Properties of Joints:
Expectation, Independence, Convolution

Noah Arthurs
CS109, Stanford University

Review

Discrete Conditional Distributions

- Recall that for *events* E and F:

$$P(E \mid F) = \frac{P(EF)}{P(F)} \quad \text{where } P(F) > 0$$

- Now, have X and Y as **discrete** random variables
  - **Conditional PMF** of X given Y:

$$P_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

Different notations,
same idea.

Continuous Conditional Distributions

Let X and Y be continuous random variables

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Epsilons are just for derivation

$$f_{X|Y}(x|y) \cdot \epsilon_x = \frac{f_{X,Y}(x, y) \cdot \epsilon_x \cdot \epsilon_y}{f_Y(y) \cdot \epsilon_y}$$

You can skip to this version

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Mixing Discrete and Continuous

Let N be a discrete random variable

Discrete X

$$P(X = x|N = n) = \frac{P(N = n|X = x)P(X = x)}{P(N = n)}$$

$$P_{X|N}(x|n) = \frac{P_{N|X}(n|x)P_X(x)}{P_N(n)}$$

Continuous X

$$f_{X|N}(x|n) \cdot \epsilon_x = \frac{P_{N|X}(n|x)f_X(x) \cdot \epsilon_x}{P_N(n)}$$

$$f_{X|N}(x|n) = \frac{P_{N|X}(n|x)f_X(x)}{P_N(n)}$$

All the Bayes Belong to Us

M,N are discrete. X, Y are continuous

OG Bayes

$$p_{M|N}(m|n) = \frac{P_{N|M}(n|m)p_M(m)}{p_N(n)}$$

Mix Bayes #1

$$f_{X|N}(x|n) = \frac{P_{N|X}(n|x)f_X(x)}{P_N(n)}$$

Mix Bayes #2

$$p_{N|X}(n|x) = \frac{f_{X|N}(x|n)p_N(n)}{f_X(x)}$$

All Continuous

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

Warmup: Bayes Revisited

Posterior belief

Likelihood of evidence

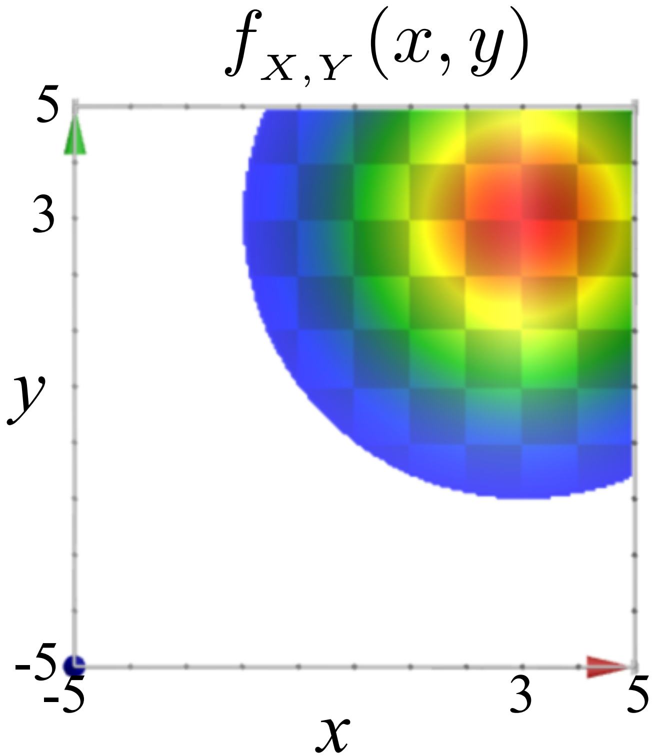
Prior belief

$$P(B|E) = \frac{P(E|B) \quad P(B)}{P(E)}$$

Normalization constant

Tracking in 2D Space: Prior

Prior belief: $$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{[(x-\mu_x)^2+(y-\mu_y)^2]}{2\cdot\sigma^2}}$$



Satellite is at (0, 0)

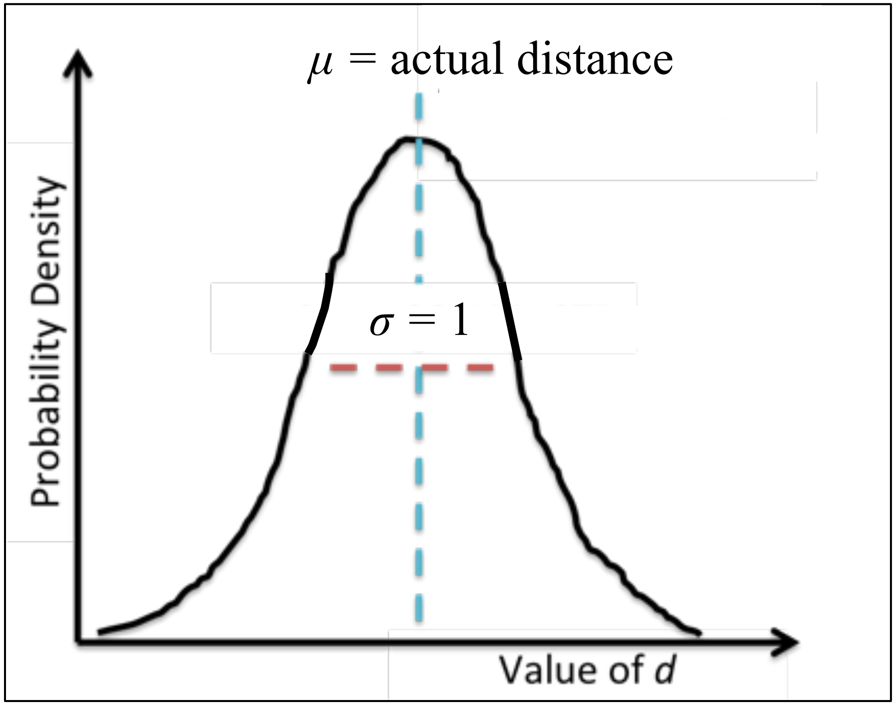
$$\mu_x = 3$$
$$\mu_y = 3$$
$$\sigma = 2$$

Prior belief with K: $$f_{X,Y}(x,y) = K \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}$$

Tracking in 2D Space: Observation!

Observe a ping of the object that is distance D away from satellite!

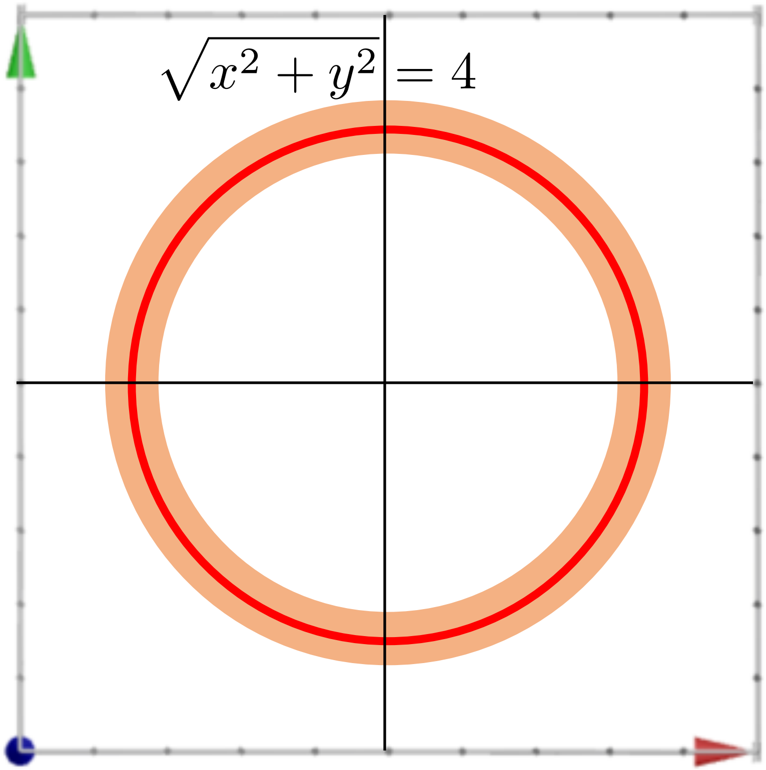
$$D|X, Y \sim N(\mu = \sqrt{x^2 + y^2}, \sigma^2 = 1)$$

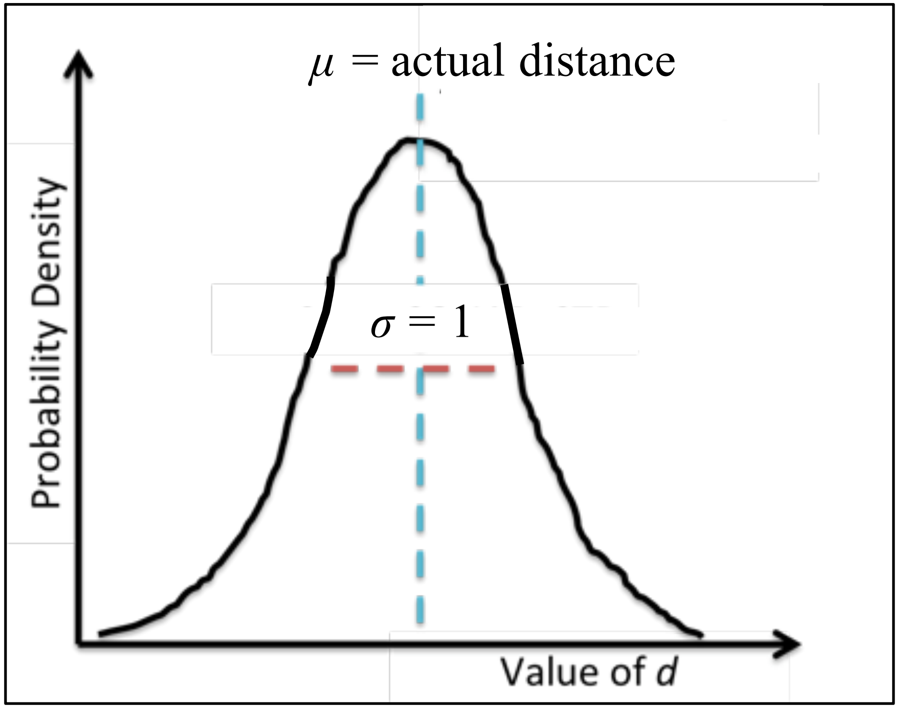


Know that the distance of a ping is normal with respect to the true distance.

Tracking in 2D Space: Observation!

Observe a ping of the object that is distance $D = 4$ away!





Know that the distance of a ping is normal with respect to the true distance

Tracking in 2D Space: New Belief

$$f(X=x,Y=y|D=4) = \frac{f(D=4|X=x,Y=y)\cdot f(X=x,Y=y)}{f(D=4)}$$

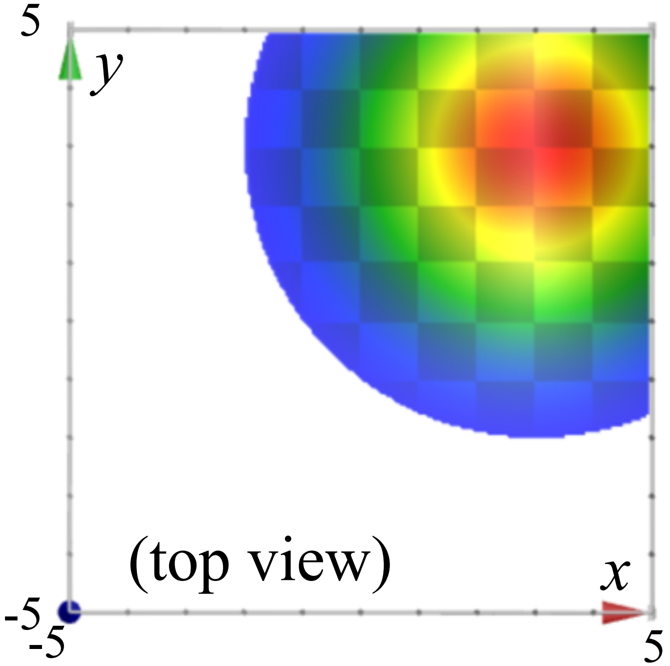
$$= \frac{K_1 \cdot e^{-\frac{[4-\sqrt{x^2+y^2})^2]}{2}} \cdot K_2 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}}{f(D=4)}$$

$$= \frac{K_3 \cdot e^{-\left[\frac{[4-\sqrt{x^2+y^2})^2]}{2}+\frac{[(x-3)^2+(y-3)^2]}{8}\right]}}{f(D=4)}$$

$$= K_4 \cdot e^{-\left[\frac{(4-\sqrt{x^2+y^2})^2}{2}+\frac{[(x-3)^2+(y-3)^2]}{8}\right]}$$

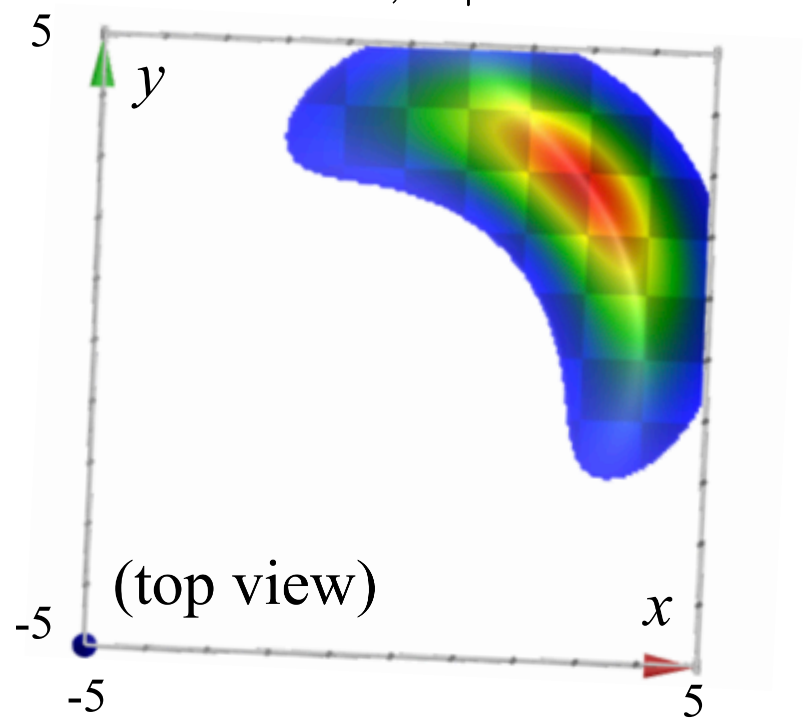
Tracking in 2D Space: Posterior

$f_{X,Y}$



Prior

$f_{X,Y|D}$



Posterior

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:

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

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:

$$P(\text{Year} = y, \text{Lunch} = s) = P(\text{Year} = y)P(\text{Lunch} = s)$$

0.50 0.68 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:

$$P(\text{Year} = y, \text{Lunch} = s) = P(\text{Year} = y)P(\text{Lunch} = s)$$

0.03   0.68   0.12

0.08

No ☹

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:

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

I owe you a (better) proof

Theorem: X and Y are independent if there exist h and g such that $f_{X,Y}(x,y) = h(x)g(y)$.

Proof: Let such X, Y, h, g exist.

$$\int_x f_{X,Y}(x,y)dx = \int_x h(x)g(y)dx \rightarrow f_Y(y) = g(y)\int_x h(x)dx = c_1 g(y)$$

$$\int_y f_{X,Y}(x,y)dy = \int_y h(x)g(y)dy \rightarrow f_X(x) = h(x)\int_y g(y)dy = c_2 h(x)$$

$$\iint_{x,y} f_{X,Y}(x,y)dydx = \iint_{x,y} \frac{f_X(x)f_Y(y)}{c_1 c_2} dxdy = \frac{1}{c_1 c_2}\int_x f_X(x)dx \int_y f_Y(y)dy$$

$$\rightarrow 1 = \frac{1}{c_1 c_2} * 1 * 1 \rightarrow c_1 c_2 = 1$$

$$f_{X,Y}(x,y) = h(x)g(y) = \frac{f_X(x)f_Y(y)}{c_1 c_2} = f_X(x)f_Y(y) \blacksquare$$

End Review

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 multiple RVS
- ☑ Use and find **independence** of multiple RVS
- ☑ 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?

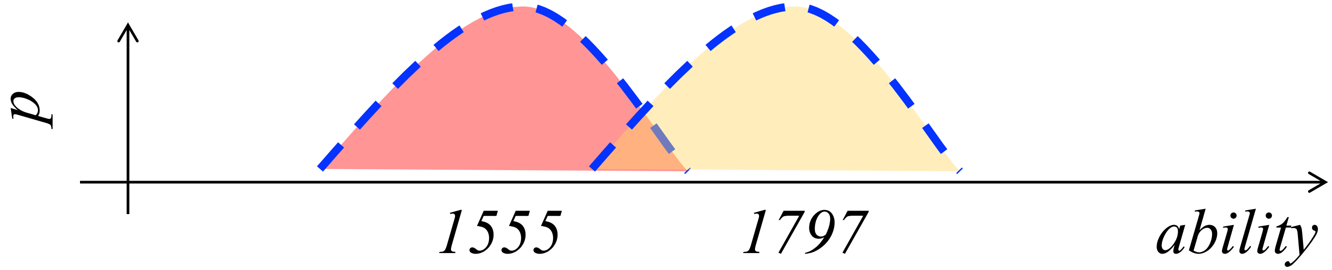
Motivating Idea: Zero Sum Games

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

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



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

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

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

What is the probability that X + Y = n?

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

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

X	Y	i	
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

What is the probability that X + Y = n?

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

Since this is the OR of mutually exclusive events

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

If the random variables are independent

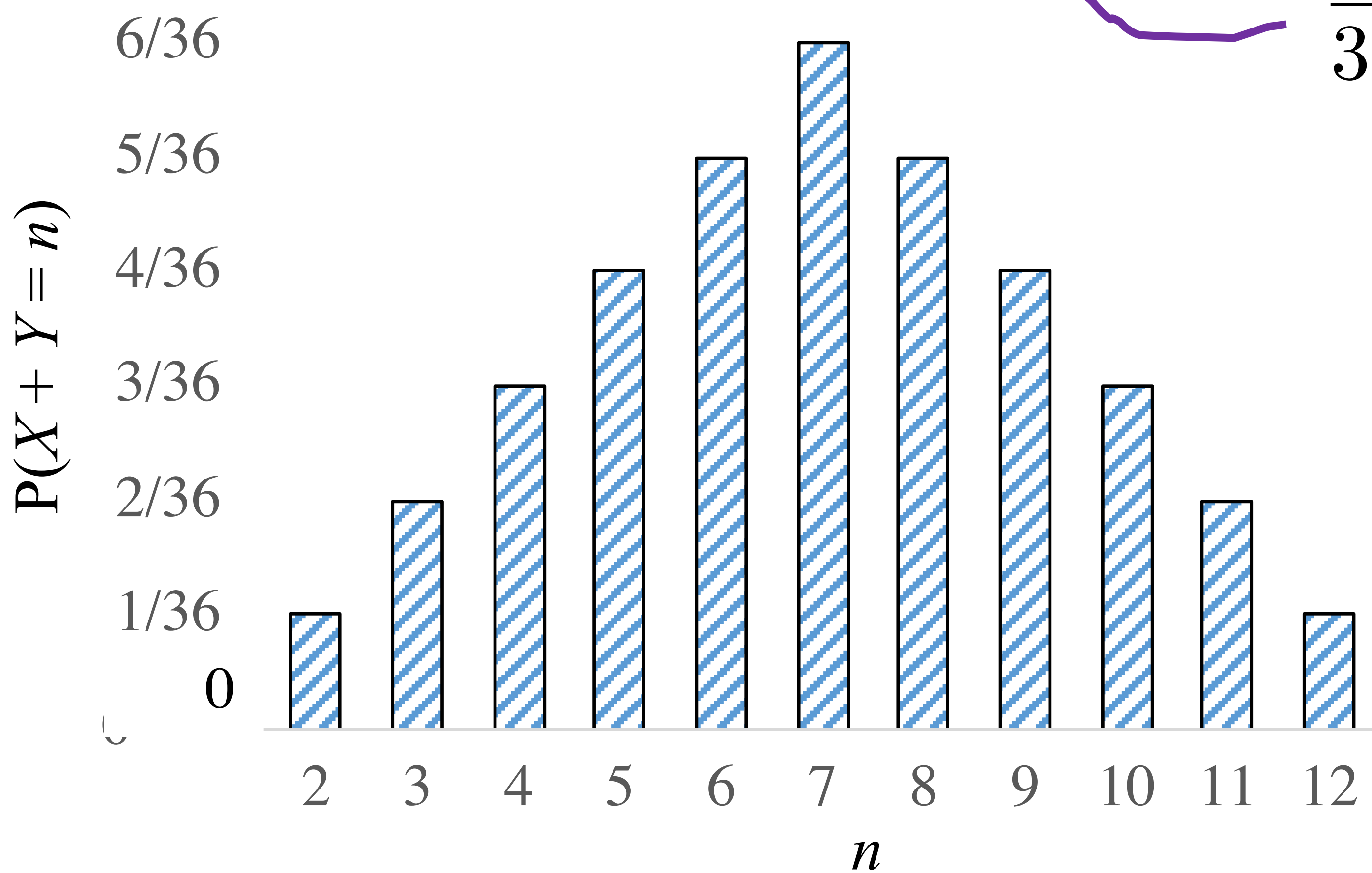
$$= \sum_{k=0}^{n} P(X = k)P(Y = n - k)$$

Sum of Two Dice

Let $X+Y$ be the value of the sum of two dice
(aka two independent random variables)

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

$\frac{1}{36}$



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 $1 \leq i \leq N$, 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 $1 \leq i \leq N$, then

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

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

Correct:

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

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

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

Incorrect:

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

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

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

X is not independent of X

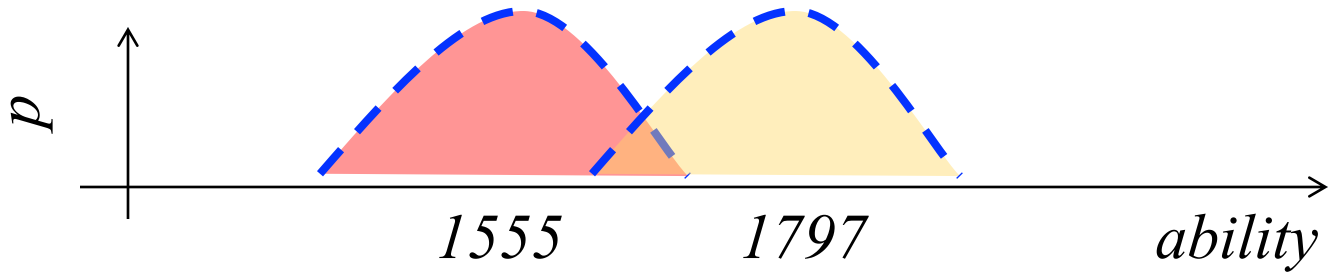
Motivating Idea: Zero Sum Games

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

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



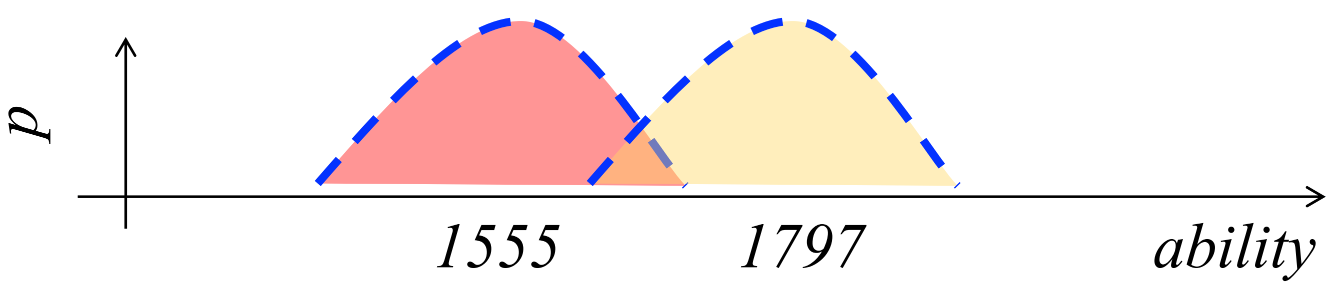
$$P(\text{Warriors win}) = P(A_W > A_B)$$

Motivating Idea: Zero Sum Games

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

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

$$D = A_W - A_B$$
$$D \sim N(\mu = 1795 - 1555, \sigma_2 = 2 \cdot 200^2)$$
$$\sim N(\mu = 240, \sigma_2 = 283)$$
$$P(D > 0) = F_D(0) = 1 - \Phi\left(\frac{0 - 240}{283}\right) \approx 0.65$$



Convolution of Probability Distributions

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

Uniform.

Generalized Convolution

Let X and Y be discrete RV's:

$$p_{X+Y}(a) = \sum_y P(X + Y = a | \mathrm{Y} = \mathrm{y})\mathrm{P}(\mathrm{Y} = \mathrm{y})$$

$$= \sum_y P(X = a - y | Y = \mathrm{y})\mathrm{P}(\mathrm{Y} = \mathrm{y})$$

If X and Y are independent…

$$= \sum_y P(X = a - y)\ \mathrm{P}(\mathrm{Y} = \mathrm{y})$$

$$= \sum_y p_X(a - y) p_Y\ (\mathrm{y})$$

Generalized Convolution

Let X and Y be continuous RV's:

$$f_{X+Y}(a) = \int_{y=-\infty}^{\infty} f(X+Y=a|Y=y)f(Y=y)\mathrm{dy}$$

$$= \int_{y=-\infty}^{\infty} f(X=a-y|Y=y)f(Y=y)\mathrm{dy}$$

If X and Y are independent…

$$= \int_{y=-\infty}^{\infty} f(X=a-y)f(Y=y)\mathrm{dy}$$

$$= \int_{y=-\infty}^{\infty} f_X(a-y)f_Y(y)\mathrm{dy}$$

Dance, Dance Convolution

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

$$F_{X+Y}(a) = P(X + Y \leq a)$$

$$= \iint_{x+y\leq a} f_X(x) f_Y(y)\, dx\, dy = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{a-y} f_X(x)\, dx\, f_Y(y)\, dy$$

$$= \int_{y=-\infty}^{\infty} F_X(a-y)\, f_Y(y)\, dy$$

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

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

X and Y are independent

$f_{X+Y}(\alpha)?$

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X=k, Y=\alpha-k)\ dk$$

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X=k)f(Y=\alpha-k)\ dk$$

Sum of Independent Uniforms

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

X and Y are independent

$f_{X+Y}(\alpha)?$

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X = k) f(Y = \alpha - k)\, dk$$

Sum of Independent Uniforms

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

X and Y are independent

$$f_{X+Y}(\alpha)?$$

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X=k)f(Y=\alpha-k)\ dk$$

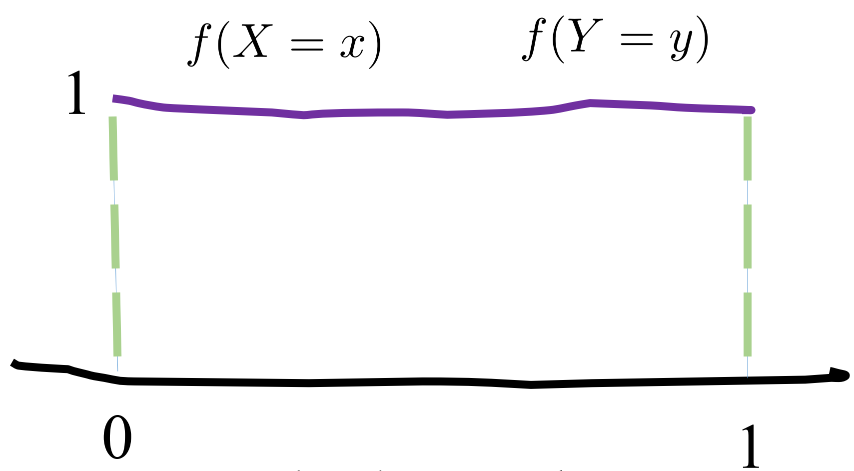
Sum of Independent Uniforms

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

X and Y are independent

$$f_{X+Y}(\alpha)?$$

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X=k) f(Y=\alpha-k) \, dk$$



For both X and Y

Sum of Independent Uniforms

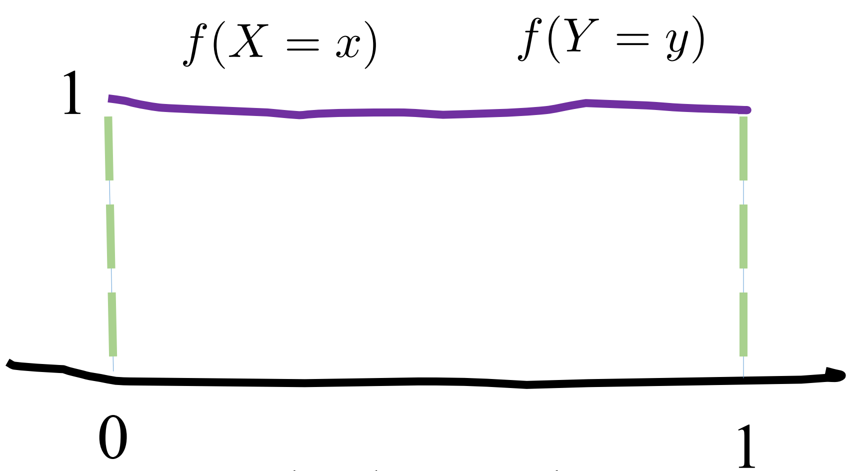
$X \sim \text{Uni}(0,1)$ $\quad Y \sim \text{Uni}(0,1)$

X and Y are independent

$f_{X+Y}(\alpha)$?

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X=k)f(Y=\alpha-k)\, dk$$

ISNT THIS JUST ONE!?!?!!!??!



For both X and Y

Sum of Independent Uniforms

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

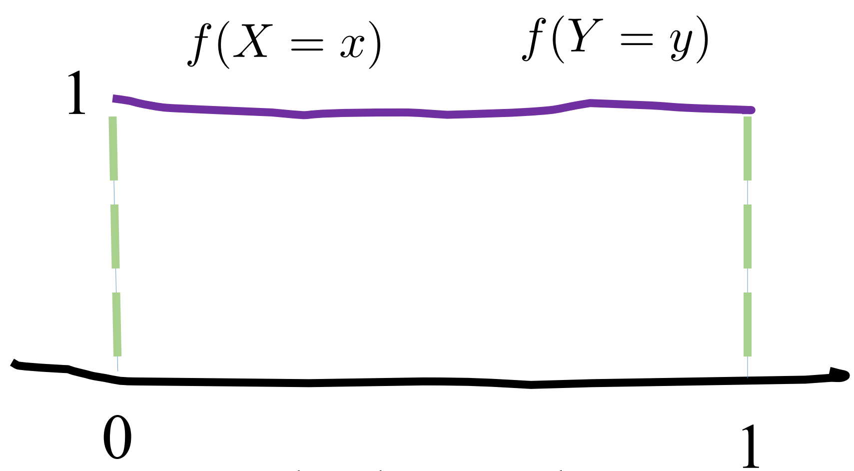
X and Y are independent

$f_{X+Y}(\alpha)?$

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X=k)f(Y=\alpha-k)\,dk$$

For these values of k, the densities of f_X and f_Y are 1

$0 < k < 1 \qquad 0 < \alpha - k < 1$



For both X and Y

Sum of Independent Uniforms

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

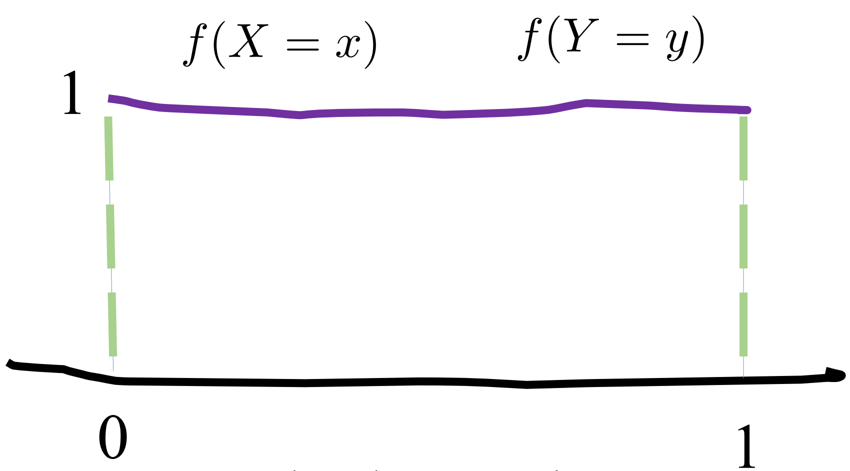
X and Y are independent

$$f_{X+Y}(\alpha)?$$

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X=k) f(Y=\alpha-k)\, dk$$

For these values of k, the densities of f_X and f_Y are 1

$$0 < k < 1 \qquad -\alpha < -k < 1-\alpha$$



For both X and Y

Sum of Independent Uniforms

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

X and Y are independent

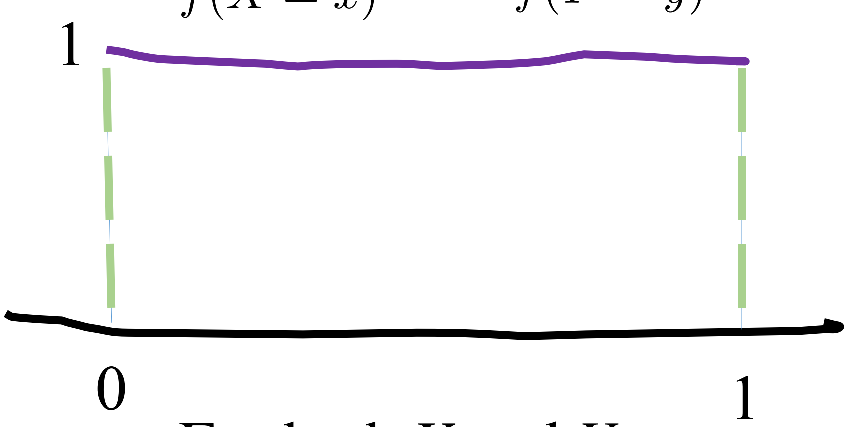
$f_{X+Y}(\alpha)$?

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X=k) f(Y=\alpha-k)\, dk$$

For these values of k, the densities of f_X and f_Y are 1

$0 < k < 1 \qquad \alpha - 1 < k < \alpha$

$f(X=x) \qquad f(Y=y)$



For both X and Y

$\alpha = ½$

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

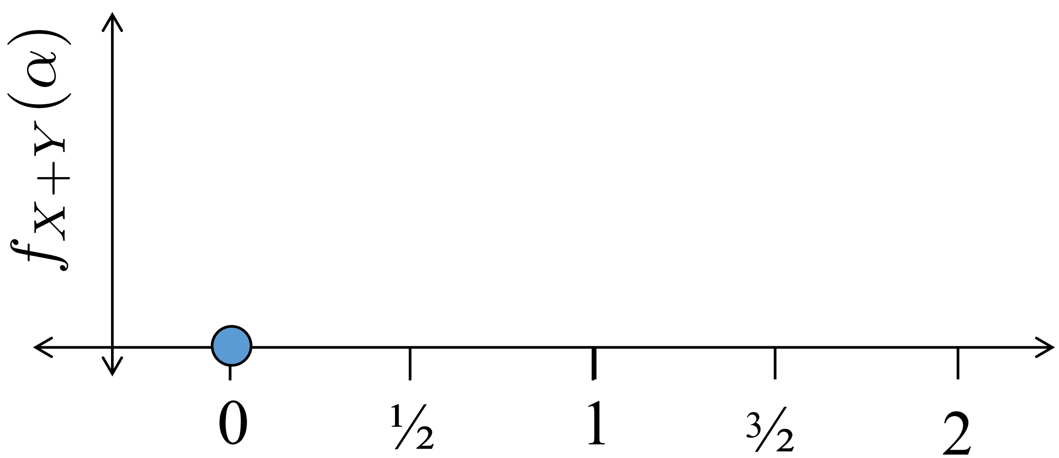
X and Y are independent

$f_{X+Y}(\alpha)?$

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X=k) f(Y=\alpha-k)\, dk$$

For these values of k, the densities of f_X and f_Y are 1

$0 < k < 1 \qquad \alpha - 1 < k < \alpha$



$\alpha = ½$

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

X and Y are independent

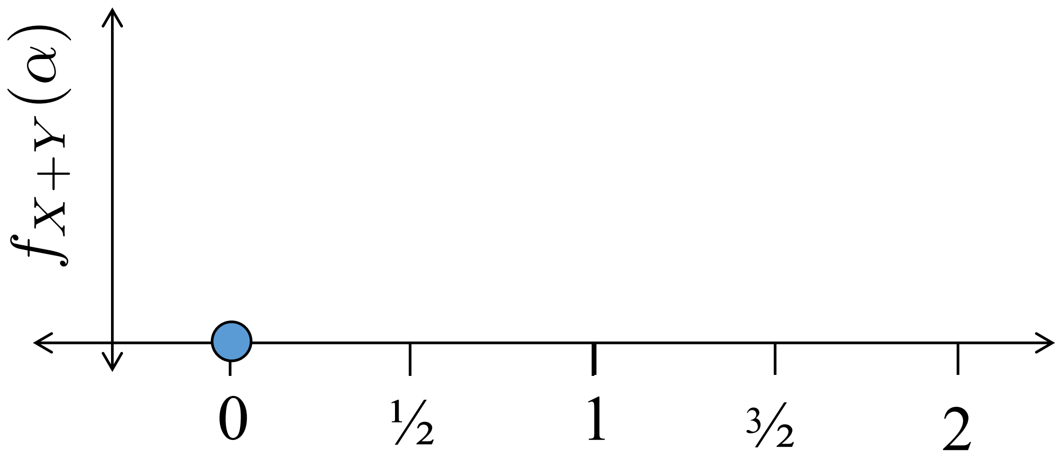
$f_{X+Y}(\alpha)?$

$$f_{X+Y}(1/2) = \int_{k=-\infty}^{\infty} f(X=k) f(Y=1/2-k)\; dk$$

$\alpha = 1/2$

For these values of k, the densities of f_X and f_Y are 1

$0 < k < 1 \qquad \alpha - 1 < k < \alpha$



$\alpha = ½$

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

X and Y are independent

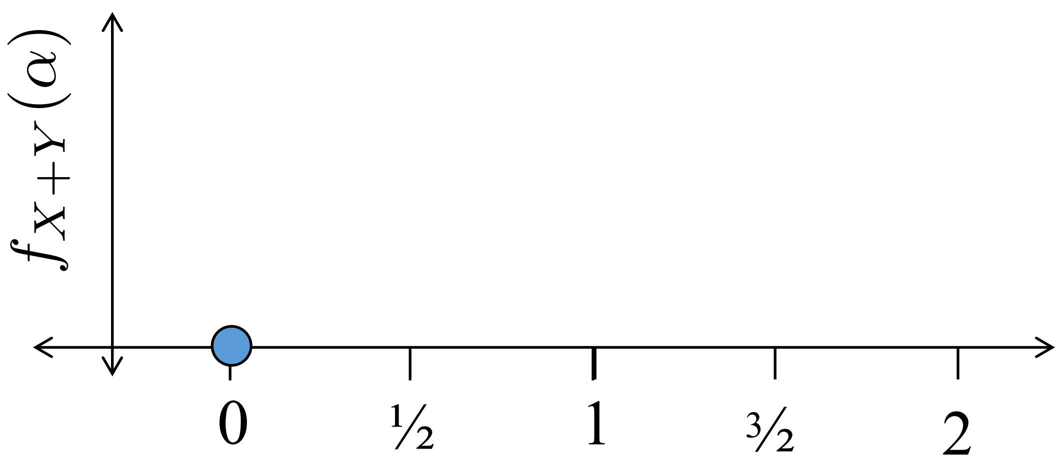
$f_{X+Y}(\alpha)?$

$$f_{X+Y}(1/2) = \int_{k=-\infty}^{\infty} f(X=k)f(Y=1/2-k)\ dk$$

$\alpha = 1/2$

For these values of k, the densities of f_X and f_Y are 1

$0 < k < 1 \qquad -1/2 < k < 1/2$



$\alpha = ½$

$X \sim \text{Uni}(0, 1) \quad Y \sim \text{Uni}(0, 1)$

X and Y are independent

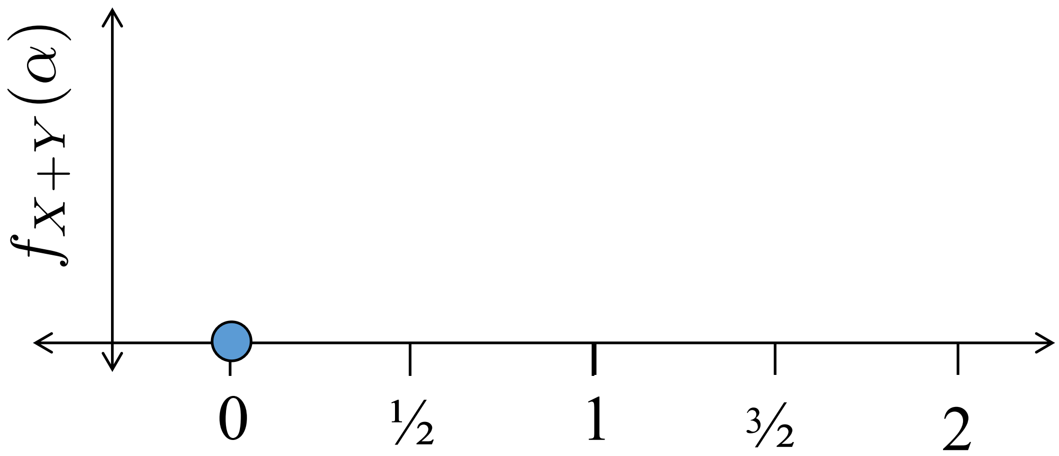
$f_{X+Y}(\alpha)?$

$$f_{X+Y}(1/2) = \int_{k=0}^{1/2} f(X = k) f(Y = 1/2 - k) \, dk$$

$\alpha = 1/2$

For these values of k, the densities of f_X and f_Y are 1

$0 < k < 1 \qquad -1/2 < k < 1/2$



$\alpha = ½$

$X \sim \text{Uni}(0,1)$ $Y \sim \text{Uni}(0,1)$

X and Y are independent

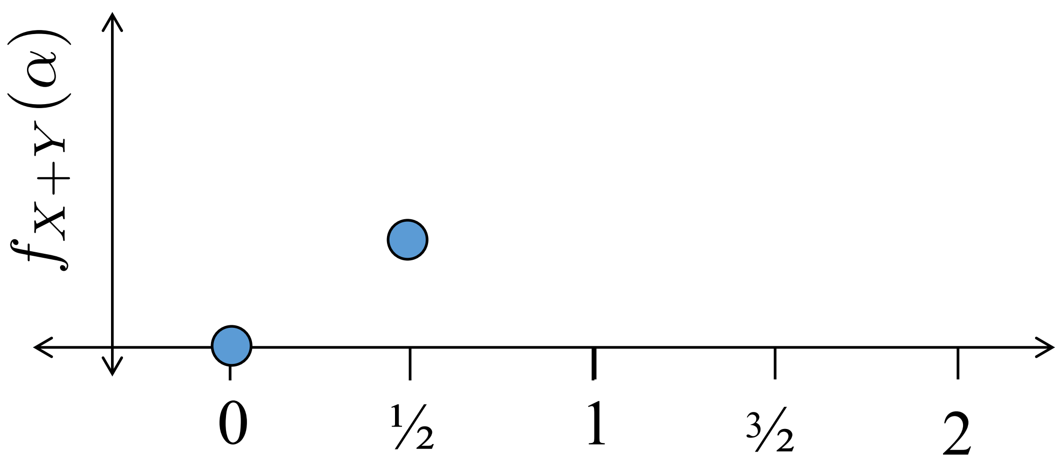
$$f_{X+Y}(\alpha)?$$

$$f_{X+Y}(1/2) = \int_{k=0}^{1/2} 1 \; dk = 0.5$$

$\alpha = 1/2$

For these values of k, the densities are 1

$0 < k < 1$ $-1/2 < k < 1/2$



$0 < \alpha < 1$

$X \sim \text{Uni}(0, 1) \quad Y \sim \text{Uni}(0, 1)$

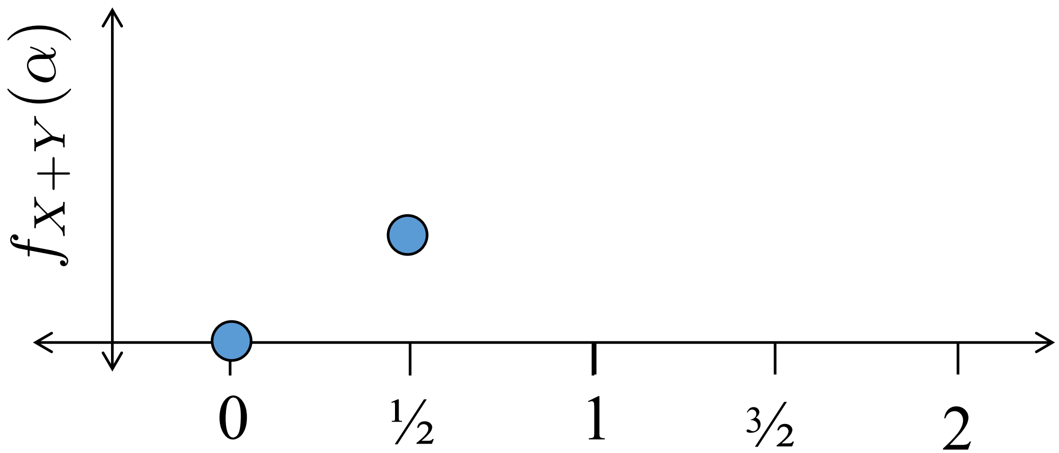
X and Y are independent

$f_{X+Y}(\alpha)?$

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X = k) f(Y = \alpha - k) \, dk$$

For these values of k, the densities are 1

$0 < k < 1 \qquad \alpha - 1 < k < \alpha$



$0 < \alpha < 1$

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

X and Y are independent

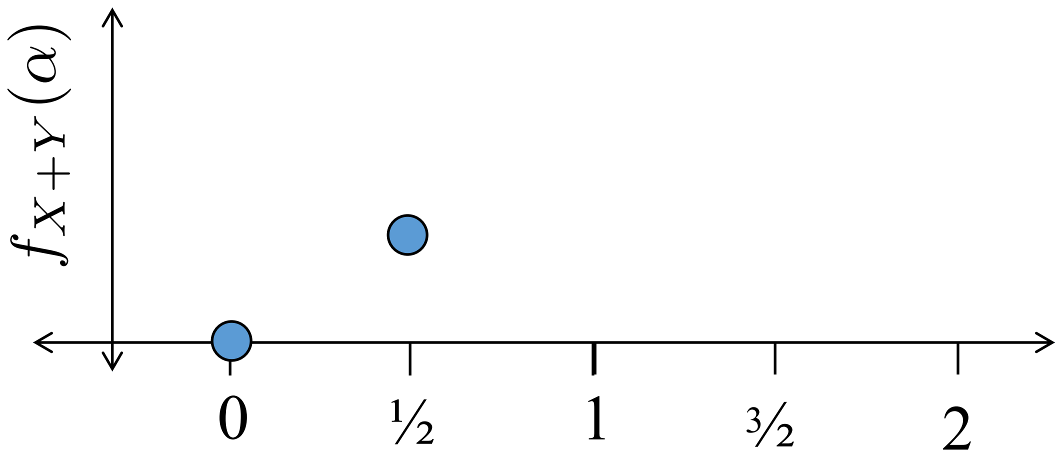
$f_{X+Y}(\alpha)$?

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X=k)f(Y=\alpha-k)\ dk$$

For these values of k, the densities are 1

$0 < k < 1 \qquad \alpha - 1 < k < \alpha$

$0 < k < \alpha$



$0 < \alpha < 1$

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

X and Y are independent

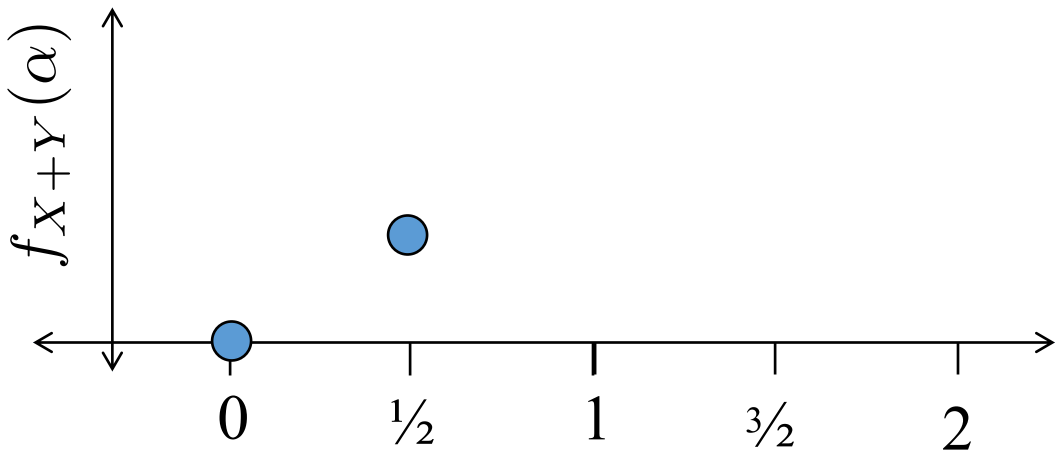
$f_{X+Y}(\alpha)?$

$$f_{X+Y}(\alpha) = \int_{k=0}^{\alpha} f(X=k)f(Y=\alpha-k)\, dk$$

For these values of k, the densities are 1

$0 < k < 1 \qquad \alpha - 1 < k < \alpha$

$0 < k < \alpha$



$0 < \alpha < 1$

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

X and Y are independent

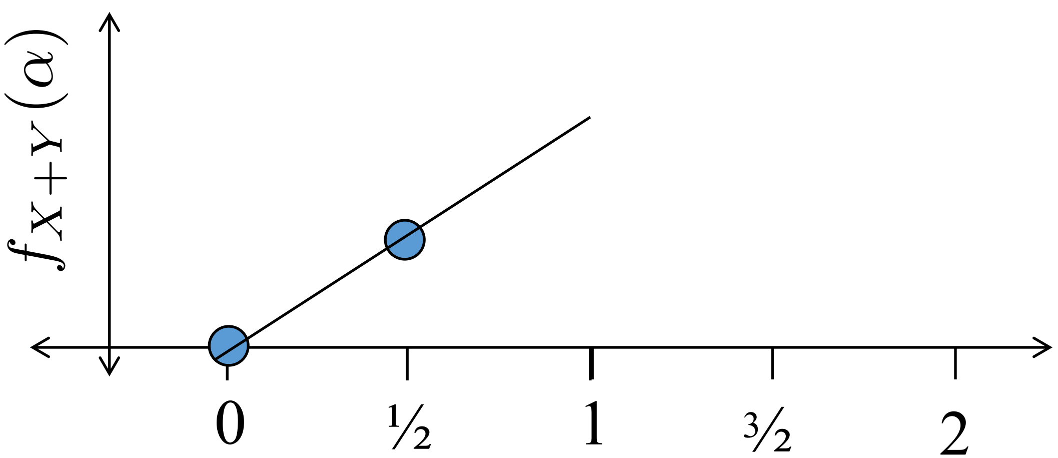
$f_{X+Y}(\alpha)?$

$$f_{X+Y}(\alpha) = \int_{k=0}^{\alpha} 1 \; dk = \alpha$$

For these values of k, the densities are 1

$0 < k < 1 \qquad \alpha - 1 < k < \alpha$

$0 < k < \alpha$



$1 < \alpha < 2$

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

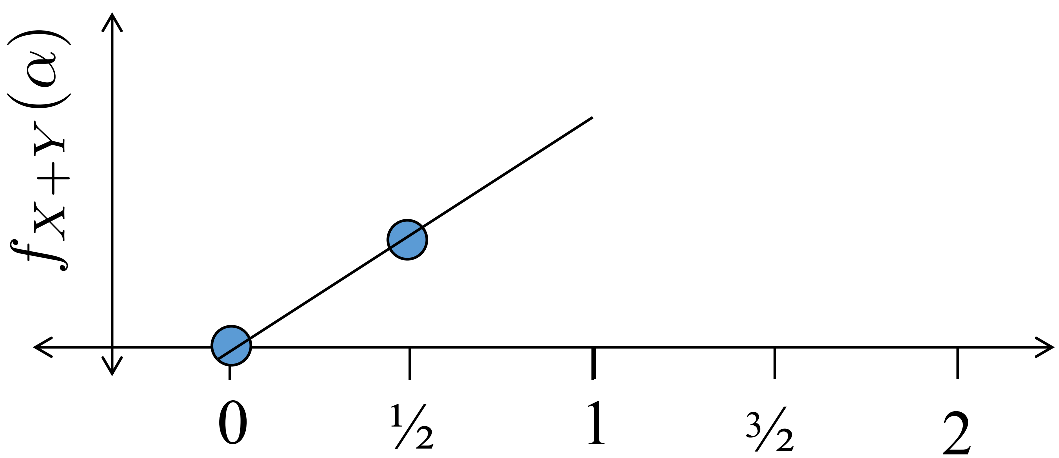
X and Y are independent

$f_{X+Y}(\alpha)$?

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X=k) f(Y=\alpha-k)\, dk$$

For these values of k, the densities are 1

$0 < k < 1 \qquad \alpha - 1 < k < \alpha$



$1 < \alpha < 2$

$X \sim \text{Uni}(0, 1)$ $Y \sim \text{Uni}(0, 1)$

X and Y are independent

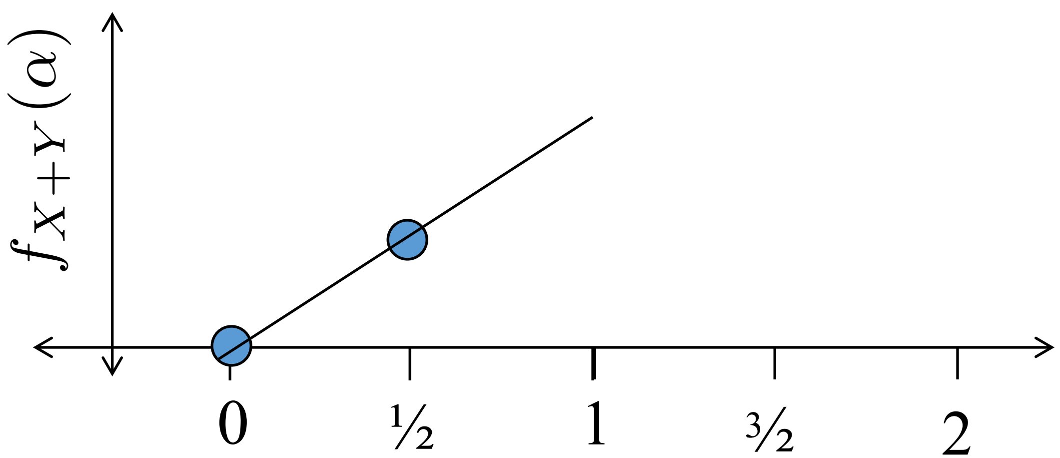
$f_{X+Y}(\alpha)$?

$$f_{X+Y}(\alpha) = \int_{k=-\infty}^{\infty} f(X = k) f(Y = \alpha - k) \; dk$$

For these values of k, the densities are 1

$0 < k < 1$ $\alpha - 1 < k < \alpha$

$\alpha - 1 < k < 1$



$1 < \alpha < 2$

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

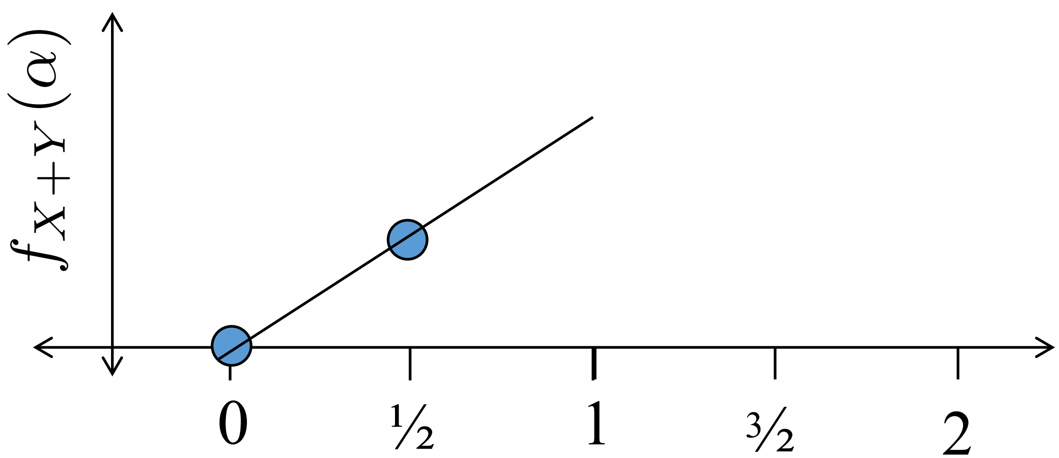
X and Y are independent

$f_{X+Y}(\alpha)$?

$$f_{X+Y}(\alpha) = \int_{k=\alpha-1}^{1} f(X=k)f(Y=\alpha-k)\ dk$$

For these values of k, the densities are 1

$0 < k < 1 \qquad \alpha - 1 < k < \alpha \qquad \alpha - 1 < k < 1$



$1 < \alpha < 2$

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

X and Y are independent

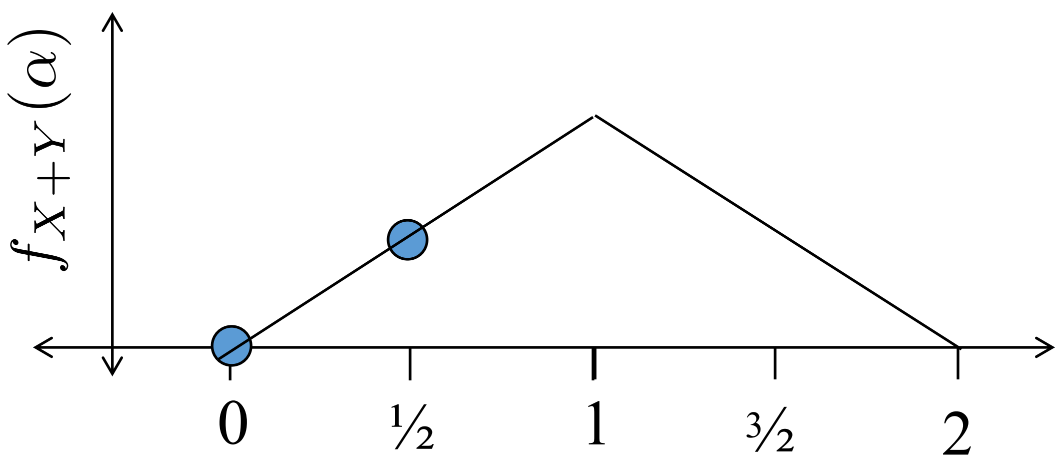
$$f_{X+Y}(\alpha)?$$

$$f_{X+Y}(\alpha) = \int_{k=\alpha-1}^{1} 1 \, dk = 2 - \alpha$$

For these values of k, the densities are 1

$0 < k < 1 \qquad \alpha - 1 < k < \alpha$

$\alpha - 1 < k < 1$



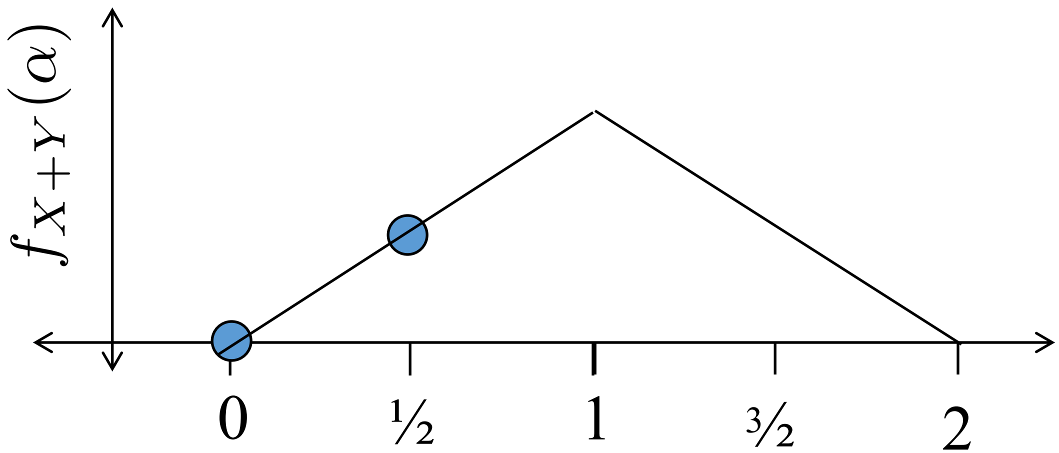
$1 < \alpha < 2$

$X \sim \text{Uni}(0,1) \quad Y \sim \text{Uni}(0,1)$

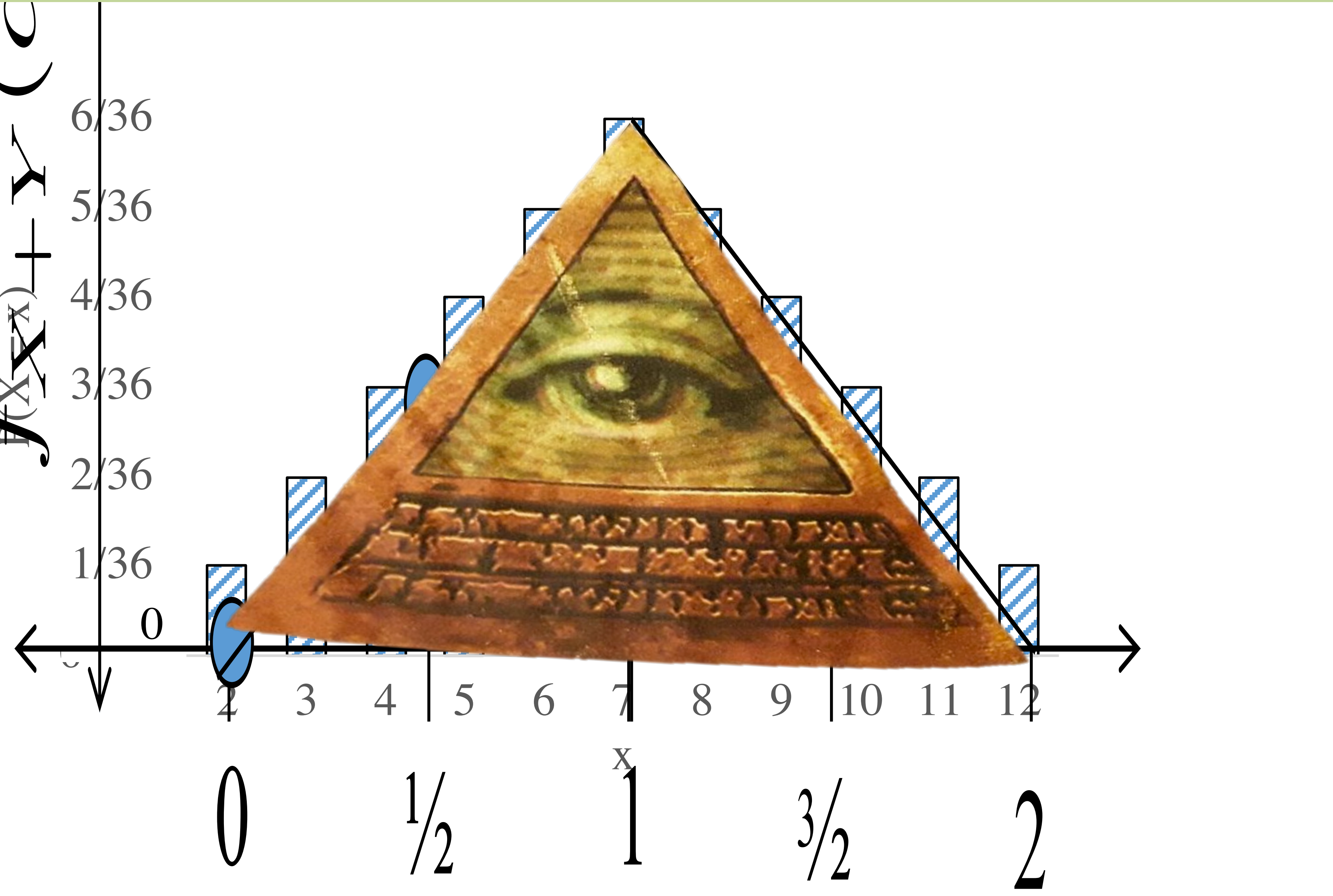
X and Y are independent

$f_{X+Y}(\alpha)$?

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



Sum of Uniforms and Sum of Dice



Stretch!

The Random Variable for Probabilities

Noah Arthurs
CS109, Stanford University

Today we are going to learn something unintuitive, beautiful and useful

Review

Continuous Conditional Distributions

- Let X be continuous random variable
- Let E be an event:

$$P(E|X = x) = \frac{P(X = x, E)}{P(X = x)}$$

$$= \frac{P(X = x|E)P(E)}{P(X = x)}$$

$$= \frac{f_X(x|E)P(E)\epsilon_x}{f_X(x)\epsilon_x}$$

You can skip to this version

$$= \frac{f_X(x|E)P(E)}{f_X(x)}$$

Continuous Conditional Distributions

- Let X be a measure of time to answer a question
- Let E be the event that the user is a human:

$$
\begin{aligned}
P(E|X = x) &= \frac{P(X = x, E)}{P(X = x)} \\
&= \frac{P(X = x|E)P(E)}{P(X = x)} \\
&= \frac{f_X(x|E)P(E)\epsilon_x}{f_X(x)\epsilon_x} \\
&= \frac{f_X(x|E)P(E)}{f_X(x)}
\end{aligned}
$$

Biometric Keystroke

- Let X be a measure of time to answer a question
- Let E be the event that the user is a human
- What if you don't know normalization term?:

Normal pdf

Prior

$$P(E|X=x) = \frac{f_X(x|E)P(E)}{f_X(x)}$$

???

$$\frac{P(E|X=x)}{P(E^C|X=x)}$$

End Review

Lets play a game

Roll a dice twice. If either time you roll a 6, I win.
Otherwise you win.

$$P(W) = \left(\frac{5}{6}\right)^2 \approx 0.69$$

Flip a Coin With Unknown Probability

We are going to think of probabilities as random variables!!!

Flip a Coin With Unknown Probability

- Flip a coin (n + m) times, comes up with n heads
  - We don't know probability X that coin comes up heads

Frequentist

$$X = \lim_{n+m \to \infty} \frac{n}{n+m} \approx \frac{n}{n+m}$$

X is a single value

Bayesian

$$f_{X|N}(x|n) = \frac{P(N = n | X = x) f_X(x)}{P(N = n)}$$

X is a random variable

What is your belief that you successfully roll a 6 on my die?

Flip a Coin With Unknown Probability

- Flip a coin (n + m) times, comes up with n heads
  - We don't know probability X that coin comes up heads
  - Our belief before flipping coins is that: X ~ Uni(0, 1)
  - Let N = number of heads
  - Given X = x, coin flips independent: (N | X) ~ Bin(n + m, x)

$$f_{X|N}(x|n) = \frac{P(N = n|X = x) f_X(x)}{P(N = n)}$$

Bayesian "posterior" probability distribution

Bayesian "prior" probability distribution

Flip a Coin With Unknown Probability

- Flip a coin (n + m) times, comes up with n heads
  - We don't know probability X that coin comes up heads
  - Our belief before flipping coins is that: X ~ Uni(0, 1)
  - Let N = number of heads
  - Given X = x, coin flips independent: (N | X) ~ Bin(n + m, x)

$$f_{X|N}(x|n) = \frac{P(N = n|X = x) f_X(x)}{P(N = n)}$$

Binomial

1 (Uniform assumption)

$$= \frac{\binom{n+m}{n} x^n (1-x)^m}{P(N = n)}$$

$$= \frac{\binom{n+m}{n}}{P(N = n)} x^n (1-x)^m$$

Move terms around

$$= \frac{1}{c} \cdot x^n (1-x)^m \qquad \text{where } c = \int_0^1 x^n (1-x)^m dx$$

Flip a Coin With Unknown Probability

If you start with a $X \sim \text{Uni}(0, 1)$ prior over probability, and observe:

n "successes" and
m "failures"…

Your new belief about the probability is:

$$f_X(x) = \frac{1}{c} \cdot x^n (1 - x)^m$$

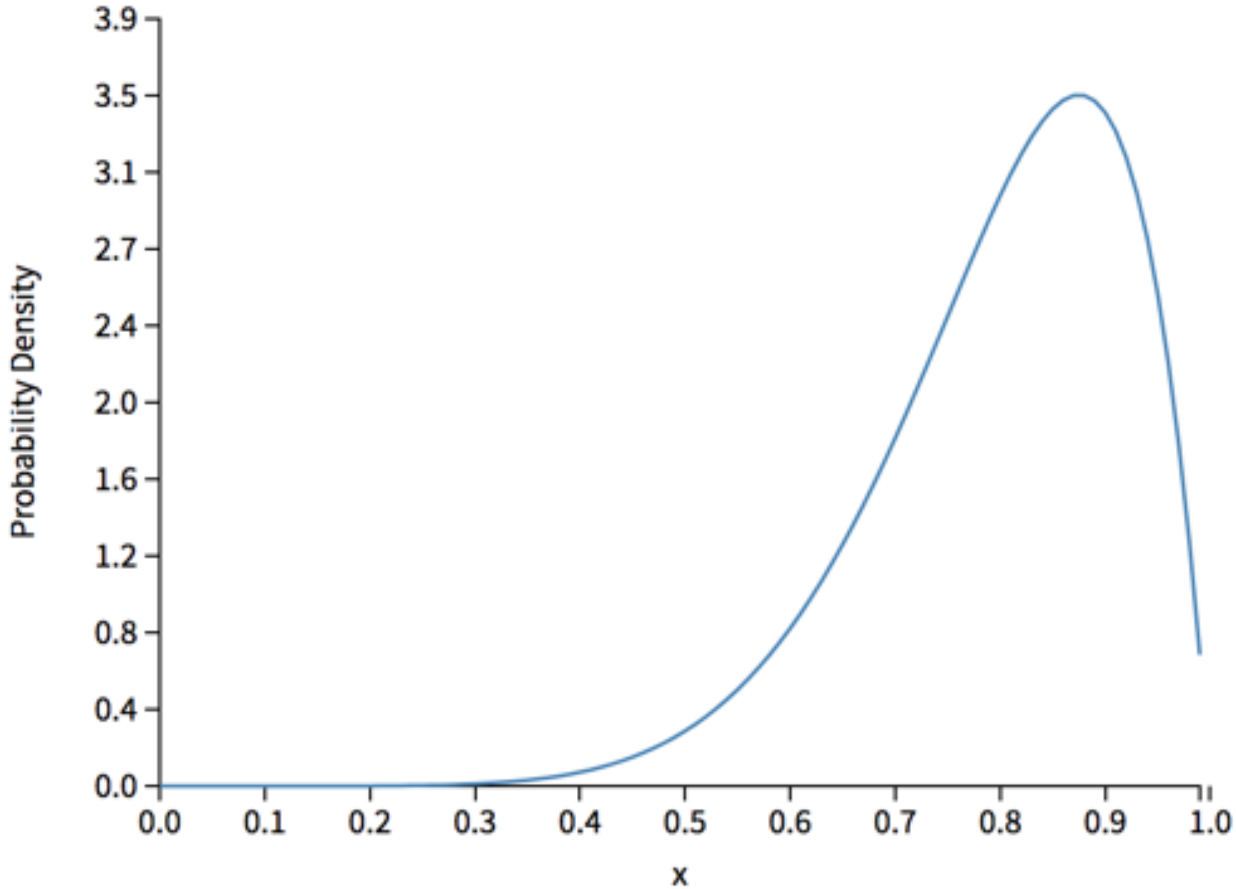
$$\text{where} \quad c = \int_0^1 x^n (1 - x)^m$$

Belief after 7 “success” and 1 “fail”

$$f_X(x) = \frac{1}{c} \cdot x^n (1 - x)^m$$

$n = 7$

$m = 1$



Equivalently

If you start with a $X \sim \text{Uni}(0, 1)$ prior over probability, and observe:

let a = num "successes" + 1
let b = num "failures" + 1

Your new belief about the probability is:

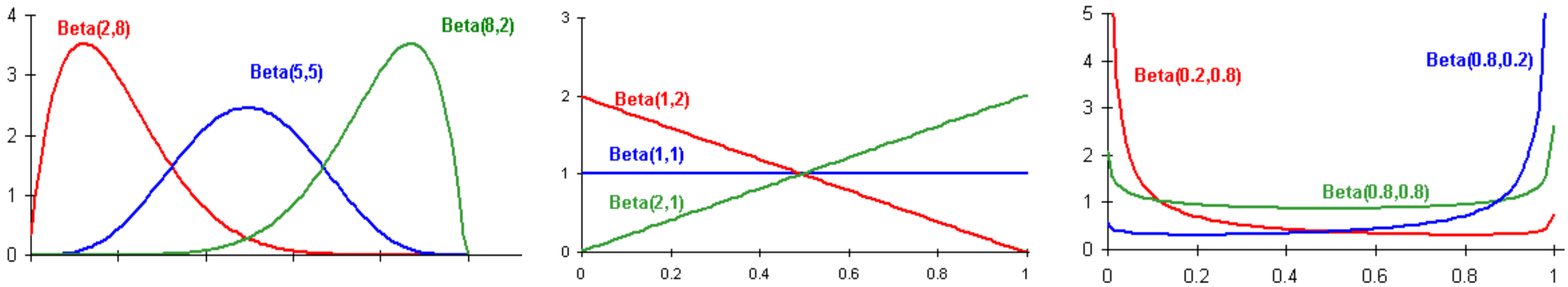
$$f_X(x) = \frac{1}{c} \cdot x^{a-1}(1-x)^{b-1}$$

where $$c = \int_0^1 x^{a-1}(1-x)^{b-1}$$

Beta Random Variable

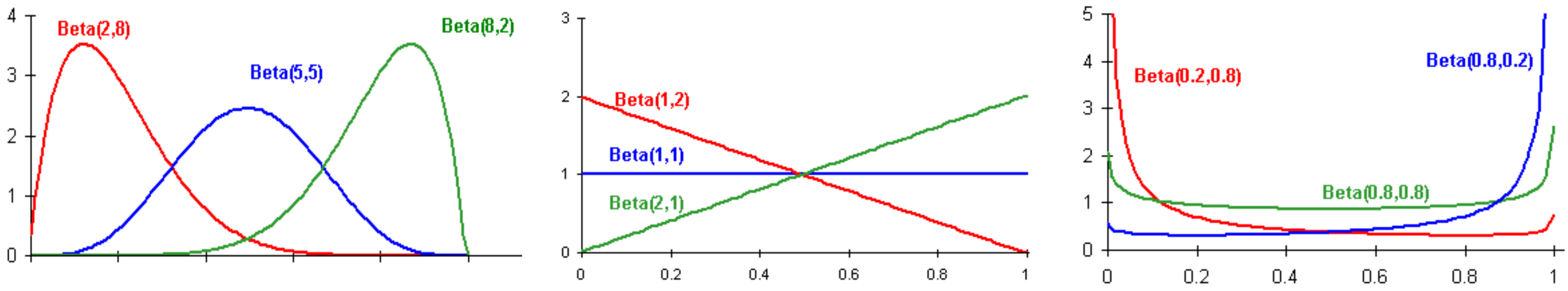
- X is a **Beta Random Variable**: X ~ Beta(*a*, *b*)
  - Probability Density Function (PDF): (where *a*, *b* > 0)

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{where} \quad B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$



  - Symmetric when *a* = *b*
  - $E[X] = \frac{a}{a+b}$ $\qquad Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$

Meta Beta



Used to represent a
distributed belief of a probability

Beta is a distribution for probabilities

Beta Parameters *can* come from experiments:

a = "successes" + 1

b = "failures" + 1

Back to flipping coins

- Flip a coin (n + m) times, comes up with n heads
  - We don't know probability X that coin comes up heads
  - Our belief before flipping coins is that: X ~ Uni(0, 1)
  - Let N = number of heads
  - Given X = x, coin flips independent: (N | X) ~ Bin(n + m, x)

$$
\begin{aligned}
f_{X|N}(x|n) &= \frac{P(N=n|X=x)f_X(x)}{P(N=n)} \\
&= \frac{\binom{n+m}{n} x^n (1-x)^m}{P(N=n)} \\
&= \frac{\binom{n+m}{n}}{P(N=n)} x^n (1-x)^m \\
&= \frac{1}{c} \cdot x^n (1-x)^m \qquad \text{where } c = \int_0^1 x^n (1-x)^m dx
\end{aligned}
$$

Understanding Beta

- X | (N = n, M = m) ~ Beta(a = n + 1, b = m + 1)
  - Prior X ~ Uni(0, 1)
  - Check this out, boss:
    - Beta(a = 1, b = 1) =?

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} = \frac{1}{B(a,b)} x^{0}(1-x)^{0}$$

$$= \frac{1}{\int_0^1 1\,dx} 1 = 1 \quad \text{where} \quad 0 < x < 1$$

    - Beta(a = 1, b = 1) = Uni(0, 1)
  - So, prior X ~ Beta(a = 1, b = 1)

N successes

M failures

If the Prior was a Beta...

X is our random variable for probability

If our **prior belief** about X was beta

$$f(X = x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

What is our **posterior belief** about X after observing *n* heads (and *m* tails)?

$$f(X = x | N = n) = ???$$

If the Prior was a Beta...

$$
\begin{aligned}
f(X = x|N = n) &= \frac{P(N = n|X = x) f(X = x)}{P(N = n)} \\
&= \frac{\binom{n+m}{n} x^n (1-x)^m \, f(X = x)}{P(N = n)} \\
&= \frac{\binom{n+m}{n} x^n (1-x)^m \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}}{P(N = n)} \\
&= K_1 \cdot \binom{n+m}{n} x^n (1-x)^m \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} \\
&= K_3 \cdot x^n (1-x)^m x^{a-1} (1-x)^{b-1} \\
&= K_3 \cdot x^{n+a-1} (1-x)^{m+b-1}
\end{aligned}
$$

$$X|N \sim \text{Beta}(n + a, m + b)$$

Understanding Beta

- If "Prior" distribution of X (before seeing flips) is Beta
- Then "Posterior" distribution of X (after flips) is Beta
- Beta is a **conjugate** distribution for Beta
  - Prior and posterior parametric forms are the same!
  - Practically, conjugate means easy update:
    - Add number of "heads" and "tails" seen to Beta parameters

Further Understanding Beta

- Can set X ~ Beta(*a*, *b*) as prior to reflect how biased you think coin is apriori
  - This is a subjective probability!
  - Prior probability for X based on seeing (*a* + *b* – 2) “imaginary” trials, where

    (*a* – 1) of them were heads.

    (b – 1) of them were tails.
  - Beta(1, 1) = Uni(0, 1) → we haven’t seen any “imaginary trials”, so apriori know nothing about coin
- Update to get posterior probability
  - X | (n heads and m tails) ~ Beta(a + n, b + m)

Enchanted Die

Let X be the probability of rolling a "6" on Chris' die.

Prior: Imagine 5 die rolls where only showed up as a "6"

Observation: Roll it a few times…

What is the updated probability density function of X after our observations?

Check out Demo!