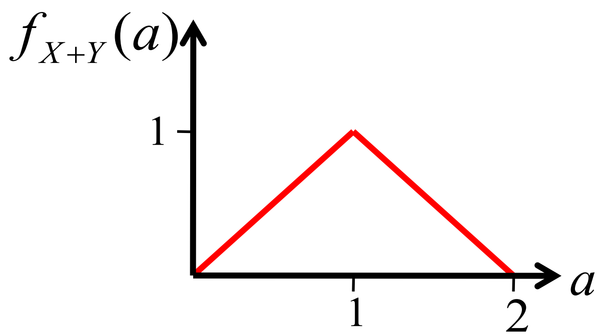
$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

  - When $0 \le a \le 1$ and $0 \le y \le a$, $0 \le a-y \le 1$ → $f_X(a-y) = 1$

$$f_{X+Y}(a) = \int_{y=0}^{a} dy = a$$

  - When $1 \le a \le 2$ and $a-1 \le y \le 1$, $0 \le a-y \le 1$ → $f_X(a-y) = 1$

$$f_{X+Y}(a) = \int_{y=a-1}^{1} dy = 2-a$$



  - Combining: $f_{X+Y}(a) = \begin{cases} a & 0 \le a \le 1 \\ 2-a & 1 < a \le 2 \\ 0 & \text{otherwise} \end{cases}$

Sum of Independent Normals

- Let X and Y be independent random variables
  - $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$
  - $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

- Generally, have *n* independent random variables $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, ..., n$:

$$\left(\sum_{i=1}^{n} X_i\right) \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Virus Infections

- Say you are working with the WHO to plan a response to a the initial conditions of a virus:
  - Two exposed groups
  - P1: 50 people, each independently infected with $p = 0.1$
  - P2: 100 people, each independently infected with $p = 0.4$
  - Question: Probability of more than 40 infections?

Sanity check: Should we use the Binomial Sum-of-RVs shortcut?

A. YES!
B. NO!
C. Other/none/more

Virus Infections

- Say you are working with the WHO to plan a response to a the initial conditions of a virus:
  - Two exposed groups
  - P1: 50 people, each independently infected with $p = 0.1$
  - P2: 100 people, each independently infected with $p = 0.4$
  - A = # infected in P1 A ~ Bin(50, 0.1) ≈ X ~ N(5, 4.5)
  - B = # infected in P2 B ~ Bin(100, 0.4) ≈ Y ~ N(40, 24)
  - What is P(≥ 40 people infected)?
  - P(A + B ≥ 40) ≈ P(X + Y ≥ 39.5)
  - X + Y = W ~ N(5 + 40 = 45, 4.5 + 24 = 28.5)

$$P(W \geq 39.5) = P\left(\frac{W-45}{\sqrt{28.5}} > \frac{39.5-45}{\sqrt{28.5}}\right) = 1 - \Phi(-1.03) \approx 0.8485$$

Linear Transform

$$X \sim N(\mu, \sigma^2)$$

$$Y = X + X = 2 \cdot X$$

$$Y \sim N(2\mu, 4\sigma^2)$$

$$Y = X + X = 2 \cdot X$$

$$X + X \sim N(\mu + \mu, \sigma^2 + \sigma^2)$$

$$Y \sim N(2\mu, 2\sigma^2)$$

X is not independent of X

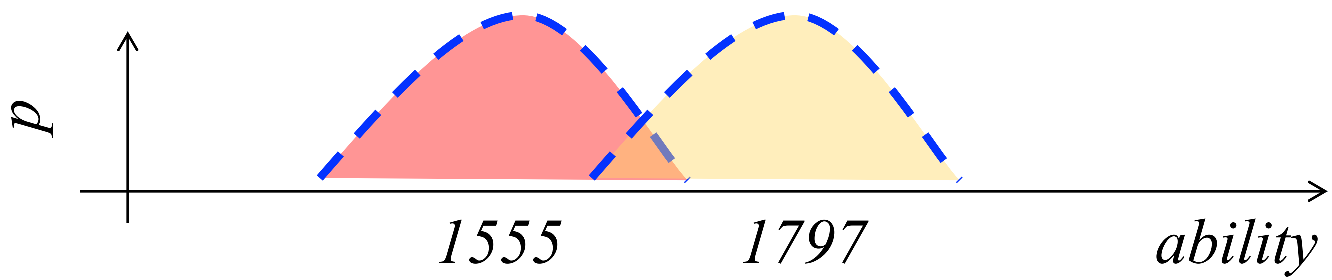
Motivating Idea: Zero Sum Games

How it works:

- Each team has an "ELO" score *S*, calculated based on their past performance.
- Each game, the team has ability $A \sim N(S, 200^2)$
- The team with the higher sampled ability wins.

Arpad Elo

$$A_B \sim \mathcal{N}(1555, 200^2) \qquad A_W \sim \mathcal{N}(1797, 200^2)$$



$$P(\text{Warriors win}) = P(A_W > A_B)$$

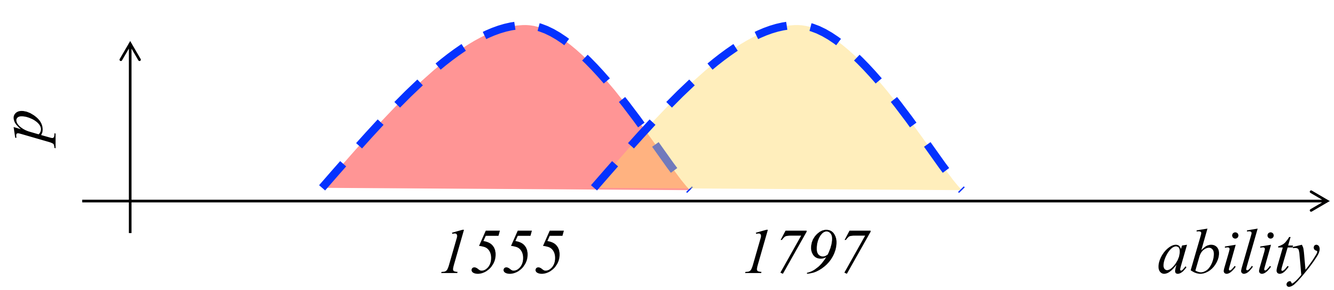
Motivating Idea: Zero Sum Games

$$A_B \sim \mathcal{N}(1555, 200^2) \qquad A_W \sim \mathcal{N}(1797, 200^2)$$

$$\begin{aligned} P(\text{Warriors win}) &= P(A_W > A_B) \\ &= P(A_W - A_B > 0) \end{aligned}$$

$$\begin{aligned} D &= A_W - A_B \\ D &\sim N(\mu = 1795 - 1555, \sigma_2 = 2 \cdot 200^2) \\ &\sim N(\mu = 240, \sigma_2 = 283) \end{aligned}$$

$$P(D > 0) = F_D(0) = 1 - \Phi\left(\frac{0 - 240}{283}\right) \approx 0.65$$



That’s all folks!