18: Central Limit Theorem

Lisa Yan

November 1, 2019

Beta random variable

<u>def</u> An **Beta** random variable X is defined as follows:

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

$$a > 0, b > 0$$

Support of X: $(0, 1)$

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

where $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$, normalizing constant

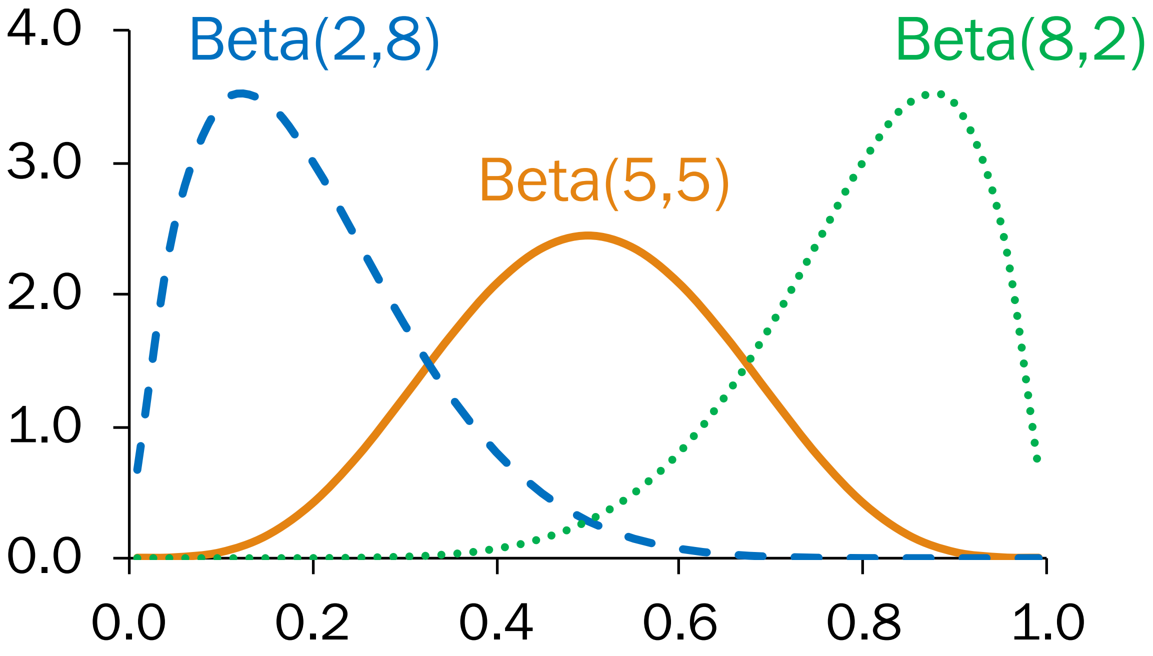
Expectation $\quad E[X] = \frac{a}{a+b}$

Variance $\quad \text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$

Beta is a distribution for probabilities.

CS109 focus: Beta where a, b both positive integers

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



If a, b are positive integers, Beta parameters a, b *could* come from an experiment:

$$a = \text{"successes"} + 1$$
$$b = \text{"failures"} + 1$$

Back to flipping coins

- Start with a $X \sim \text{Uni}(0,1)$ over probability
- Observe $n = 7$ successes and $m = 1$ failures
- Your new belief about the probability of X is:

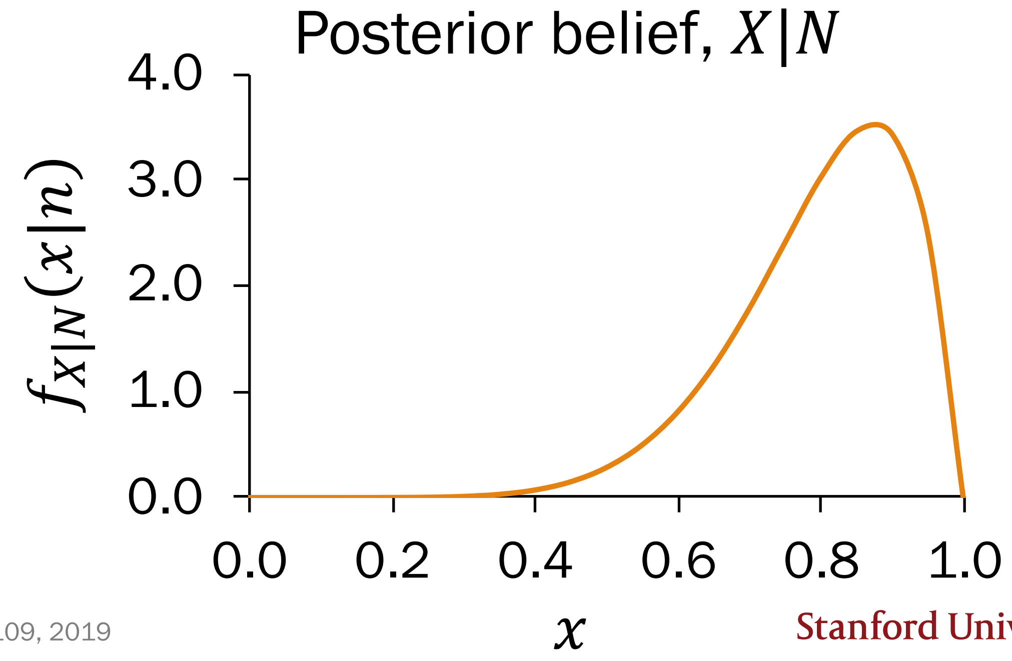
$$f_{X|N}(x|n) = \frac{1}{c}\, x^{7}(1-x)^{1}, \text{where } c = \int_0^1 x^{7}(1-x)^{1} dx$$

Posterior belief, $X|N$:

Beta$(a = 8, b = 2)$

$$f_{X|N}(x|n) = \frac{1}{c}\, x^{8-1}(1-x)^{2-1}$$

Beta$(a = n+1, b = m+1)$



Understanding Beta

- Start with a $X \sim \text{Uni}(0,1)$ over probability
- Observe n successes and m failures
- Your new belief about the probability of X is:

$$X|N \sim \text{Beta}(a = n + 1, b = m + 1)$$

Understanding Beta

- Start with a $X \sim \text{Uni}(0,1)$ over probability
- Observe n successes and m failures
- Your new belief about the probability of X is:

$$X|N \sim \text{Beta}(a = n + 1, b = m + 1)$$

Check this out:

Beta$(a = 1, b = 1)$ has PDF:

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} = \frac{1}{B(a,b)} x^0(1-x)^0 = \frac{1}{\int_0^1 1dx}$$

where $0 < x < 1$

So our prior $X \sim \text{Beta}(a = 1, b = 1)$!

If the prior is a Beta…

Let X be our random variable for probability of success and N

- If our prior belief about X is beta: $X \sim \text{Beta}(a, b)$

likelihood

- …and if we observe n successes and m failures: $N|X \sim \text{Bin}(n + m, x)$

- …then our posterior belief about X is also beta. $X|N \sim \text{Beta}(a + n, b + m)$

This is the main takeaway of Beta.

If the prior is a Beta…

Let X be our random variable for probability of success and N

- If our **prior belief** about X is beta: $X \sim \text{Beta}(a, b)$

likelihood

- …and if we observe n successes and m failures: $N|X \sim \text{Bin}(n+m, x)$

- …then our **posterior belief** about X is also beta. $X|N \sim \text{Beta}(a+n, b+m)$

Proof:

$$f_{X|N}(x|n) = \frac{p_{N|X}(n|x) f_X(x)}{p_N(n)} = \frac{\binom{n+m}{n} x^n (1-x)^m \cdot \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}}{p_N(n)}$$

constants that don't depend on x

$$= C \cdot x^n (1-x)^m \cdot x^{a-1}(1-x)^{b-1}$$

$$= C \cdot x^{n+a-1}(1-x)^{m+b-1}$$ ✅

If the prior is a Beta…

Let X be our random variable for probability of success and N

- If our prior belief about X is beta: $X \sim \text{Beta}(a, b)$
- likelihood ...and if we observe n successes and m failures: $N|X \sim \text{Bin}(n+m, x)$
- ...then our posterior belief about X is also beta. $X|N \sim \text{Beta}(a+n, b+m)$

Beta is a conjugate distribution.

- Prior and posterior parametric forms are the same
- Practically, conjugate means easy update:
  Add number of "heads" and "tails" seen
  to Beta parameter.

If the prior is a Beta…

Let X be our random variable for probability of success and N

- If our **prior belief** about X is beta: $X \sim \text{Beta}(a, b)$

likelihood

- …and if we observe n successes and m failures: $N|X \sim \text{Bin}(n + m, x)$
- …then our **posterior belief** about X is also beta. $X|N \sim \text{Beta}(a + n, b + m)$

You can set the prior to reflect how biased you think the coin is apriori.

- This is a subjective probability!
- $X \sim \text{Beta}(a, b)$: have seen $(a + b - 2)$ imaginary trials, where $(a - 1)$ are heads, $(b - 1)$ tails
- Then $\text{Beta}(1, 1) = \text{Uni}(0, 1)$ means we haven't seen any imaginary trials

If the prior is a Beta…

Let X be our random variable for probability of success and N

- If our **prior belief** about X is beta: $X \sim \text{Beta}(a, b)$

likelihood

- …and if we observe n successes and m failures: $N|X \sim \text{Bin}(n+m, x)$
- …then our **posterior belief** about X is also beta. $X|N \sim \text{Beta}(a+n, b+m)$

Prior $\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$

Posterior $\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

👉 This is the main takeaway of Beta.

The enchanted die

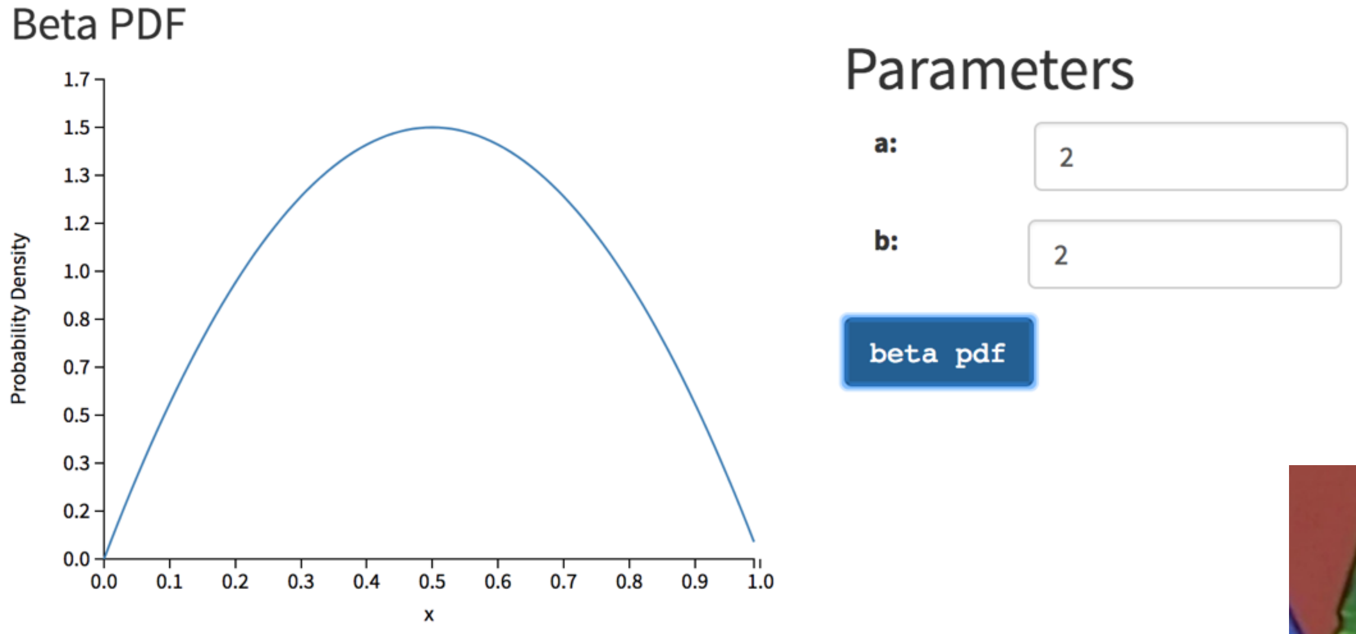
Prior	$\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$
Posterior	$\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

Let X be the probability of rolling a 6 on Lisa's die.

- Prior: Imagine 5 die rolls where only 6 showed up
- Observation: roll it a few times…

What is the updated distribution of X after our observation?

Check out the demo!



Medicinal Beta

- Before being tested, a medicine is believed to "work" 80% of the time.
- The medicine is tried on 20 patients.
- It "works" for 12, "doesn't work" for 8.

What is your new belief that the drug "works"?

Frequentist

Let p be the probability your drug works.

$$p \approx \frac{12}{20} = 0.6$$

Bayesian

A frequentist view will not incorporate prior/expert belief about probability.

Medicinal Beta

- Before being tested, a medicine is believed to "work" 80% of the time.
- The medicine is tried on 20 patients.
- It "works" for 12, "doesn't work" for 8.

What is your new belief that the drug "works"?

Frequentist

Let p be the probability your drug works.

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

Bayesian

Let X be the probability your drug works.

X is a random variable.

Medicinal Beta

Prior	$\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$
Posterior	$\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

- Before being tested, a medicine is believed to "work" 80% of the time.
- The medicine is tried on 20 patients.
- It "works" for 12, "doesn't work" for 8.

What is your new belief that the drug "works"? (Bayesian interpretation)

What is the prior distribution of X? (select all that apply)

A. $X \sim \text{Beta}(1, 1) = \text{Uni}(0, 1)$

B. $X \sim \text{Beta}(81, 101)$

C. $X \sim \text{Beta}(80, 20)$

D. $X \sim \text{Beta}(81, 21)$

E. $X \sim \text{Beta}(5, 2)$

Medicinal Beta

Prior	$\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$
Posterior	$\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

- Before being tested, a medicine is believed to "work" 80% of the time.
- The medicine is tried on 20 patients.
- It "works" for 12, "doesn't work" for 8.

What is your new belief that the drug "works"? (Bayesian interpretation)

What is the prior distribution of X? (select all that apply)

A. $X \sim \text{Beta}(1,1) = \text{Uni}(0,1)$

B. $X \sim \text{Beta}(81,101)$

C. $X \sim \text{Beta}(80,20)$

D. $X \sim \text{Beta}(81,21)$ Interpretation: 80 successes / 100 imaginary trials

E. $X \sim \text{Beta}(5,2)$ Interpretation: 4 successes / 5 imaginary trials

(you can choose either; we choose E on next slide)

Medicinal Beta

Prior	$\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$
Posterior	$\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

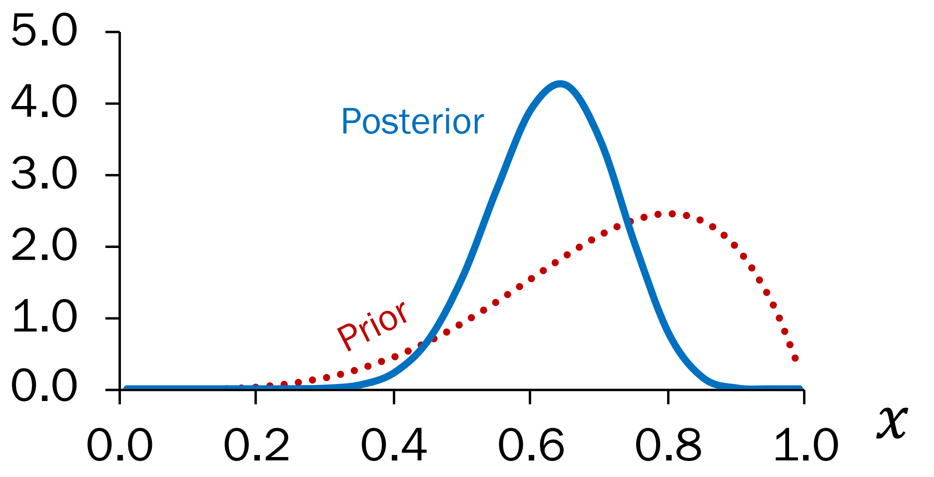
- Before being tested, a medicine is believed to "work" 80% of the time.
- The medicine is tried on 20 patients.
- It "works" for 12, "doesn't work" for 8.

What is your new belief that the drug "works"? (Bayesian interpretation)

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

Posterior: $X \sim \text{Beta}(a = 5 + 12, b = 2 + 8)$

$\sim \text{Beta}(a = 17, b = 10)$



Medicinal Beta

Prior	$\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$
Posterior	$\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

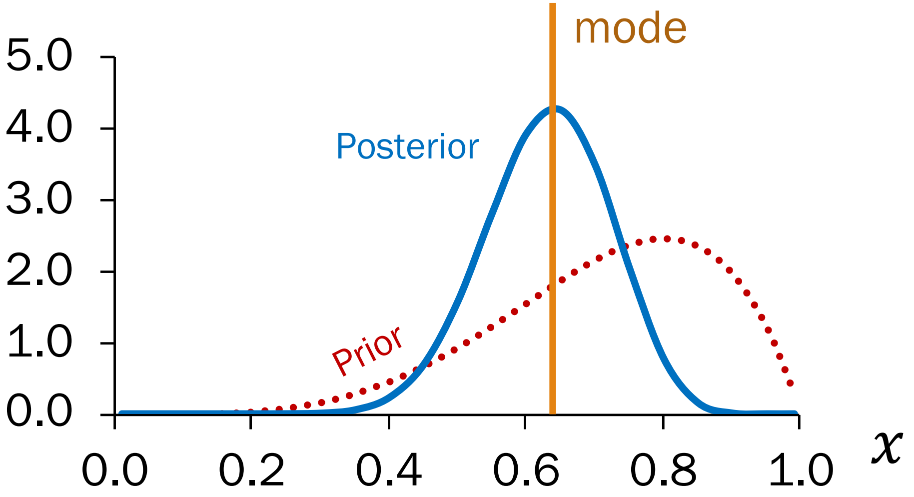
- Before being tested, a medicine is believed to “work” 80% of the time.
- The medicine is tried on 20 patients.
- It “works” for 12, “doesn’t work” for 8.

What is your new belief that the drug “works”?

(Bayesian interpretation)

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

Posterior: $X \sim \text{Beta}(a = 5 + 12, b = 2 + 8)$
$\sim \text{Beta}(a = 17, b = 10)$



What do you report to pharmacists?

A. Expectation of posterior
B. Mode of posterior
C. Distribution of posterior
D. Nothing

Medicinal Beta

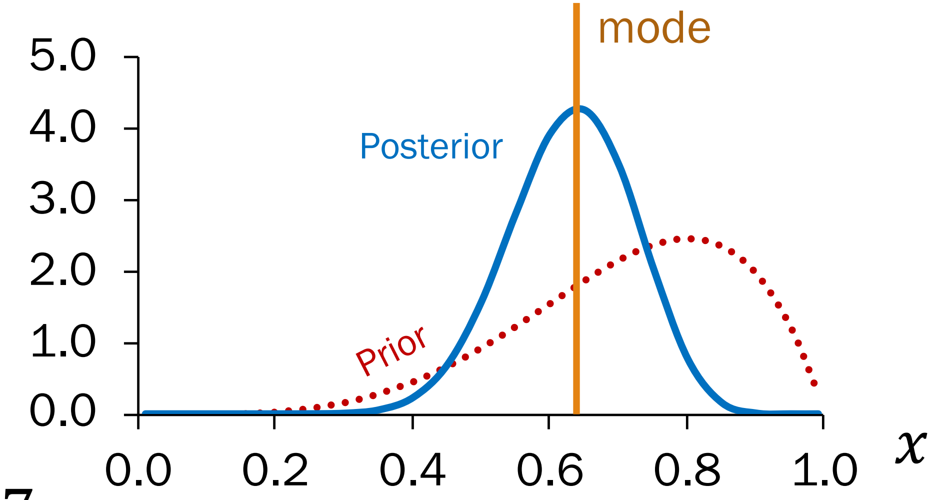
Prior	$\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$
Posterior	$\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

- Before being tested, a medicine is believed to "work" 80% of the time.
- The medicine is tried on 20 patients.
- It "works" for 12, "doesn't work" for 8.

What is your new belief that the drug "works"? (Bayesian interpretation)

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

Posterior: $X \sim \text{Beta}(a = 5 + 12, b = 2 + 8)$
$\sim \text{Beta}(a = 17, b = 10)$



What do you report to pharmacists?

A. Expectation of posterior
B. Mode of posterior
C. Distribution of posterior
D. Nothing

$$E[X] = \frac{a}{a+b} = \frac{17}{17+10} \approx 0.63$$

$$\text{mode}(X) = \frac{a-1}{a+b-2} = \frac{16}{16+9} = 0.64$$

Food for thought

In this lecture:

$$Y \sim \text{Ber}(p)$$

If we don't know the **parameter** p,
Bayesian statisticians will:

- Treat the parameter as a random variable X with a Beta prior distribution
- Perform an experiment
- Based on experiment outcomes, update the posterior distribution of X

Food for thought:

Any parameter for a "parameterized" random variable can be thought of as a random variable.

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

Today's plan

Finish Beta

Central Limit Theorem (CLT)

CLT exercises

(silent drumroll)

Central Limit Theorem

Consider n independent and identically distributed (i.i.d.) variables $X_1, X_2, \ldots, X_n$ with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$.

As $n \to \infty$

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

The sum of n i.i.d. random variables is normally distributed with mean $n\mu$ and variance $n\sigma^2$.

True happiness

Central Limit Theorem

Consider n **independent and identically distributed (i.i.d.)** variables $X_1, X_2, \ldots, X_n$ with $E[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$.

As $n \to \infty$

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

The sum of n **i.i.d.** random variables is normally distributed with mean $n\mu$ and variance $n\sigma^2$.

Quick check

What dimensions are the following random variables?

1. X_1
2. $(X_1, X_2, \dots, X_n)$
3. $\sum_{i=1}^{n} X_i$
4. $\frac{1}{n}\sum_{i=1}^{n} X_i$

A. 1-D random variable
B. n -D random variable (a vector)
C. not a random variable

Quick check

What dimensions are the following random variables?

1. X_1 (A)
2. $(X_1, X_2, \dots, X_n)$ (B) (aka a **sample**)
3. $\sum_{i=1}^{n} X_i$ (A)
4. $\frac{1}{n}\sum_{i=1}^{n} X_i$ (A) (aka the **sample mean**)

A. 1-D random variable
B. n -D random variable (a vector)
C. not a random variable

Central Limit Theorem

Consider n independent and identically distributed (i.i.d.) variables $X_1, X_2, \dots, X_n$ with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$.

As $n \to \infty$

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

The sum of n **i.i.d.** random variables is normally distributed with mean $n\mu$ and variance $n\sigma^2$.

i.i.d. random variables

Consider n variables $X_1, X_2, \dots, X_n$.

$X_1, X_2, \dots, X_n$ are **independent and identically distributed** if

- $X_1, X_2, \dots, X_n$ are independent, and
- All have the same PMF (if discrete) or PDF (if continuous).

$\Rightarrow E[X_i] = \mu$ for $i = 1, \dots, n$

$\Rightarrow \text{Var}(X_i) = \sigma^2$ for $i = 1, \dots, n$

Same thing: **i.i.d.** **iid** **IID**

Side note: Multiple random variables $X_1, X_2, \dots, X_n$ are independent iff

$$P(X_1 \le x_1, X_2 \le x_2, \dots, X_k \le x_k) = \prod_{i=1}^{k} P(X_i \le x_i) \quad \text{for all subsets } X_1, \dots, X_k$$

Quick check

Are $X_1, X_2, \ldots, X_n$ i.i.d. with the following distributions?

1. $X_i \sim \text{Exp}(\lambda)$, X_i independent

2. $X_i \sim \text{Exp}(\lambda_i)$, X_i independent

3. $X_i \sim \text{Exp}(\lambda)$, $X_1 = X_2 = \cdots = X_n$

4. $X_i \sim \text{Bin}(n_i, p)$, X_i independent

Quick check

Are $X_1, X_2, \ldots, X_n$ i.i.d. with the following distributions?

1. $X_i \sim \text{Exp}(\lambda)$, X_i independent ✅
2. $X_i \sim \text{Exp}(\lambda_i)$, X_i independent (unless λ_i equal)
3. $X_i \sim \text{Exp}(\lambda)$, $X_1 = X_2 = \cdots = X_n$ dependent: $X_1 = X_2 = \cdots = X_n$
4. $X_i \sim \text{Bin}(n_i, p)$, X_i independent (unless n_i equal)
   Note underlying Bernoulli RVs are i.i.d.! 🤔

Central Limit Theorem

Consider n independent and identically distributed (i.i.d.) variables $X_1, X_2, \ldots, X_n$ with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$.

As $n \to \infty$

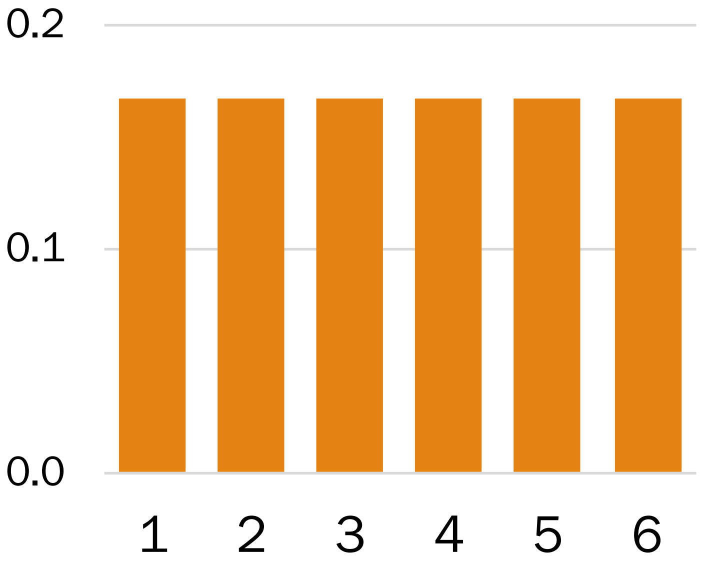
$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

The sum of n i.i.d. random variables is normally distributed with mean $n\mu$ and variance $n\sigma^2$.

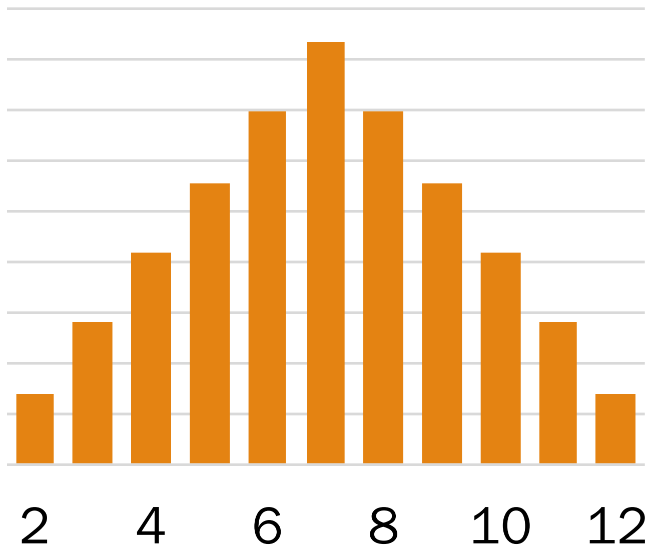
(demo)

Sum of dice rolls

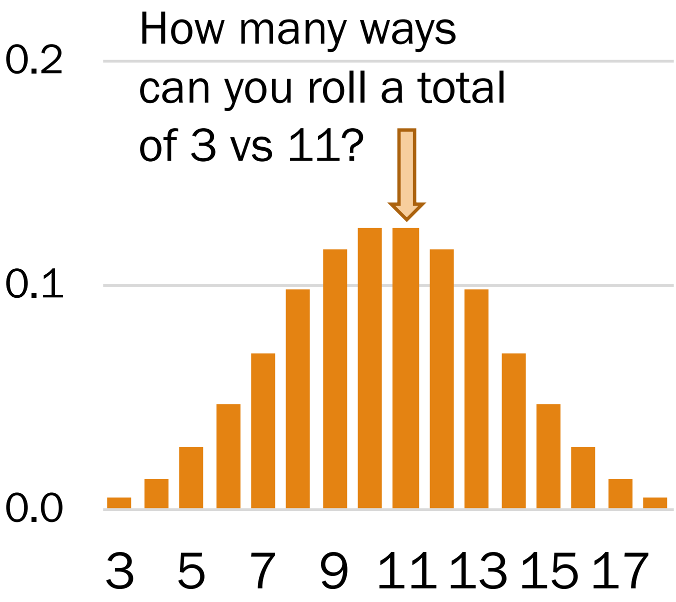
Roll n independent dice. Let X_i be the outcome of roll i. X_i are i.i.d.



$$\sum_{i=1}^{1} X_i$$ Sum of 1 die roll



$$\sum_{i=1}^{2} X_i$$ Sum of 2 die rolls



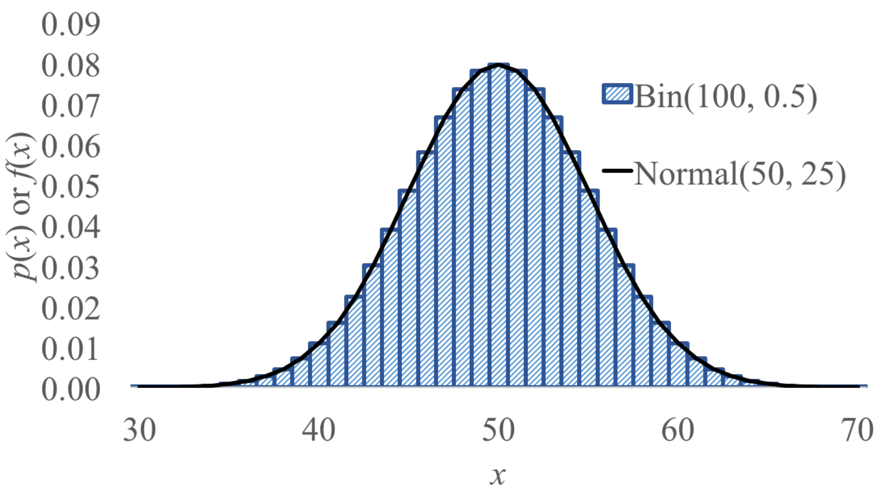
$$\sum_{i=1}^{3} X_i$$ Sum of 3 die rolls

CLT explains a lot

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

As $n \to \infty$

The sum of n **i.i.d.** random variables is normally distributed with mean $n\mu$ and variance $n\sigma^2$.



Normal approximation of Binomial

Sum of i.i.d. Bernoulli RVs ≈ Normal

Proof:

Let $X_i \sim \text{Ber}(p)$ for $i = 1, \ldots, n$, where X_i are i.i.d.

$E[X_i] = p$, $\text{Var}(X_i) = p(1-p)$

$$X = \sum_{i=1}^{n} X_i \qquad (X \sim \text{Bin}(n, p))$$

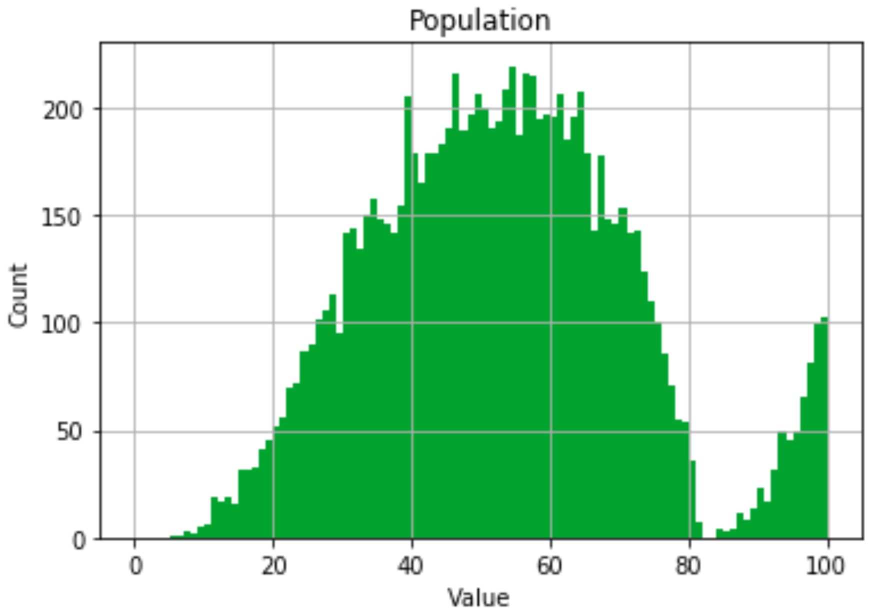
$$X \sim \mathcal{N}(n\mu, n\sigma^2) \qquad (\text{CLT, as } n \to \infty)$$

$$X \sim \mathcal{N}(np, np(1-p)) \qquad (\text{substitute mean, variance of Bernoulli})$$

CLT explains a lot

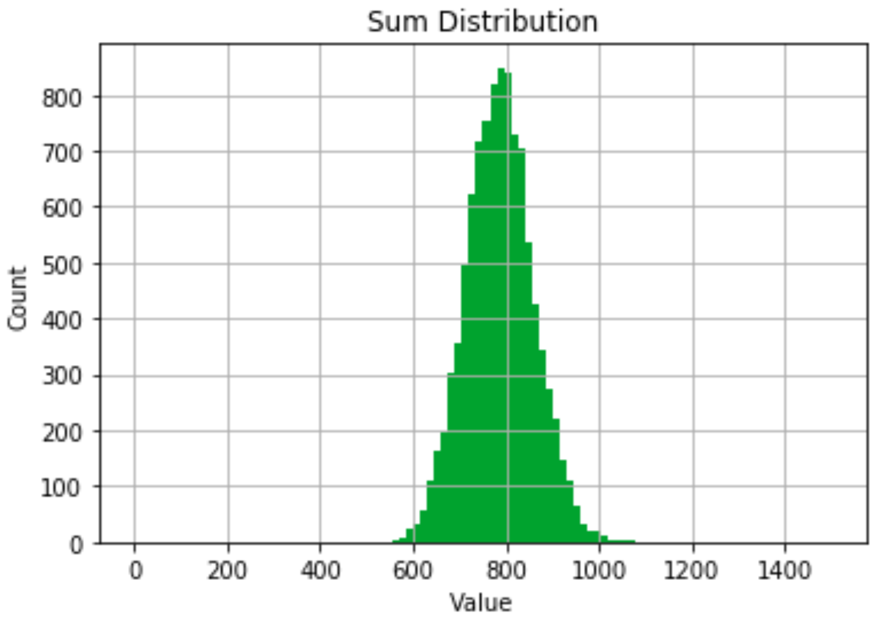
$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2) \quad \text{As } n \to \infty$$

The sum of n i.i.d. random variables is normally distributed with mean $n\mu$ and variance $n\sigma^2$.



Distribution of X_i

Sample of size 15, sum values



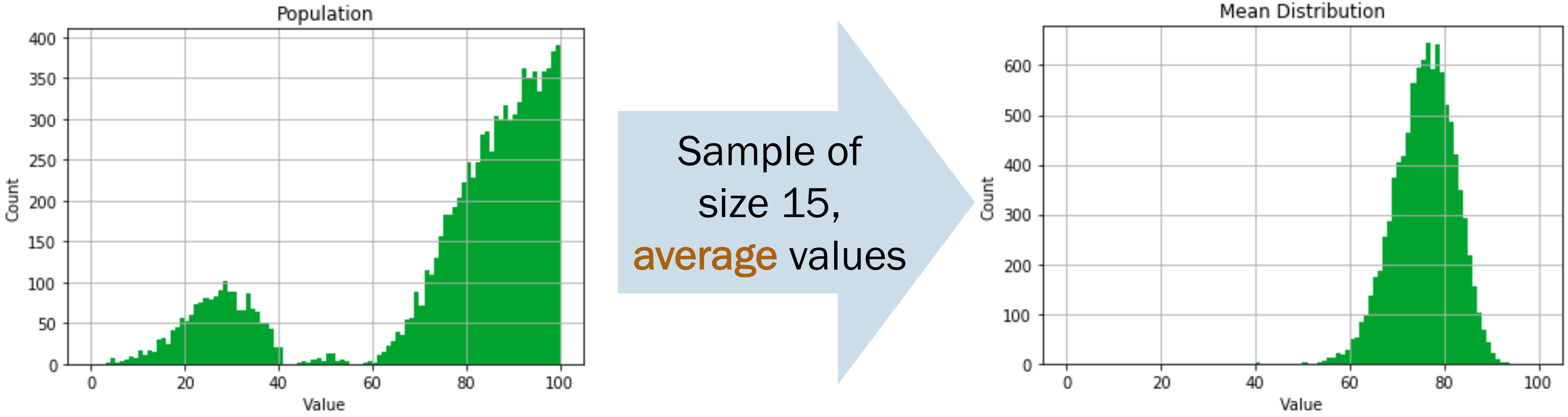
Distribution of $\sum_{i=1}^{15} X_i$

CLT explains a lot

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

As $n \to \infty$

The sum of n **i.i.d.** random variables is normally distributed with mean $n\mu$ and variance $n\sigma^2$.



Distribution of X_i

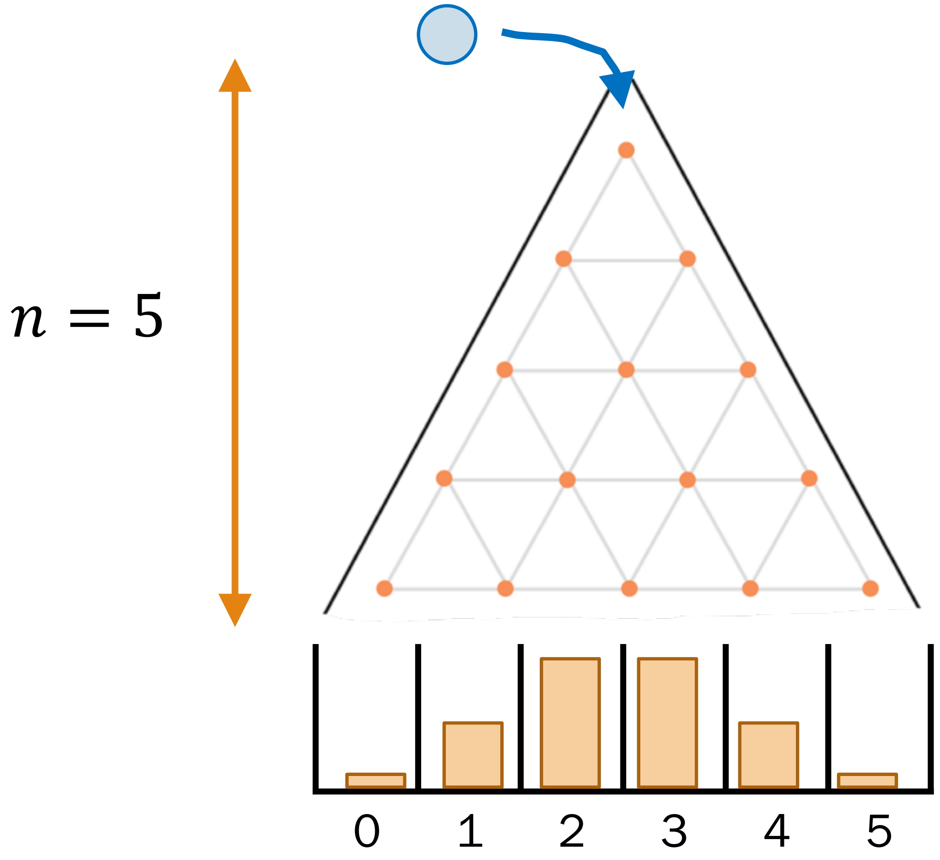
(sample mean)

Distribution of $\frac{1}{15}\sum_{i=1}^{15} X_i$

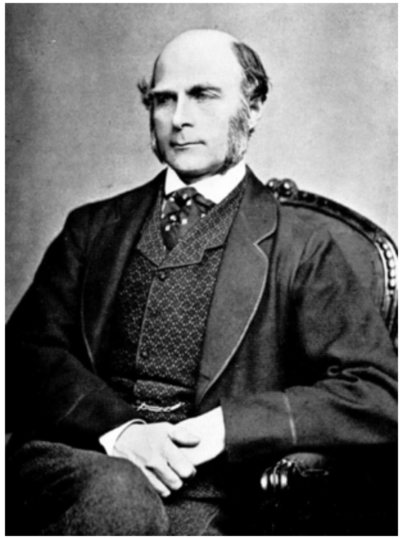
CLT explains a lot

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2) \quad \text{As } n \to \infty$$

The sum of n **i.i.d.** random variables is normally distributed with mean $n\mu$ and variance $n\sigma^2$.



Galton Board, by Sir Francis Galton (1822-1911)



Break for Friday/ announcements

Announcements

<u>Midterm exam</u>

It's done!

Grades:	Friday 11/1
Solutions:	Friday 11/1

<u>Problem Set 4</u>

Due:	Wednesday 11/6
Covers:	Up to Law of Total Expectation

Late day reminder: No late days permitted past last day of the quarter, 12/7

Proof of CLT

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$ As $n \to \infty$

The sum of n i.i.d. random variables is normally distributed with mean $n\mu$ and variance $n\sigma^2$.

Proof:

- The Fourier Transform of a PDF is called a characteristic function.
- Take the characteristic function of the probability mass of the sample distance from the mean, divided by standard deviation
- Show that this approaches an exponential function in the limit as $n \to \infty$: $$f(x) = e^{-\frac{x^2}{2}}$$
- This function is in turn the characteristic function of the Standard Normal, $Z \sim \mathcal{N}(0,1)$.

(this proof is beyond the scope of CS109)

Implications of CLT

Anything that is a sum/average of independent random variables is normal

…meaning in real life, many things are normally distributed:

- Movie ratings: averages of independent viewer scores
- Polling:
  - Ask 100 people if they will vote for candidate 1
    - p_1 = # "yes"/100
    - Sum of Bernoulli RVs (each person independently says "yes" w.p. p)
  - Repeat this process with different groups to get $p_1, \dots, p_n$ (different sample statistics)
  - Normal distribution over sample means p_k
  - Confidence interval: "How likely is it that an estimate for true p is close?"

What about other functions?

Let $X_1, X_2, \ldots, X_n$ i.i.d., where $E[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$. As $n \to \infty$:

$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$	Sum of i.i.d. RVs
?	Average of i.i.d. RVs (sample mean)
?	Max of i.i.d. RVs

What about other functions?

Let $X_1, X_2, \ldots, X_n$ i.i.d., where $E[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$. As $n \to \infty$:

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

Sum of i.i.d. RVs

?

Average of i.i.d. RVs

(sample mean)

?

Max of i.i.d. RVs

Distribution of sample mean

Let $X_1, X_2, \ldots, X_n$ i.i.d., where $E[X_i] = \mu, \text{Var}(X_i) = \sigma^2$. As $n \to \infty$:

Define: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (sample mean) $Y = \sum_{i=1}^{n} X_i$ (sum)

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

$\bar{X} = \frac{1}{n} Y$

$\bar{X} \sim \mathcal{N}(\,?\,,\,?\,)$ (Linear transform of a Normal)

A. $\bar{X} \sim \mathcal{N}(n\mu, n\sigma^2)$
B. $\bar{X} \sim \mathcal{N}(\frac{\mu}{n}, \frac{\sigma^2}{n})$
C. $\bar{X} \sim \mathcal{N}(\mu, \sigma^2)$
D. $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Distribution of sample mean

Let $X_1, X_2, \ldots, X_n$ i.i.d., where $E[X_i] = \mu, \text{Var}(X_i) = \sigma^2$. As $n \to \infty$:

Define: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (sample mean) $Y = \sum_{i=1}^{n} X_i$ (sum)

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

$\bar{X} = \frac{1}{n} Y$

$\bar{X} \sim \mathcal{N}(\ ?\ ,\ ?\ )$ (Linear transform of a Normal)

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

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

C. $\bar{X} \sim \mathcal{N}(\mu, \sigma^2)$

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

Distribution of sample mean

Let $X_1, X_2, \ldots, X_n$ i.i.d., where $E[X_i] = \mu, \text{Var}(X_i) = \sigma^2$. As $n \to \infty$:

Define: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (sample mean) $\quad Y = \sum_{i=1}^{n} X_i$ (sum)

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

$$\bar{X} = \frac{1}{n} Y$$

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

$$\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

The average of **i.i.d.** random variables (i.e., sample mean) is normally distributed with mean μ and variance σ^2/n.

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

What about other functions?

Let $X_1, X_2, \ldots, X_n$ i.i.d., where $E[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$. As $n \to \infty$:

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

Sum of i.i.d. RVs

$$\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

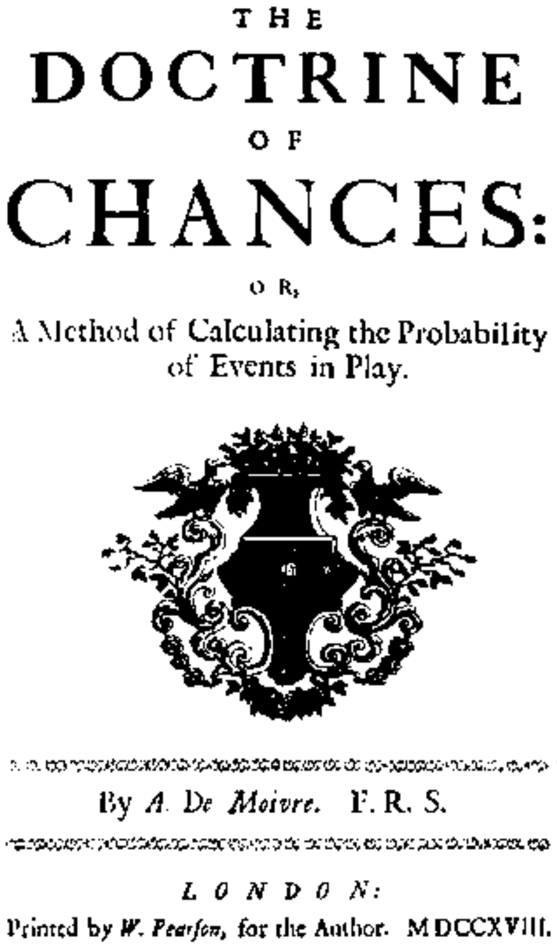
Average of i.i.d. RVs
(sample mean)

Gumbel

Max of i.i.d. RVs

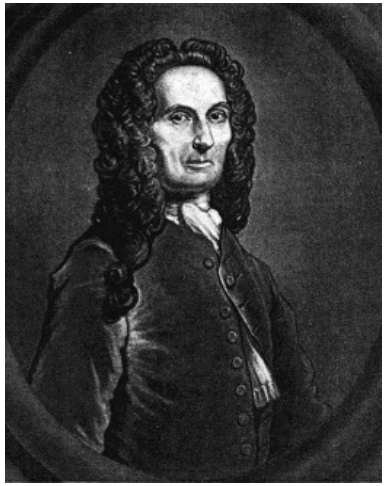
(see Fisher-Tippett Gnedenko Theorem)

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.

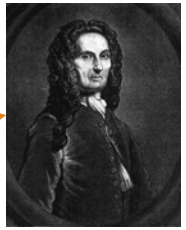


Abraham de Moivre
CLT for $X \sim \text{Ber}(1/2)$
1733

Aubrey Drake Graham
(Drake)

A short history of the CLT

1700

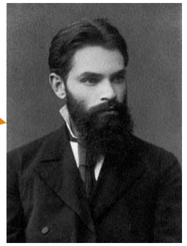


1733: CLT for $X \sim \text{Ber}(1/2)$
postulated by Abraham de Moivre

1800

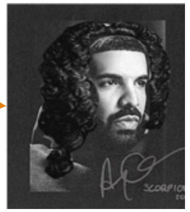
1823: Pierre-Simon Laplace extends de Moivre's work to approximating $\text{Bin}(n, p)$ with Normal

1900



1901: Alexandr Lyapunov provides precise definition and rigorous proof of CLT

2000



2018: Drake releases *Scorpion*

- It was his 5th studio album, bringing his total # of songs to 190
- Mean quality of subsamples of songs is normally distributed (thanks to the Central Limit Theorem)

Today's plan

Central Limit Theorem (CLT)

CLT exercises

Working with the CLT

Let $X_1, X_2, \ldots, X_n$ i.i.d., where $E[X_i] = \mu, \text{Var}(X_i) = \sigma^2$. As $n \to \infty$:

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

Sum of i.i.d. RVs

$$\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

Average of i.i.d. RVs
(sample mean)

If X_i is discrete:

Use the continuity correction on Y!

Dice game

As $n \to \infty$: $\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$

You will roll 10 6-sided dice $(X_1, X_2, \dots, X_{10})$.

- Let $X = X_1 + X_2 + \cdots + X_{10}$, the total value of all 10 rolls.
- You win if $X \le 25$ or $X \ge 45$.

And now the truth (according to the CLT)…

1. Define RVs and state goal.
2. Solve.

$E[X_i] = 3.5$,
$\text{Var}(X_i) = 35/12$
$X \approx Y \sim \mathcal{N}(10(3.5), 10(35/12))$

Want:

$P(X \le 25 \text{ or } X \ge 45)$

A. $P(25 \le Y \le 45)$
B. $P(Y \le 25.5) + P(Y \ge 44.5)$
C. $1 - P(25 \le Y \le 45)$
D. $1 - P(25.5 \le Y \le 44.5)$

Dice game

As $n \to \infty$: $\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$

You will roll 10 6-sided dice $(X_1, X_2, \dots, X_{10})$.

- Let $X = X_1 + X_2 + \cdots + X_{10}$, the total value of all 10 rolls.
- You win if $X \le 25$ or $X \ge 45$.

And now the truth (according to the CLT)…

1. Define RVs and state goal.

$E[X_i] = 3.5$,
$\text{Var}(X_i) = 35/12$
$X \approx Y \sim \mathcal{N}(10(3.5), 10(35/12))$

Want:

$P(X \le 25 \text{ or } X \ge 45)$
$\approx P(Y \le 25.5) + P(Y \ge 44.5)$

2. Solve.

$$P(Y \le 25.5) + P(Y \ge 44.5)$$

$$= \Phi\left(\frac{25.5 - 35}{\sqrt{10(35/12)}}\right) - \left(1 - \Phi\left(\frac{44.5 - 35}{\sqrt{10(35/12)}}\right)\right)$$

$$\approx \Phi(-1.76) - \big(1 - \Phi(1.76)\big)$$

$$\approx (1 - 0.9608) - (1 - 0.9608)$$

$$= 0.0784$$

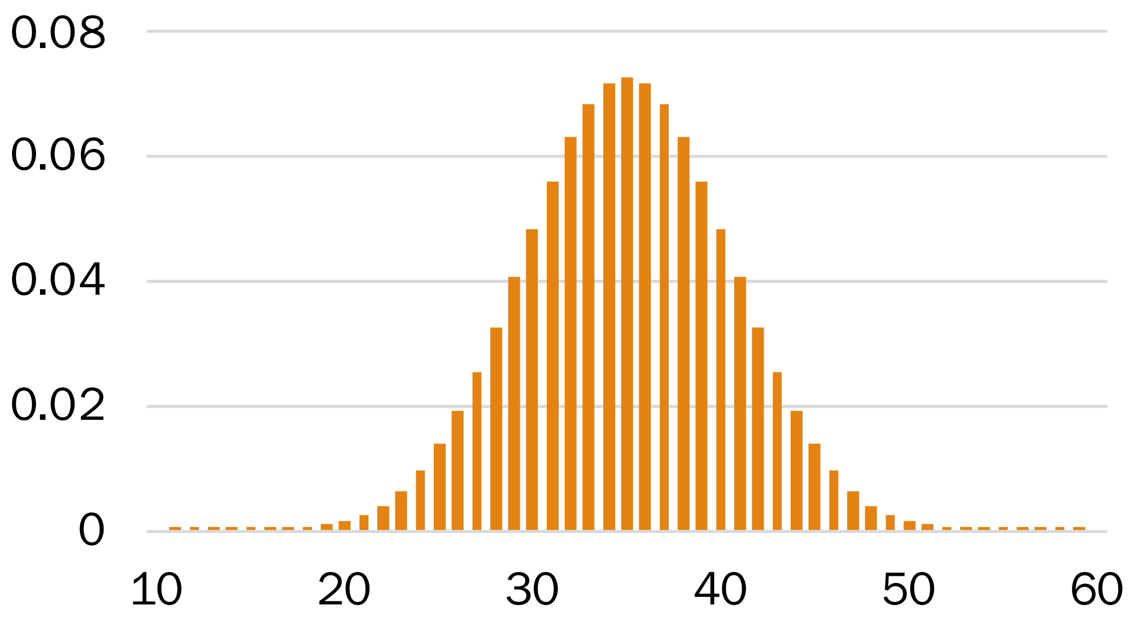
Dice game

As $n \to \infty$: $\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$

You will roll 10 6-sided dice $(X_1, X_2, \ldots, X_{10})$.

- Let $X = X_1 + X_2 + \cdots + X_{10}$, the total value of all 10 rolls.
- You win if $X \leq 25$ or $X \geq 45$.

And now the truth (according to the CLT)…



(by computer)

$$P(X \leq 25 \text{ or } X \geq 45) \approx 0.0780$$

(by CLT)

$$\approx P(Y \leq 25.5) + P(Y \geq 44.5)$$

$$\approx 0.0784$$

Clock running time

As $n \to \infty$: $\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Want to find the mean (clock) runtime of an algorithm, $\mu = t$ sec.

- Suppose variance of runtime is $\sigma^2 = 4$ sec².

Run algorithm repeatedly (i.i.d. trials):

- X_i = runtime of i-th run (for $1 \le i \le n$)
- Estimate runtime to be **average** of n trials, $\bar{X}$

How many trials do we need s.t. estimated time $= t \pm 0.5$ with 95% certainty?

1. Define RVs and state goal.
2. Solve.

(CLT) $\bar{X} \sim \mathcal{N}\left(t, \frac{4}{n}\right)$

Want: $P(t - 0.5 \le \bar{X} \le t + 0.5) = 0.95$

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

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

Clock running time

As $n \to \infty$: $\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Want to find the mean (clock) runtime of an algorithm, $\mu = t$ sec.

- Suppose variance of runtime is $\sigma^2 = 4$ sec².

Run algorithm repeatedly (i.i.d. trials):

- X_i = runtime of i-th run (for $1 \le i \le n$)
- Estimate runtime to be **average** of n trials, $\bar{X}$

How many trials do we need s.t. estimated time $= t \pm 0.5$ with 95% certainty?

1. Define RVs and state goal.

$$\bar{X} - t \sim \mathcal{N}\left(0, \frac{4}{n}\right)$$

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

2. Solve.

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

$$= \Phi\left(\frac{0.5 - 0}{\sqrt{4/n}}\right) - \Phi\left(\frac{-0.5 - 0}{\sqrt{4/n}}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1$$

Clock running time

As $n \to \infty$: $\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Want to find the mean (clock) runtime of an algorithm, $\mu = t$ sec.

- Suppose variance of runtime is $\sigma^2 = 4$ sec².

Run algorithm repeatedly (i.i.d. trials):

- X_i = runtime of i-th run (for $1 \leq i \leq n$)
- Estimate runtime to be **average** of n trials, $\bar{X}$

How many trials do we need s.t. estimated time $= t \pm 0.5$ with 95% certainty?

1. Define RVs and state goal.

$\bar{X} - t \sim \mathcal{N}\left(0, \frac{4}{n}\right)$

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

2. Solve.

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

$$= \Phi\left(\frac{0.5 - 0}{\sqrt{4/n}}\right) - \Phi\left(\frac{-0.5 - 0}{\sqrt{4/n}}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1$$

$$0.975 = \Phi(\sqrt{n}/4)$$

$$\sqrt{n}/4 = \Phi^{-1}(0.975) \approx 1.96 \Rightarrow n \approx 62$$

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
(of the Galton Board)

Extra slides

Extra CLT exercises

Crashing website

- Let X = number of visitors to a website, where $X \sim \text{Poi}(100)$.
- The server crashes if there are ≥ 120 requests/minute.

What is P(server crashes in next minute)?

Strategy: Poisson (exact)

State goal

Want: $P(X \geq 120)$

Solve

$$P(X \geq 120)$$

$$= \sum_{k=120}^{\infty} \frac{(120)^k e^{-100}}{k!}$$

$$\approx 0.0282$$

Strategy: CLT (approximate)

Crashing website

- Let X = number of visitors to a website, where $X \sim \text{Poi}(100)$.
- The server crashes if there are ≥ 120 requests/minute.

What is P(server crashes in next minute)?

Strategy: Poisson (exact)

State goal

Want: $P(X \geq 120)$

Solve

$$P(X \geq 120)$$

$$= \sum_{k=120}^{\infty} \frac{(120)^k e^{-100}}{k!}$$

$$\approx 0.0282$$

Strategy: CLT (approximate)

State approx. goal

$$\text{Poi}(100) \sim \sum_{i=1}^{n} \text{Poi}(100/n)$$

(sum of IID Poisson = Poisson)

$$X \approx Y \sim \mathcal{N}(n\mu, n\sigma^2)$$

$$\mu = \frac{100}{n} = \sigma^2$$

Want: $P(X \geq 120) \approx P(Y \geq 119.5)$

Solve

$$P(Y \geq 119.5) = 1 - \Phi\left(\frac{119.5 - 100}{\sqrt{100}}\right) = 1 - \Phi(1.95)$$

$$\approx 0.0256$$