# The Random Variable for Probabilities

Noah Arthurs
CS109, Stanford University

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?

# 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 $i = 1, 2, ..., n$:

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

# 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

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

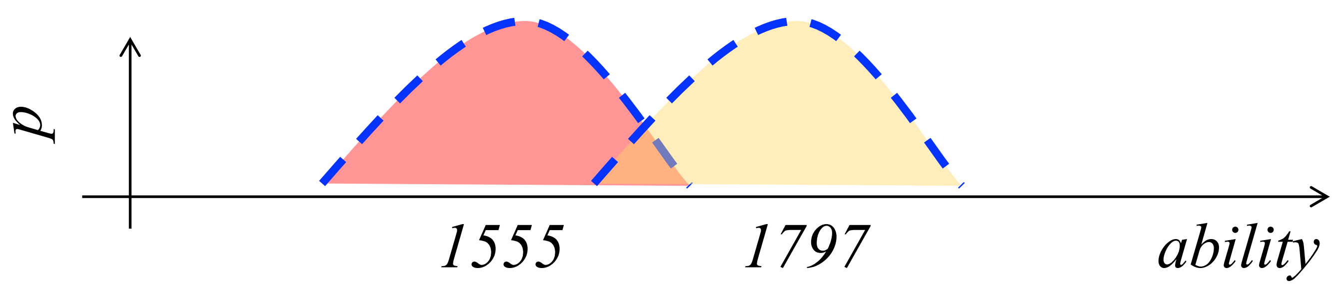
$$\begin{aligned} P(\text{Warriors win}) &= P(A_W > A_B) \\ &= P(A_W - A_B > 0) \end{aligned}$$

---

$$D = A_W - A_B$$

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

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



# Generalized Convolution

Let X and Y be discrete RV's:

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

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

If X and Y are independent…

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

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

# 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 $p(y)$

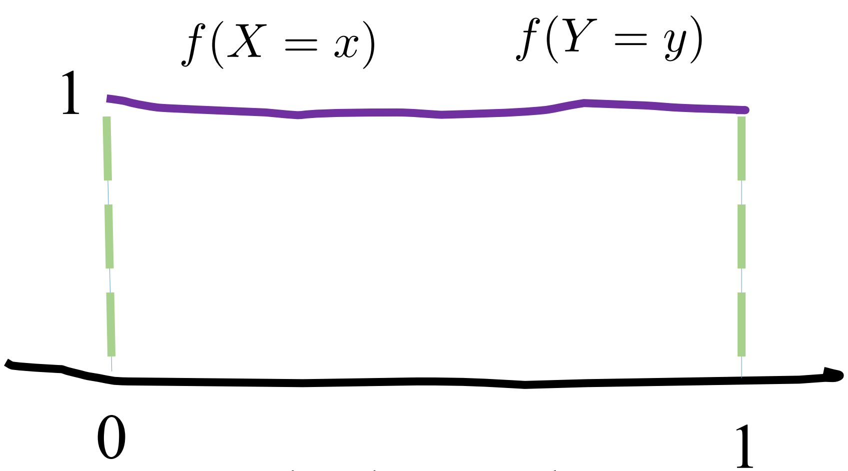
# Sum of Independent Uniforms

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

$X$ and $Y$ are independent

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

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




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

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

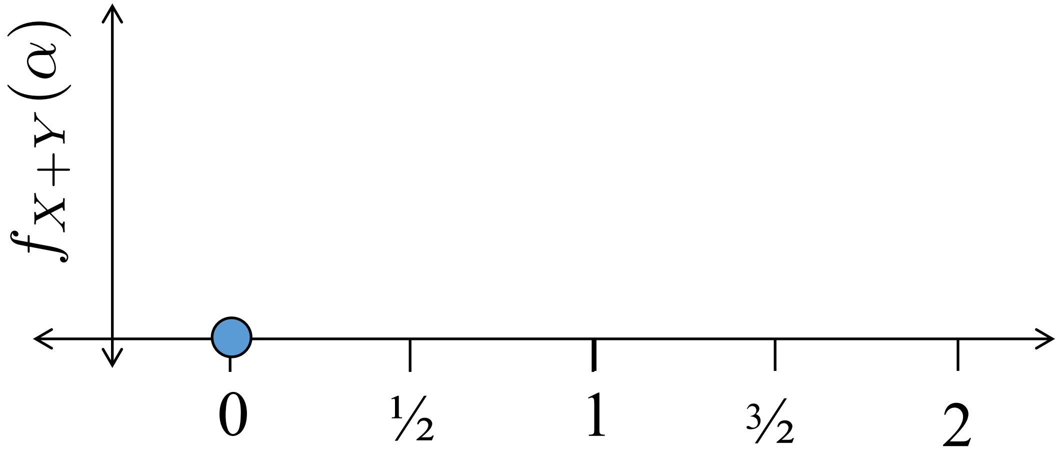
---

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

$\alpha = 1/2$

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

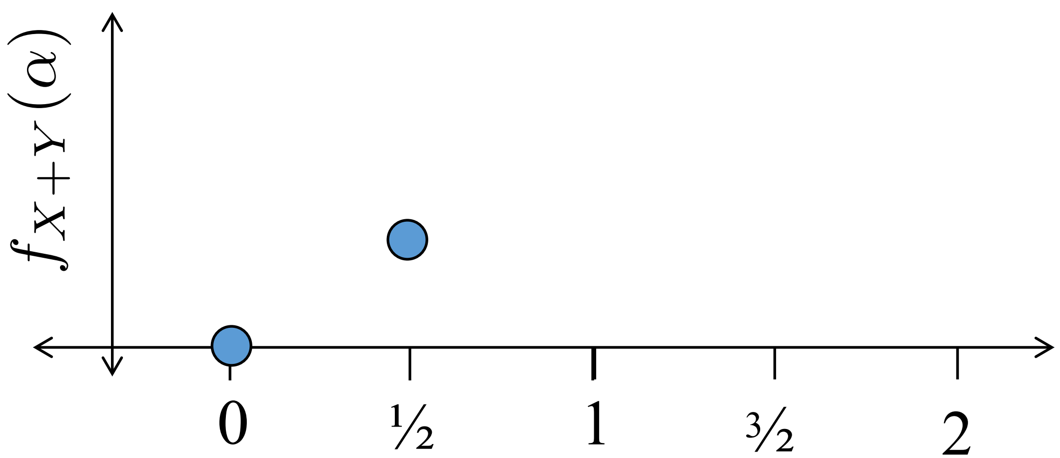
$f_{X+Y}(\alpha)?$

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

α = 1/2

For these values of k, the densities are 1

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



# $0 < \alpha < 1$

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

$X$ and $Y$ are independent

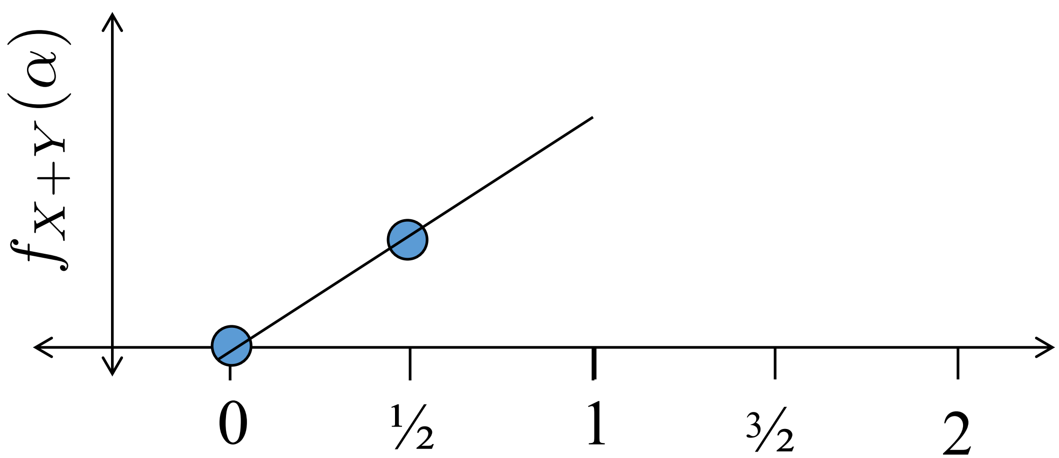
$f_{X+Y}(\alpha)?$

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

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

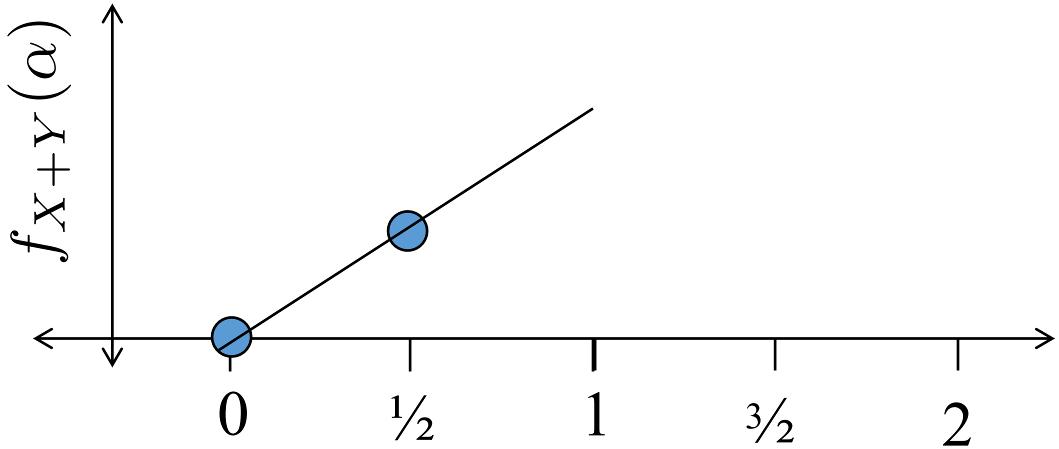
$f_{X+Y}(\alpha)?$

---

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

For these values of k, the densities are 1

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



# $1 < \alpha < 2$

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

$X$ and $Y$ are independent

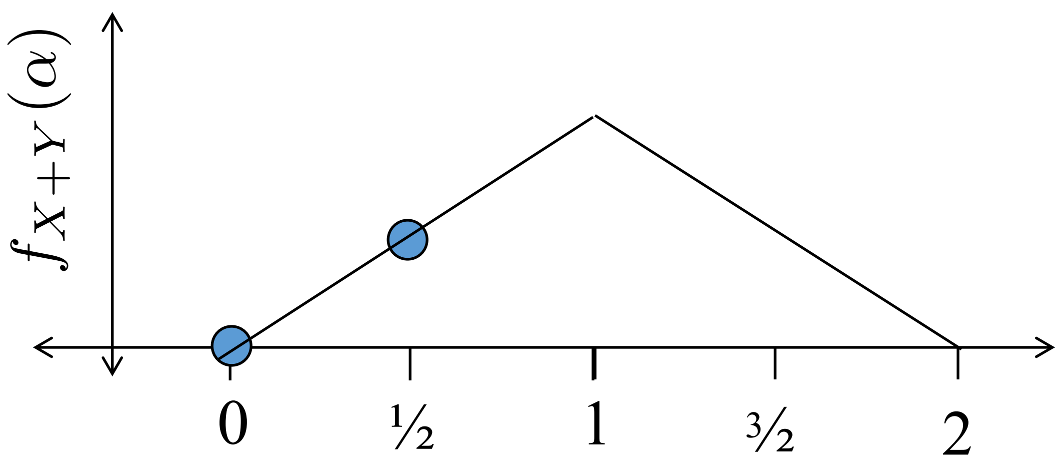
$f_{X+Y}(\alpha)?$

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

For these values of k, the densities are 1

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

$\alpha - 1 < k < 1$



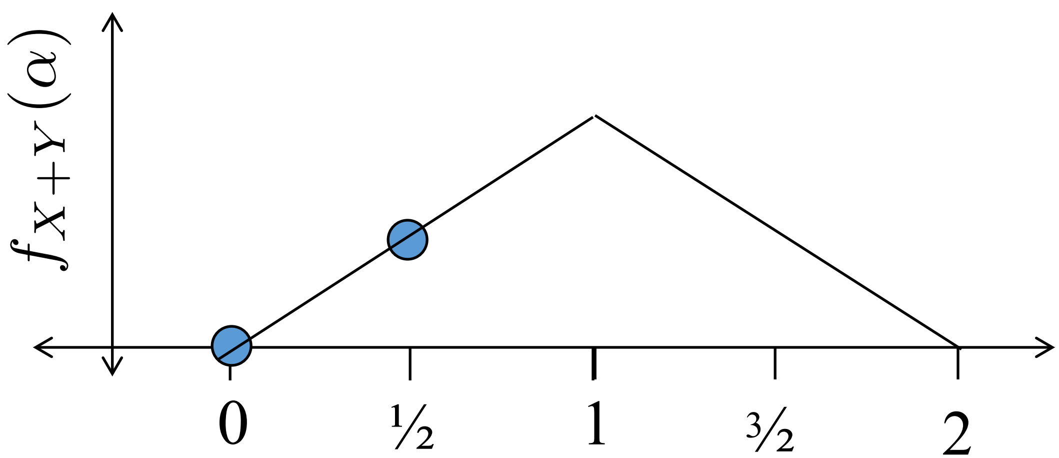
# $1 < \alpha < 2$

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

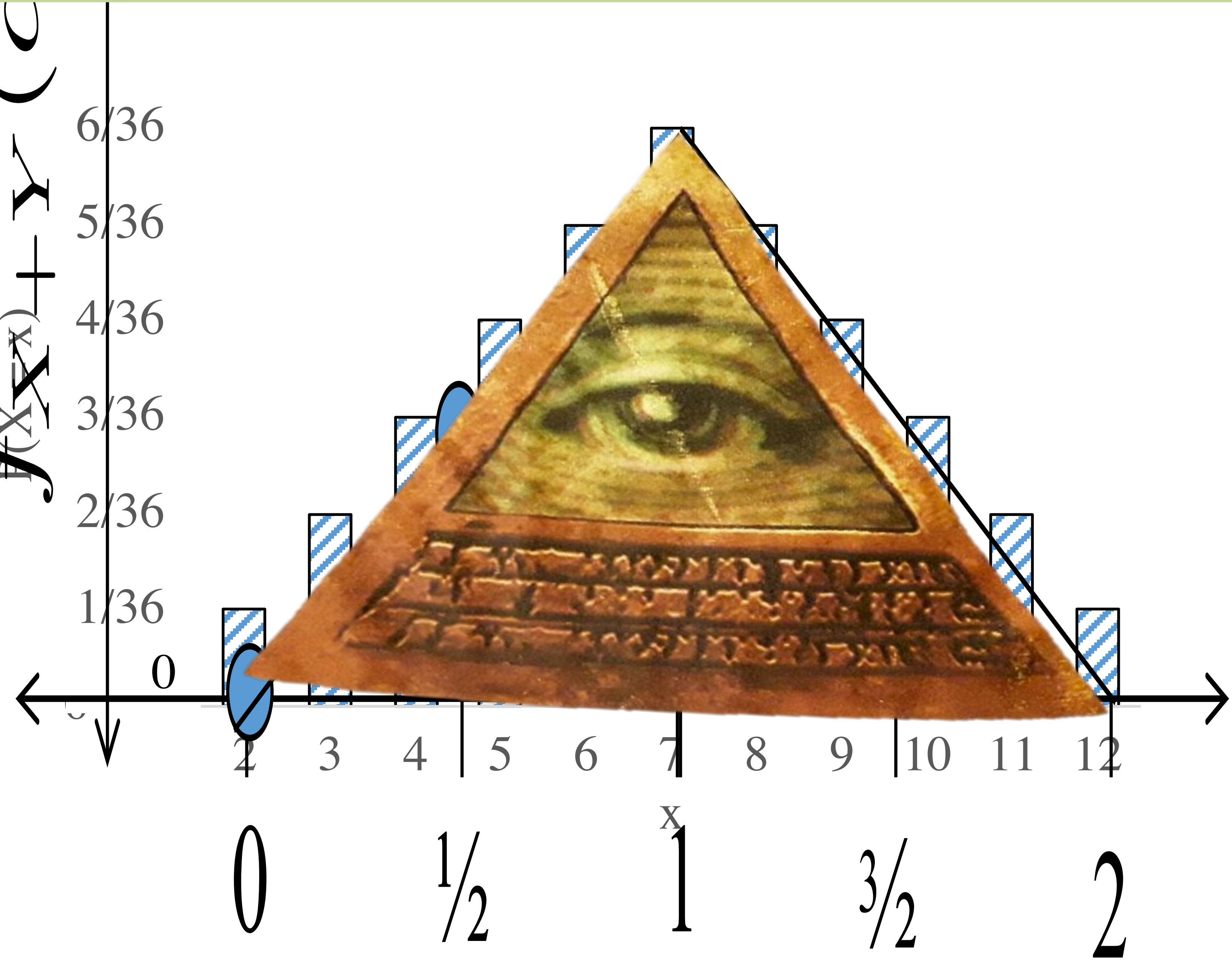
$X$ and $Y$ are independent

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

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



# Sum of Uniforms and Sum of Dice



# 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

# 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)

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

1 (Uniform assumption)

Binomial

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

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

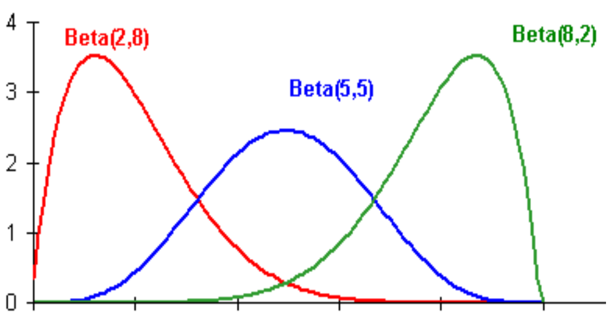
Move terms around

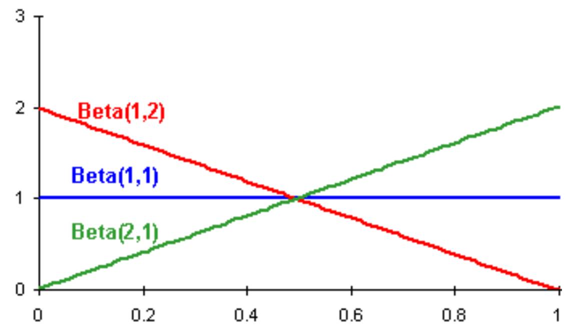
$$= \frac{1}{c} \cdot x^n (1-x)^m \qquad \text{where } c = \int_0^1 x^n (1-x)^m dx$$

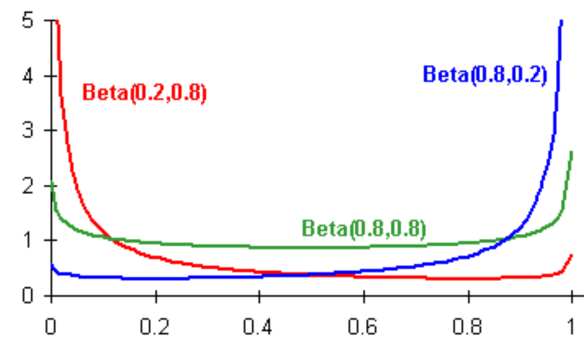
# Beta Random Variable

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

Beta is a distribution for probabilities

Beta Parameters *can* come from experiments:

$a$ = "successes" + 1

$b$ = "failures" + 1

# 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)

End Review

# Enchanted Die

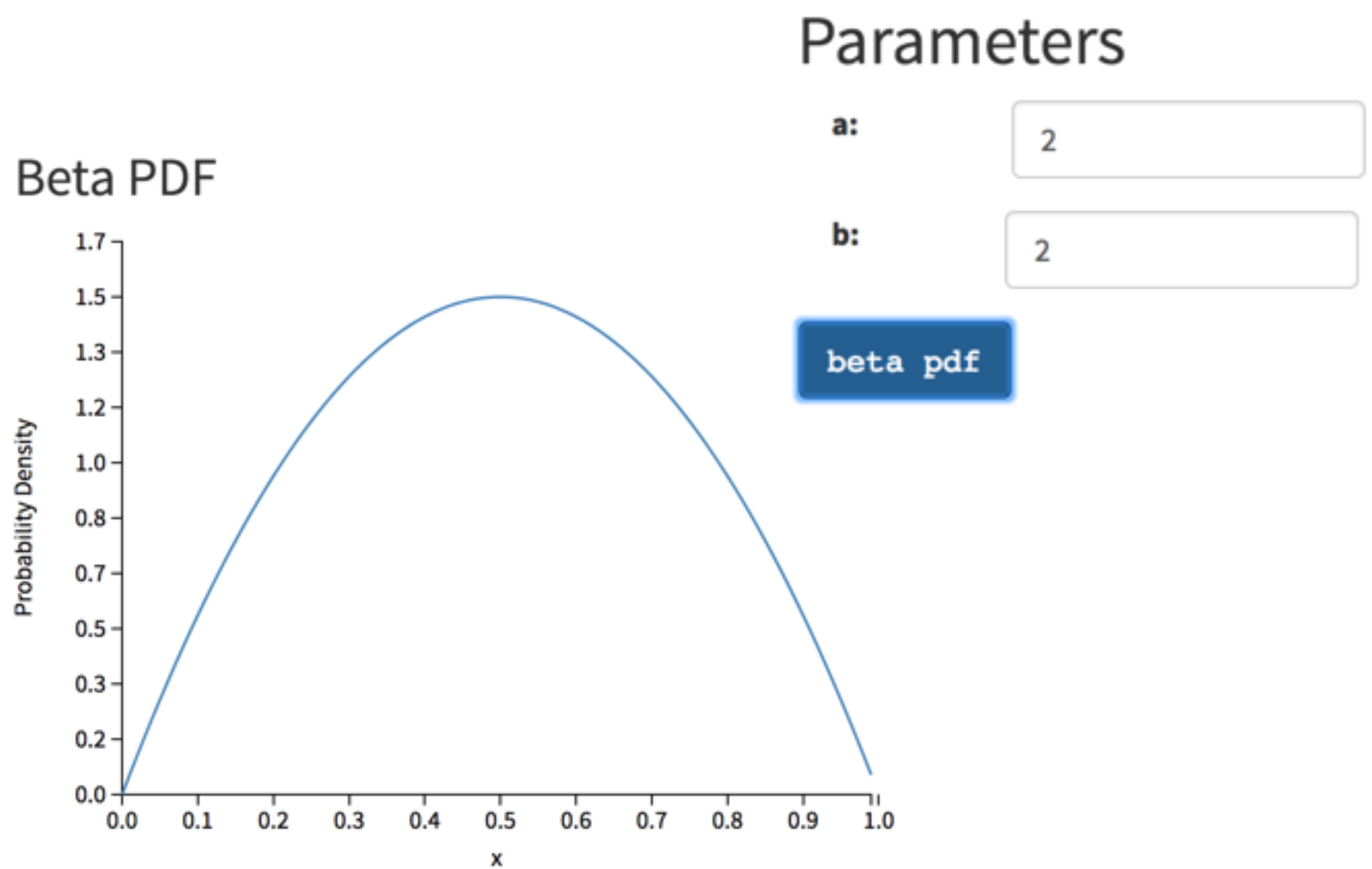
Let $X$ be the probability of rolling a "6" on Noah's die.

**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!



# Beta Example

Before being tested, a medicine is believed to "work" about 80% of the time. The medicine is tried on 20 patients. It "works" for 14 and "doesn't work" for 6. What is your new belief that the drug works?

---

Frequentist:

$$p \approx \frac{14}{20} = 0.7$$

# Beta Example

Before being tested, a medicine is believed to "work" about 80% of the time. The medicine is tried on 20 patients. It "works" for 14 and "doesn't work" for 6. What is your new belief that the drug works?

---

Bayesian: $X \sim \text{Beta}$

| Prior: | Interpretation: |
|---|---|
| $X \sim \text{Beta}(a = 81, b = 21)$ | 80 successes / 100 trials |
| $X \sim \text{Beta}(a = 9, b = 3)$ | 8 successes / 10 trials |
| $X \sim \text{Beta}(a = 5, b = 2)$ | 4 successes / 5 trials |

# Beta Example

Before being tested, a medicine is believed to "work" about 80% of the time. The medicine is tried on 20 patients. It "works" for 14 and "doesn't work" for 6. What is your new belief that the drug works?

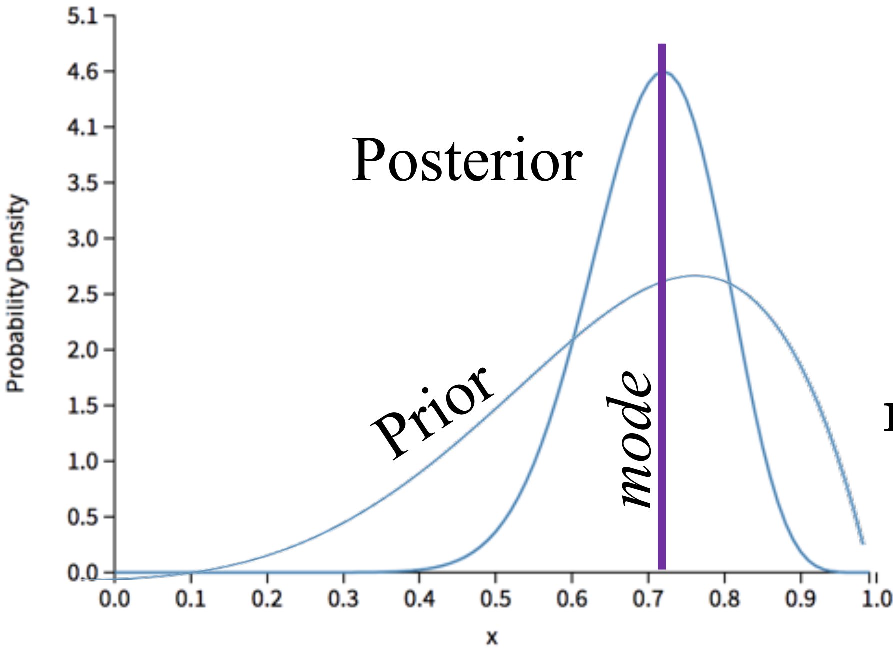
---

$$\text{Bayesian:} \quad X \sim \text{Beta}$$

$$\text{Prior:} \quad X \sim \text{Beta}(a = 5, b = 2)$$

$$\text{Posterior:} \quad X \sim \text{Beta}(a = 5 + 14, b = 2 + 6)$$

$$\sim \text{Beta}(a = 19, b = 8)$$




$$E[X] = \frac{a}{a+b} = \frac{19}{19+8} \approx 0.70$$

$$\text{mode}(X) = \frac{a-1}{a+b-2}$$

$$= \frac{19}{18+7} \approx 0.72$$

# Multi Armed Bandit

# Multi Armed Bandit

Drug A

Drug B

Which one do you give to a patient?

# Lets Play!

Drug A

Drug B

Which one do you give to a patient?

# Lets Play!

sim.py

```python
import pickle
import random

def main():
  X1, X2  = pickle.load(open('probs.pkl', 'rb'))

  print("Welcome to the drug simulator. There are two drugs")

  while True:
    choice = getChoice()
    prob = X1 if choice == "a" else X2
    success = bernoulli(prob)
    if success:
      print('Success. Patient lives!')
    else:
      print('Failure. Patient dies!')
    print('')

```

# Optimal Decision Making

You try drug B, 5 times. It is successful 2 times.

If you had a uniform prior, what is your posterior belief about the likelihood of success?

---

2 successes

3 failures

$$X \sim \text{Beta}(a = 3, b = 4)$$

# Optimal Decision Making

You try drug B, 5 times. It is successful 2 times.
$X$ is the probability of success.

$$X \sim \text{Beta}(a = 3, b = 4)$$

---

What is expectation of X?

$$E[X] = \frac{a}{a+b} = \frac{3}{3+4} \approx 0.43$$

# Optimal Decision Making

You try drug B, 5 times. It is successful 2 times.
$X$ is the probability of success.

$$X \sim \text{Beta}(a = 3, b = 4)$$

---

What is the probability that $X > 0.6$

$$P(X > 0.6) = 1 - P(X < 0.6) = 1 - F_X(0.6)$$

Wait! What's the Beta CDF??

```
stats.beta.cdf(x, a, b)
```

$$P(X > 0.6) = 1 - F_X(0.6) = 0.1792$$

# How can we choose a drug?

We have a big problem! There is a tradeoff between:

**Exploration** – trying new drugs in order to find out if they work well

**Exploitation** – giving the drugs that have been proven to be successful.

**Thompson Sampling** addresses this issue by choosing drugs probabilistically based on your current beliefs (higher chance of choosing a drug the more strongly you believe it works well).

You will learn more on PS5, but check out the Wikipedia article if you're impatient!

# Which Tycho are you today?

Stretch!

# Central Theorem

Noah Arthurs
CS109, Stanford University

As n approaches infinity, The sum of n independent, identically distributed variables:

$$Y = \sum_{i=0}^{n} X_i$$

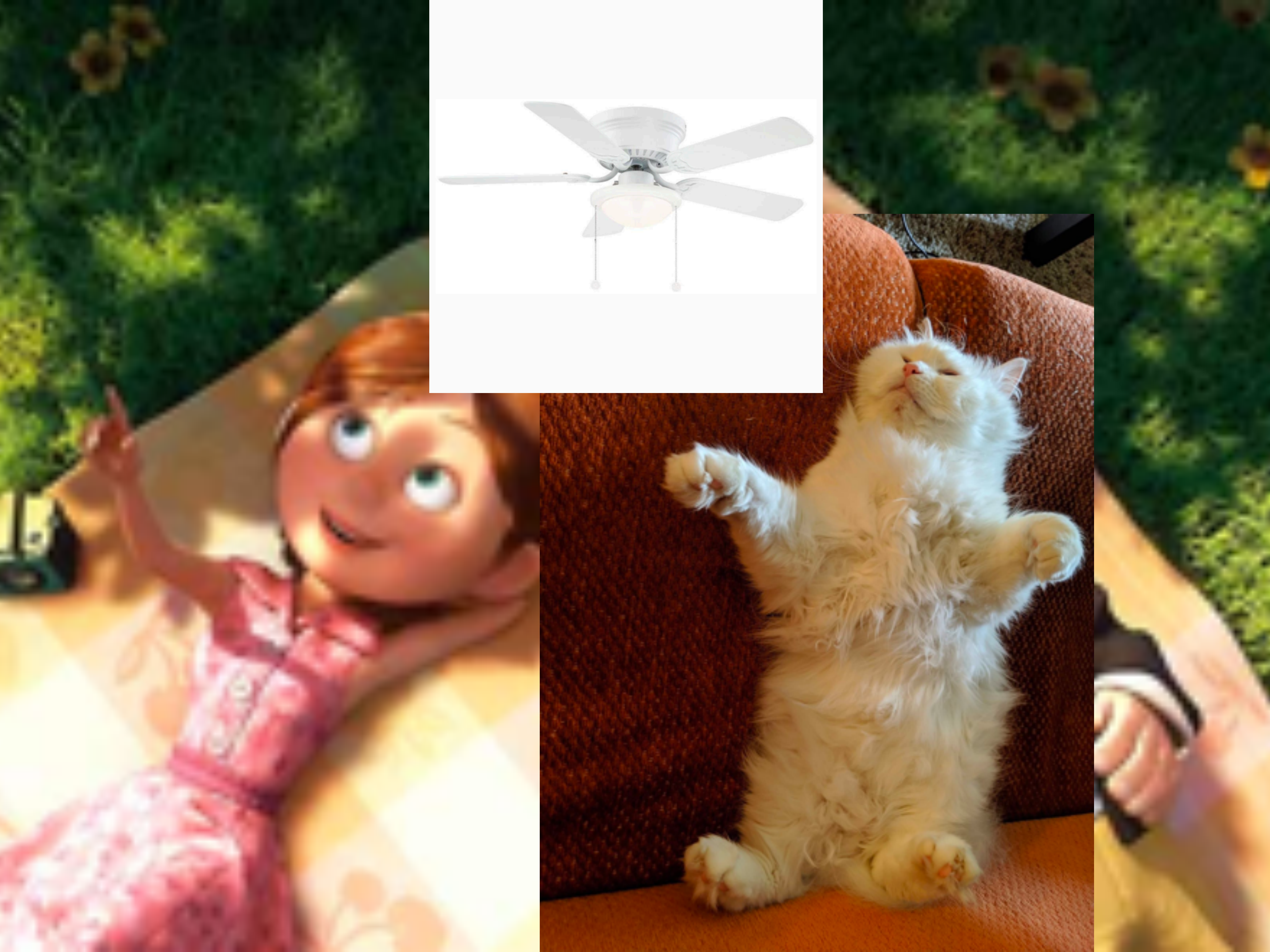
Is normally distributed:

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

---

where $\mu = E[X_i]$

$$\sigma^2 = \mathrm{Var}(X_i)$$



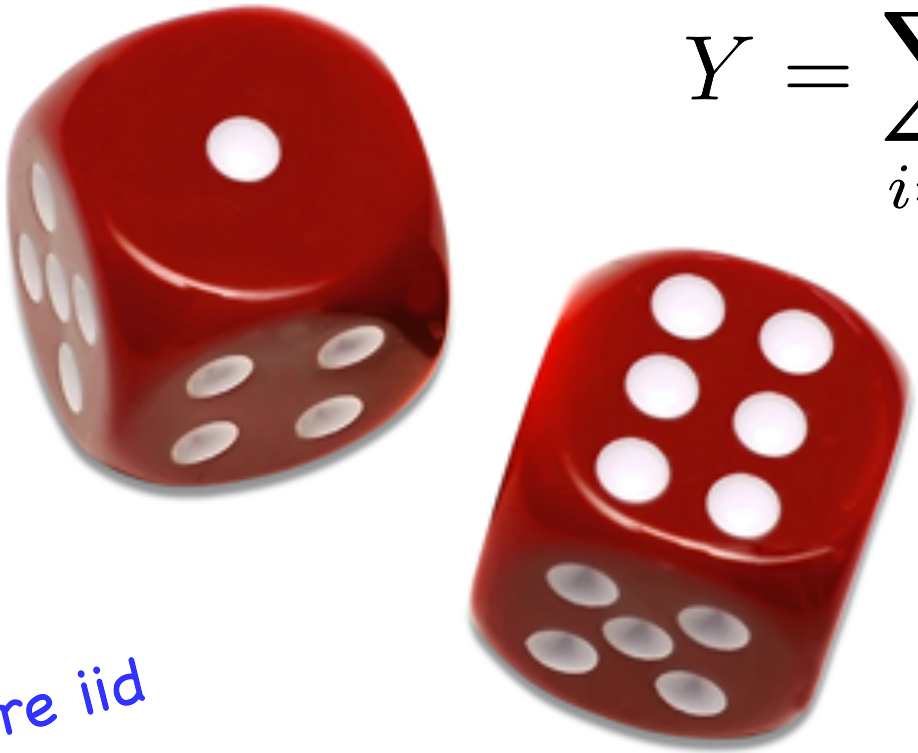
# IID Random Variables

- Consider *n* random variables $X_1$, $X_2$, ... $X_n$
  - $X_i$ are all independently and identically distributed (I.I.D.)
  - All have the same PMF (if discrete) or PDF (if continuous)
  - All have the same expectation
  - All have the same variance

IID　　　　iid

# Sum of Two Dice

$$Y = \sum_{i=0}^{2} X_i$$

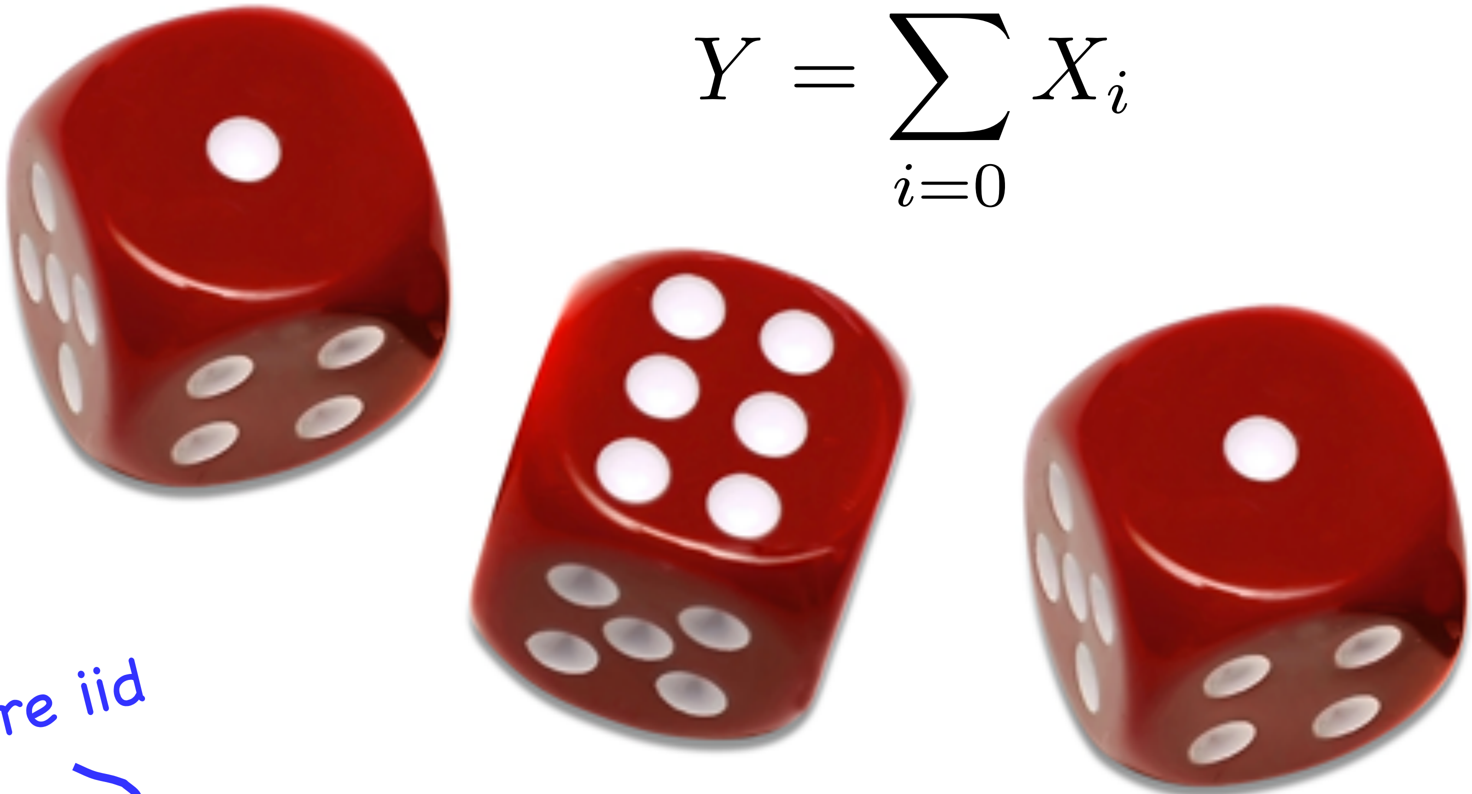


$X_i$ s are iid

$X_i$ is the outcome of dice roll $i$

# Sum of Three Dice

$$Y = \sum_{i=0}^{3} X_i$$



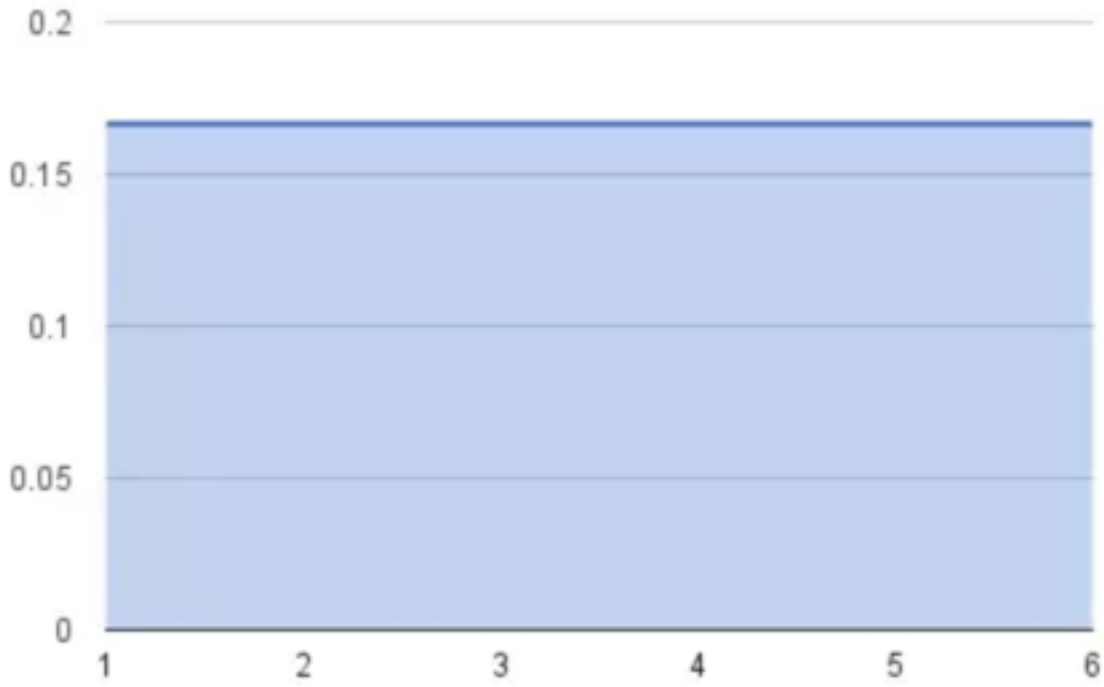
$X_i$ s are iid

$X_i$ is the outcome of dice roll $i$

Demo

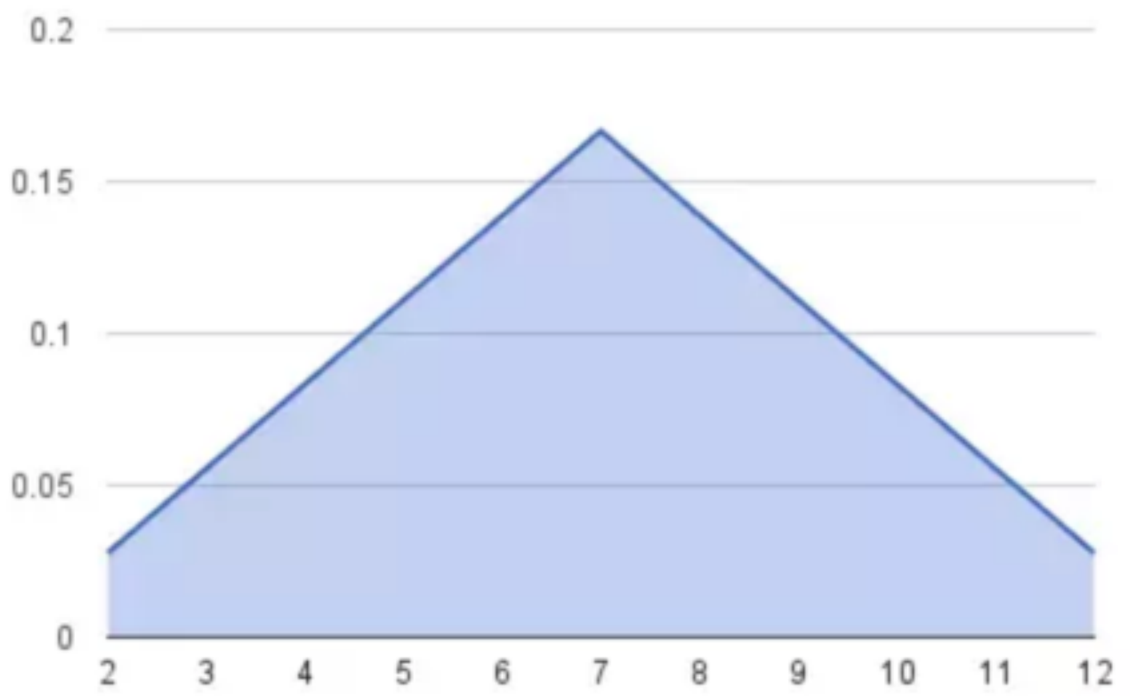
# C.L.T. Intuition

This is the PMF of the sum of one dice



# C.L.T. Intuition

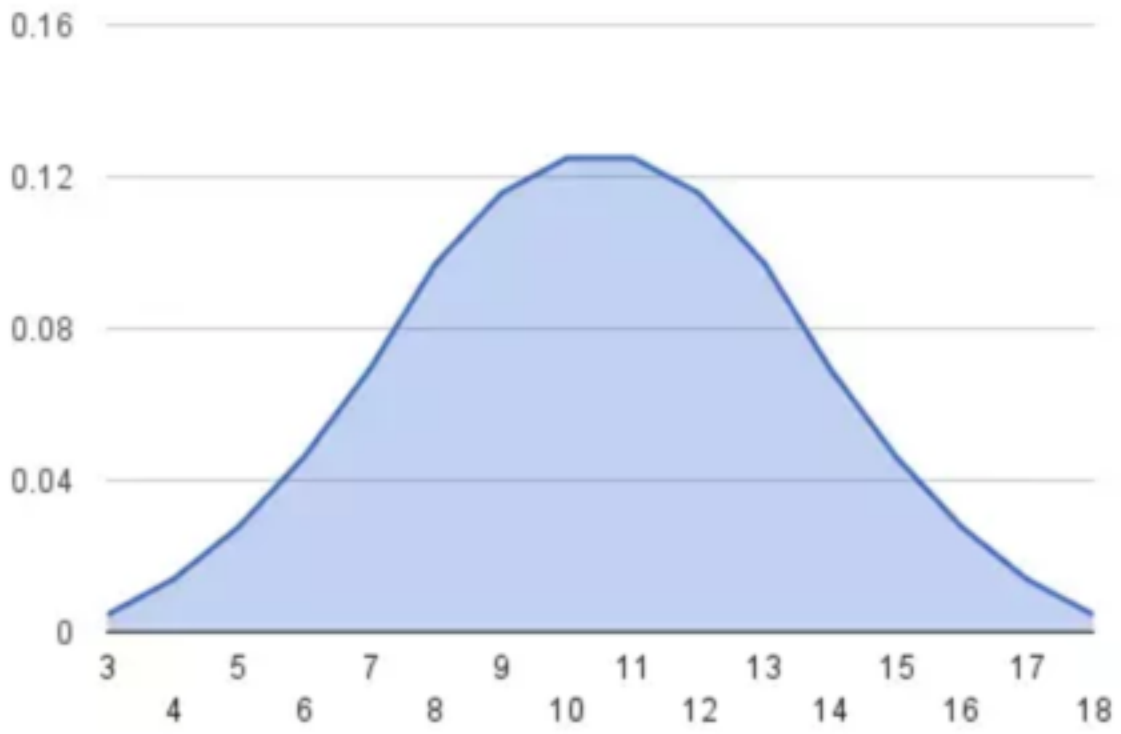
This is the PMF of the sum of two dice




Why is there more mass in the middle?

# C.L.T. Intuition

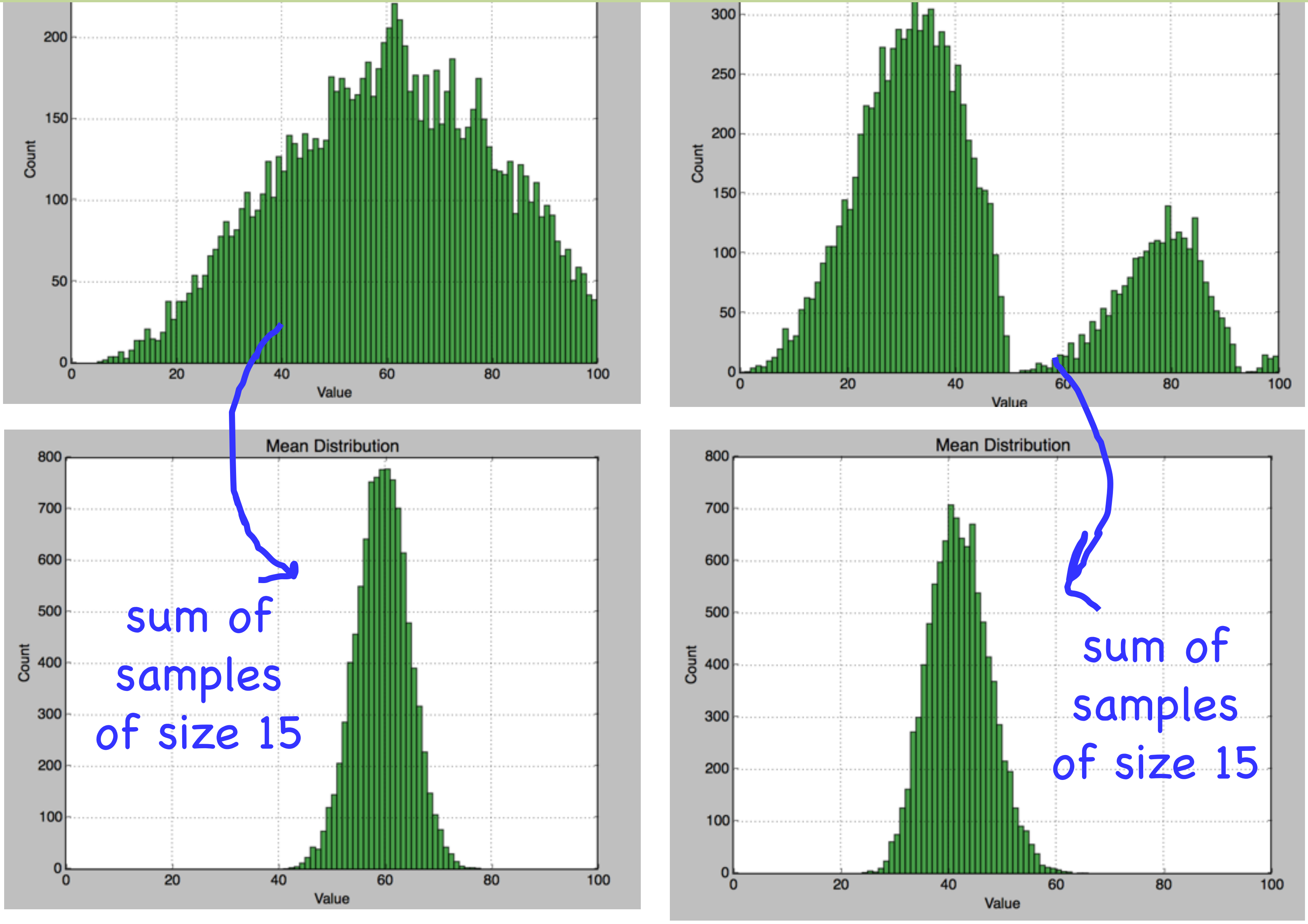
This is the PMF of the sum of three dice



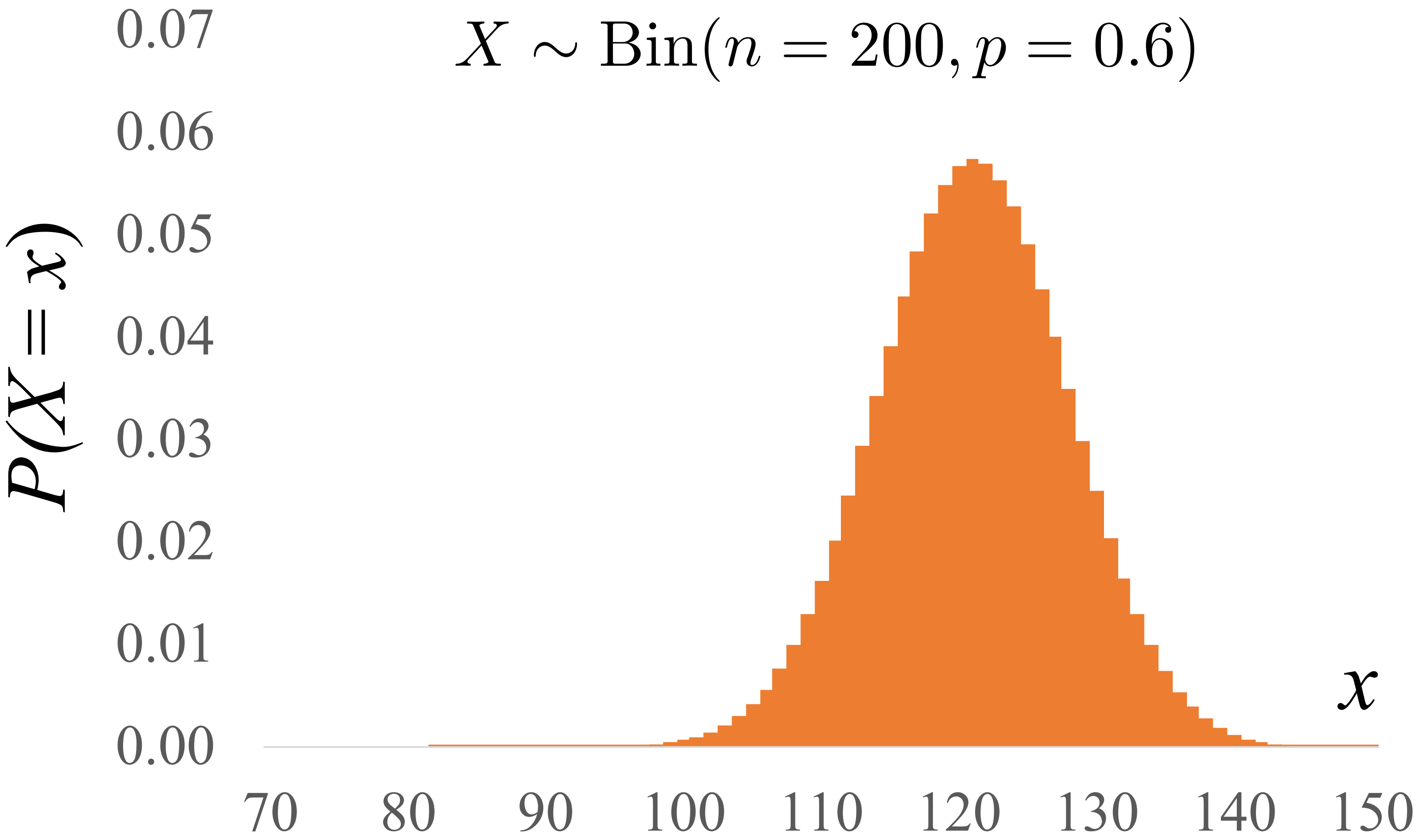

Why is there more mass in the middle?

# Other Functions?

# C.L.T. Explains This



# C.L.T. Explains This



# Binomial Approximation

- Consider I.I.D. Bernoulli variables $X_1$, $X_2$, ... With probability *p*
  - $X_i$ have $E[X_i] = p$ and $Var(X_i) = p(1-p)$

$$Y = \sum_{i=0}^{n} X_i$$

Y is the sum of the Bernoullis

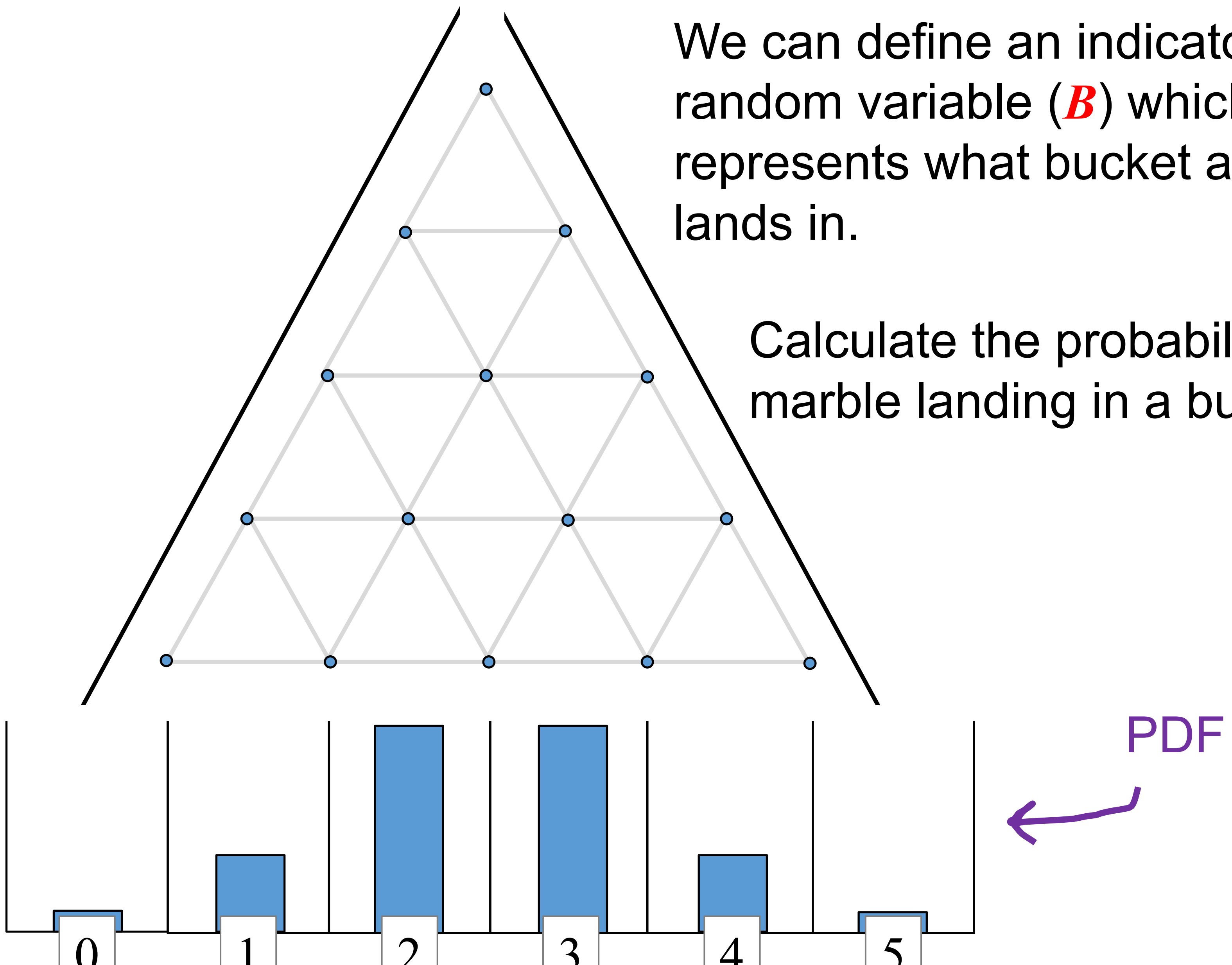
$$Y \sim N(n\mu, n\sigma^2) \quad \text{as } n \to \infty$$

Central Limit Theorem

$$Y \sim N(np, np(1-p))$$

Substituting mean and variance of Bernoilli

# C.L.T. Explains This




We can define an indicator random variable (***B***) which represents what bucket a marble lands in.

Calculate the probability of a marble landing in a bucket.

As n approaches infinity, The sum of n independent, identically distributed variables:

$$Y = \sum_{i=0}^{n} X_i$$

Is normally distributed:

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

---

where $\mu = E[X_i]$

$$\sigma^2 = \mathrm{Var}(X_i)$$

Since n is never actually infinite, the CLT is always an approximation. It's a very good one though!

# On the Proof of the CLT

- The proof of the CLT uses the Fourier transform of the probability mass of the sample distance from the mean, divided by standard deviation, and shows that this approaches an exponential function in the limit:

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

- That exponential function is in turn the Fourier transform of the Standard Normal. The Fourier transform of a probability density function is called a *Characteristic Function*.

- The proof is beyond the scope of CS109.

# Central Limit Theorem in the Real World

- CLT is why some things in "real world" appear Normally distributed
  - Many quantities are sum of independent variables
  - Exams scores
    - Sum of individual problems on the SAT
    - Why does the CLT not apply to our midterm?
  - Election polling
    - Ask 100 people if they will vote for candidate X ($p_1$ = # "yes"/100)
    - Repeat this process with different groups to get $p_1, \ldots, p_n$
    - Will have a normal distribution over $p_i$
    - Can produce a "confidence interval"
      - How likely is it that estimate for true p is correct

# What about other functions?

Sum of iid? Normal

Average of iid?

Max of iid?

# The Central Limit Theorem

- Consider I.I.D. random variables $X_1, X_2, \ldots$
  - $X_i$ have same distribution with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$
  - Let: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - Central Limit Theorem:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \text{as} \quad n \to \infty$$

Demo

http://onlinestatbook.com/stat_sim/sampling_dist/

# But Wait! There is More

- Consider I.I.D. random variables $X_1$, $X_2$, ...
  - $X_i$ have distribution *F* with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$

$$\bar{X} = \frac{1}{n}\sum_{i}^{n} X_i \qquad Y = \sum_{i}^{n} X_i \qquad \bar{X} = \frac{1}{n} Y$$

---

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

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$ Linear transform of a normal

By the Central Limit Theorem, the sample mean of IID variables are distributed normally.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

# What about other functions?

Sum of iid? Normal

Average of iid? Normal

Max of iid?

# What about other functions?

Sum of iid? Normal

Average of iid? Normal

Max of iid? Gumbel

See Fisher Trippett Gnedenko Theorem

# Once Upon a Time…


THE
DOCTRINE
OF
CHANCES:
OR,
A Method of Calculating the Probability
of Events in Play.

By *A. De Moivre.* F. R. S.

*LONDON:*
Printed by *W. Pearson*, for the Author. MDCCXVIII.

1733

Abraham De Moivre

# Once Upon a Time…

- History of the Central Limit Theorem
  - 1733: CLT for X ~ Ber(1/2) postulated by Abraham de Moivre

  - 1823: Pierre-Simon Laplace extends de Moivre's work to approximating Bin(n, p) with Normal

  - 1901: Aleksandr Lyapunov provides precise definition and rigorous proof of CLT
  - 2017: Tycho is born
    - The average of the sizes of the Siberian cats in a litter is normally distributed (probably)

# Estimating Clock Running Time

- Have new algorithm to test for running time
  - Mean (clock) running time: $\mu = t$ sec.
  - Variance of running time: $\sigma^2 = 4$ sec$^2$.
  - Run algorithm repeatedly (I.I.D. trials), measure time
    - How many trials s.t. estimated time = $t \pm 0.5$ with 95% certainty?
    - $X_i$ = running time of $i$-th run (for $1 \leq i \leq n$), $\bar{X}$ is the mean

---

$$0.95 = P(-0.5 < \bar{X} - t < 0.5)$$

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \sim N(t, \frac{4}{n})$$ By CLT

$$\bar{X} - t \sim N(0, \frac{4}{n})$$ By linear transform of a normal

$$0.95 = P(-0.5 < \bar{X} - t < 0.5) \qquad \bar{X} - t \sim N(0, \frac{4}{n})$$

---

$$0.95 = F_{\bar{X}-t}(0.5) - F_{\bar{X}-t}(-0.5)$$

$$= \Phi\Big(\frac{0.5 - 0}{\sqrt{4/n}}\Big) - \Phi\Big(\frac{-0.5 - 0}{\sqrt{4/n}}\Big)$$

$$= 2\phi(\frac{\sqrt{n}}{4}) - 1$$

$$0.95 \;= 2\phi(\frac{\sqrt{n}}{4}) - 1$$

---

$$0.975 = \phi(\frac{\sqrt{n}}{4})$$

$$\phi^{-1}(0.975) = \frac{\sqrt{n}}{4}$$

$$1.96 = \frac{\sqrt{n}}{4}$$

$$n = 61.4$$


William Sealy Gosset
(aka Student)

It’s play time!

# Sum of Dice

- You will roll 10 6-sided dice ($X_1, X_2, \ldots, X_{10}$)
  - $X$ = total value of all 10 dice = $X_1 + X_2 + \ldots + X_{10}$
  - Win if: $X \leq 25$ or $X \geq 45$
  - Roll!
- And now the truth (according to the CLT)…

# Sum of Dice

- You will roll 10 6-sided dice ($X_1, X_2, \ldots, X_{10}$)
  - $X$ = total value of all 10 dice = $X_1 + X_2 + \ldots + X_{10}$
  - Win if: $X \le 25$ or $X \ge 45$

---

- Recall CLT: $$X = \sum_i^n X_i \to N(n\mu, n\sigma^2) \qquad \text{As } n \to \infty$$
  - Determine P($X \le 25$ or $X \ge 45$) using CLT:

$$\mu = E[X_i] = 3.5 \qquad \sigma^2 = \mathrm{Var}(X_i) = \frac{35}{12} \qquad X \approx N(35, 29.2)$$

$$1 - P(25.5 < X < 44.5) = 1 - P\left(\frac{25.5 - 35}{\sqrt{29.2}} < Z < \frac{44.5 - 35}{\sqrt{29.2}}\right)$$

$$\approx 1 - (2\Phi(1.76) - 1) \approx 2(1 - 0.9608) = 0.0784$$

# Wonderful Form of Cosmic Order

I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the ”[Central limit theorem]". The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.

-Sir Francis Galton