Central Limit Theorem
CSE 312 24Su
Lecture 15
Logistics

• Reminder about concept checks 12, 13, and 14, late due date tonight
• Midterm grades released on ~Wednesday
• Updated lecture notes for last Wed and Fri lecture on website
  \[\text{will be soon}\]
Normal Distributions

A normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ has two parameters:

- $\mu = \mathbb{E}[X]$ is the mean
- $\sigma^2 = \text{Var}(X)$ is the variance ($\sigma = \sqrt{\text{Var}(X)}$ is standard deviation)

and follows this **probability density function** (a bell curve!):

$$f_X(k) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-\mu)^2}{2\sigma^2}}$$
Normal Distributions

The CDF has no closed form, so instead, we have a table containing values of the CDF for a standard normal random variable $\mathcal{N}(0,1)$.

To find the probability of a normal RV $X \sim \mathcal{N}(\mu, \sigma^2)$ being in some range...

1. **Standardize** the normal random variable: $Z = \frac{X - \mu}{\sigma}$

   note: when we standardize, the numbers left are called z-scores (the number of standard deviations away from the mean (e.g., $\mathbb{P}(Z \geq 2)$ means we're finding probability of being more than 2 standard deviations away from the mean)

2. Write probability expression in terms of $\Phi(z) = \mathbb{P}(Z \leq z)$

3. Look up the value(s) in the table
Normal distributions show up everywhere!
But...why?

This is because of what we call, the **central limit theorem**!

“The sum of any independent random variables approaches a normal distribution. It becomes closer to normal as we sum more RVs together.”
More formally, the Central Limit Theorem!

This is because of what we call, the *central limit theorem*!

“The sum of *any* independent random variables *approaches* a normal distribution. It becomes closer to normal as we sum more RVs together.”

---

**Central Limit Theorem**

If $X_1, X_2, \ldots, X_n$ are i.i.d. random variables, each with mean $\mu$ and variance $\sigma^2$

Let $Y_n = X_1 + X_2 + \cdots + X_n$

As $n \to \infty$, $Y_n$ approaches a normal distribution $\mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$ (i.e., CDF of $Y_n$ converges to the CDF of $\mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$)
What does i.i.d mean?

**Independent and Identically Distributed (i.i.d)**

For random variables $X_1, X_2, \ldots, X_n$ to be i.i.d., they must

- **Be mutually independent**
  "knowing value of one random variable doesn't give us info about the value of others"

- **All have the same PMF (if discrete) or PDF (if continuous)**
  "they follow the same probability distribution"
CLT with \textbf{RVs that are NOT i.i.d}

(R.A. Fisher (1918))

Height may be expressed as the \textit{sum of many independent, random factors}:
- How much milk you drank per day as a child
- Which variant of GH1 (a growth hormone) you have
- How much protein/calcium/vitamins/minerals you had as a child
- How many hours of sleep you averaged
- How many hours of physical activity you averaged

\[ H = H_1 + H_2 + \ldots + H_n \rightarrow H \sim N(...) \]

A version of CLT does work here! But it’s outside the scope of this class.
CLT with **RVs that are i.i.d**

“number of firing neurons” – sum of indicator random variables for whether each neuron fired
Assume: each neuron is independent, and has the same probability

“number of people who voted for someone” – sum of indicator random variables for (assume people are independent)
Assume: each person makes independent, and has the same probability

“total amount invested in a year” – sum of random variables four amount invested each day (assume each investment is independent)
Assume: each day’s investment is independent and follows the same distribution

We will use CLT in this class on problems like this.
A Sum of i.i.d Random Variables

If we have $X_1, X_2, \ldots, X_n$ as i.i.d RVs each with mean $\mu$ and variance $\sigma^2$

$S_n = X_1 + X_2 + \cdots + X_n$ is the sum of those RVs. Then...

> **Expectation.** by linearity of expectation...

$$\mathbb{E}[S_n] = \mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] = n\mu$$

> **Variance.** by linearity of variance because of independence

$$Var(S_n) = Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n) = n\sigma^2$$

$$S_n \sim N(n\mu, n\sigma^2)$$
Proof of the CLT?

We’re not going to cover the proof here.

How is the proof done?

Step 1: Prove that for all positive integers $k$, $\left(Y_n\right)^k \rightarrow \mathbb{E}[Z^k]$

Step 2: Prove that if $\mathbb{E}\left[(Y_n)^k\right] = \mathbb{E}[Z^k]$ for all $k$ then $F_{Y_n}(z) = F_Z(z)$
“Proof by example”

CLT says a sum of $n$ i.i.d RVs approaches a normal distribution as $n$ gets larger

$n$ is the number of i.i.d RVs summed

The dotted lines show an “empirical PMF” – a PMF estimated by running the experiment a large number of times.

The blue line is the normal RV that the CLT predicts.

Shown are $n = 1, 2, 3, 10$
“Proof by example” -- uniform

https://www.desmos.com/calculator/2n2m05a9km
“Proof by example” -- uniform \( n=1 \)

\[ X = X_1 \]

where each \( X_i \sim \text{Unif}(0,1) \) and is independent
“Proof by example” -- uniform

\[ X = X_1 \]

where each \( X_i \sim \text{Unif}(0,1) \) and is independent
"Proof by example" -- uniform

\[ X = X_1 + X_2 \]

where each \( X_i \sim \text{Unif}(0,1) \) and is independent
“Proof by example” -- uniform

\[ X = X_1 + X_2 + X_3 \]

where each \( X_i \sim \text{Unif}(0,1) \) and is independent
“Proof by example” -- uniform

\[ X = X_1 + X_2 + X_3 + X_4 \]
where each \( X_i \sim \text{Unif}(0,1) \) and is independent
“Proof by example” -- uniform

\[ X = X_1 + X_2 + X_3 + X_4 + X_5 \]
where each \( X_i \sim \text{Unif}(0,1) \) and is independent
“Proof by example” -- uniform

\[ X = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 \]

where each \( X_i \sim \text{Unif}(0,1) \) and is independent
“Proof by example” -- uniform

\[ X = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 \]
where each \( X_i \sim \text{Unif}(0,1) \) and is independent
“Proof by example” -- uniform

\[ X = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 \]
where each \( X_i \sim \text{Unif}(0,1) \) and is independent
“Proof by example” -- uniform

\[ X = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 \]

where each \( X_i \sim \text{Unif}(0,1) \) and is independent
“Proof by example” -- uniform

\[ X = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} \]

where each \( X_i \sim \text{Unif}(0,1) \) and is independent
A lot of real-world bell-curves can be explained as:

1. The random variable comes from a combination of independent factors.
2. The CLT says the distribution will become like a bell curve.
Theory vs. Practice

> The formal theorem statement is “in the limit”

You might not get exactly a normal distribution for any finite $n$ (e.g. if you sum discrete, the sum is always discrete and will be discontinuous for every finite $n$.

> In practice, the approximations get very accurate very quickly (at least with a few tricks we’ll see soon).

They won’t be exact (unless the $X_i$ are normals) but it’s close enough to use even with relatively small $n$. 
Using the Central Limit Theorem

Let’s start with the case when we are using CLT to approximate a sum of continuous i.i.d random variables as normal.
Outline of CLT steps

1. Setup the problem (e.g., \( X = \sum_{i=1}^{n} X_i \), \( X_i \) are i.i.d., and we want \( \mathbb{P}(X \leq k) \))
   Write event you are interested in, in terms of sum of random variables.

   ✔️ we’re going to be adding one more step here when we talk about discrete RVs!

2. Apply CLT (e.g., approx \( X \) as \( Y \sim N(n\mu, n\sigma^2) \) -\( \mathbb{P}(X \leq k) \approx \mathbb{P}(Y \leq k) \))
   Approximate sum of RVs as normal with appropriate mean and variance

   from here, we’re working with a normal distribution, which we’ve worked with before!

3. Compute probability approximation using Phi table

   > **Standardize** \( (Z = \frac{N-\mu}{\sigma}) \) -\( \mathbb{P}(Y \leq k) = \mathbb{P}\left(\frac{Y-\mu}{\sigma} \leq \frac{k-\mu}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{k-\mu}{\sigma}\right) \)

   > **Write in terms of** \( \Phi(z) = \mathbb{P}(Z \leq z) \)

   > **Look up in table**
You buy lightbulbs that each burn out according to an exponential distribution with parameter of $\lambda = 1.8$ lightbulbs per year.

You buy a 10 pack of (independent) light bulbs. What is the probability that your 10-pack lasts at least 5 years?

Let $X_i$ be the time it takes for lightbulb $i$ to burn out. Let $X$ be the total time. Estimate $\mathbb{P}(X \geq 5)$. 

You buy lightbulbs that each burn out according to distribution $\text{Exp}(1.8)$ lightbulbs per year. Estimate the probability your 10-pack (independent) lasts at least 5 years?

1. Setup the Problem:
   \[ X = \sum_{i=1}^{10} X_i \text{ burn } \sim \text{Exp}(1.8) \]
   \[ \mathbb{E}(X_i) = \mu = \frac{1}{1.8}, \quad \text{Var}(X_i) = \frac{1}{1.8^2} \]

2. Apply CLT.
   \[ Y \sim N\left(10 \cdot \frac{1}{1.8}, 10 \cdot \frac{1}{1.8^2}\right) \]
   \[ X \sim Y \]
   \[ P(X \geq 5) \approx P(Y \geq 5) \]

   \[ P(Y \geq 5) = P\left(Z \geq \frac{5 - \frac{10}{1.8}}{\sqrt{10/1.8^2}}\right) = P(Z \geq -0.32) \]
   \[ = 1 - P(Z \leq -0.32) \]
   \[ = 1 - P(Z \geq 0.32) \]
   \[ = 1 - P(Z \geq 0.32) \]
You buy lightbulbs that each burn out according to distribution \( \text{Exp}(1.8) \) lightbulbs per year. Estimate the probability your 10-pack (independent) lasts at least 5 years?

1. **Setup the Problem:** Let \( X_i \) be the time it takes for lightbulb \( i \) to burn out. \( X_i \sim \text{Exp}(1.8) \) and \( \mu = \mathbb{E}[X_i] = \frac{1}{1.8} \) and \( \sigma^2 = \text{Var}(X_i) = \frac{1}{1.8^2} \). Let \( X \) be the total time and \( X = \sum_{i=1}^{10} X_i \). We are interested in \( \mathbb{P}(X \geq 5) \).

2. **Apply CLT.** Because the \( X_i \)'s are i.i.d., we can apply CLT and \( X \) can be approximated by \( Y \sim \mathcal{N}(10 \cdot \frac{1}{1.8}, 10 \cdot \frac{1}{1.8^2}) \). \( \mathbb{P}(X \geq 5) \approx \mathbb{P}(Y \geq 5) \)

3. **Compute Probability.**

\[
\mathbb{P}(Y \geq 5) = \mathbb{P}\left( Z \geq \frac{5 - 10/1.8}{\sqrt{10/1.8}} \right) \text{ standardize }
\]

\[
\approx \mathbb{P}(Z \geq -0.32) = \mathbb{P}(Z \leq 0.32) \text{ (symmetry)}
\]

\[
= \Phi(0.32) \approx 0.62552 \text{ plug into z-table}
\]

(True value (uses a distribution not in our zoo) is \( \approx 0.58741 \))
Using the Central Limit Theorem

Now, let’s try the case when we are using CLT to approximate a sum of \textit{discrete} i.i.d random variables as normal.
Suppose you are managing a factory, that produces widgets. Each widget produced is defective (independently) with probability 5%.

Your factory will produce 1000 (possibly defective) widgets. What is the probability of producing at most 940 non-defective widgets?
Suppose you are managing a factory, that produces widgets. Each widget produced is defective (independently) with probability 5%.

Your factory will produce 1000 (possibly defective) widgets. What is the probability of producing at most 940 non-defective widgets?

\[ X \] is the number of non-defective widgets. Let \( X \sim \text{Bin}(1000, .95) \)

Our goal: \( \mathbb{P}(X \leq 940) \)?

That's a big summation: \( \sum_{k=0}^{940} \binom{1000}{k} (.95)^k \cdot (.05)^{1000-k} \approx .08673 \)
Suppose you are managing a factory, that produces widgets. Each widget produced is defective (independently) with probability 5%.

Your factory will produce 1000 (possibly defective) widgets. What is the probability of producing at most 940 non-defective widgets?

\[ X \] is the number of non-defective widgets. Let \( X \sim \text{Bin}(1000, .95) \)

Our goal: \( \mathbb{P}(X \leq 940) \)?

That’s a big summation: \( \sum_{k=0}^{940} \binom{1000}{k} (.95)^k \cdot (.05)^{1000-k} \approx .08673 \)

What does the CLT give? Binomial is sum of i.i.d bernoullis -> can use CLT!
Suppose you are managing a factory, that produces widgets. Each widget produced is defective (independently) with probability 5%.

Your factory will produce 1000 (possibly defective) widgets. What is the probability of producing at most 940 non-defective widgets?

1. Setup the Problem:

2. Apply CLT.

Factory Widgets - CLT

Suppose you are managing a factory, that produces widgets. Each widget produced is defective (independently) with probability 5%.

Your factory will produce 1000 (possibly defective) widgets. What is the probability of producing at most 940 non-defective widgets?

1. Setup the Problem: $X$ is the number of non-defective widgets.
   $X = \sum_{i=1}^{1000} X_i$ where $X_i$ is 1 if the $i$'th widget is non-defective. **Goal**: $\mathbb{P}(X \leq 940)$

2. Apply CLT. $X$ is sum of i.i.d RVs each with $\mu = \mathbb{E}[X_i] = p = .95$ and $\text{Var}(X_i) = p(1-p) = .0475$, we can approximate $X$ with $Y \sim \mathcal{N}(1000 \cdot .95, 1000 \cdot .0475)$. So, $\mathbb{P}(X \leq 940) \approx \mathbb{P}(Y \leq 940)$

   
   $\mathbb{P}(Y \leq 940) = \mathbb{P}\left( Z \leq \frac{940-1000\cdot .95}{\sqrt{1000\cdot .0475}} \right)$
   
   $\approx \Phi(-1.45) = 1 - \Phi(1.45)$ write in terms of $\Phi$

   $\approx 1 - .92647 = .07353$ plug into z-table
Suppose you are managing a factory, that produces widgets. Each widget produced is defective (independently) with probability 5%.

Your factory will produce 1000 (possibly defective) widgets. What is the probability of producing at most 940 non-defective widgets?

1. **Setup the Problem**: $X$ is the number of non-defective widgets. $X = \sum_{i=1}^{1000} X_i$ where $X_i$ is 1 if the $i$’th widget is non-defective. **Goal**: $\mathbb{P}(X \leq 940)$

2. **Apply CLT**: $X$ is sum of i.i.d RVs each with $\mu = \mathbb{E}[X_i] = p = .95$ and $\text{Var}(X_i) = p(1-p) = .0475$, we can approximate $X$ with $Y \sim \mathcal{N}(1000 \cdot .95, 1000 \cdot .0475)$. So, $\mathbb{P}(X \leq 940) \approx \mathbb{P}(Y \leq 940)$

3. **Compute Probability**.

   $\mathbb{P}(Y \leq 940) = \mathbb{P} \left( Z \leq \frac{940 - 1000 \cdot .95}{\sqrt{1000 \cdot .0475}} \right)$  
   
   standardize
   
   $\approx \Phi(-1.45) = 1 - \Phi(1.45)$  
   
   write in terms of $\Phi$
   
   $\approx 1 - .92647 = .07353$  
   
   plug into z-table

The exact probability is .08673. We’re off by ~1.3%!
There’s are some problems 😞

When approximating a discrete distribution like binomial with a continuous normal distribution, there are some problems that arise!

> \( \mathbb{P}(X = 2) > 0 \) (we can use the binomial PMF).

But, when we approximate to the normal, continuous, \( Y \), \( \mathbb{P}(Y = 2) = 0 \)

> \( X \) only takes on integers, so \( \mathbb{P}(X \leq 1) + \mathbb{P}(X \geq 2) = 1 \).

But, when we approximate to the normal, continuous, \( Y \),
\( \mathbb{P}(Y \leq 1) + \mathbb{P}(Y \geq 2) < 1 \)
Continuity Correction

The binomial distribution is **discrete**, but the normal is **continuous**. Let’s correct for that (called a "**continuity correction**")

⭐ Assign each value in the discrete range to a **continuous** interval

*Here the support of $X$ is $\{..., -2, -1, 0, 1, 2, ...\}$*
Continuity Correction

The binomial distribution is **discrete**, but the normal is **continuous**. Let’s correct for that (called a “continuity correction”)

Assign each value in the discrete range to a continuous interval

*Here the support of $X$ is $\{..., -2, -1, 0, 1, 2, ... \}$*

\[
\begin{align*}
\Pr(X = 2) &\rightarrow \Pr(1.5 \leq X \leq 2.5) \\
\Pr(X \geq 2) &\rightarrow \Pr(X \geq 1.5) \\
\Pr(X > 1) &\rightarrow \Pr(X \geq 1.5) \\
\Pr(X \leq 1) &\rightarrow \Pr(X \leq 1.5)
\end{align*}
\]
Continuity Correction

The binomial distribution is **discrete**, but the normal is **continuous**. Let’s correct for that (called a “*continuity correction*”)

Assign each value in the discrete range to a continuous interval

*Here the support of $X$ is $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$*

\[
\begin{align*}
\mathbb{P}(X = 2) &\rightarrow \mathbb{P}(1.5 \leq X \leq 2.5) \\
\mathbb{P}(X \geq 1) &\rightarrow \mathbb{P}(X \geq 0.5) \\
\mathbb{P}(X > 1) &\rightarrow \mathbb{P}(X \geq 1.5) \\
\mathbb{P}(X \leq 1) &\rightarrow \mathbb{P}(X \leq 1.5)
\end{align*}
\]
**Continuity Correction**

The binomial distribution is **discrete**, but the normal is **continuous**. Let’s correct for that (called a "continuity correction")

⭐ Assign each value in the discrete range to a continuous interval

*Here the support of $X$ is \{..., $-2$, $5$, $12$, $19$ ...\})*

*e.g.,

$\mathbb{P}(X = -2) \rightarrow$

$\mathbb{P}(X \geq 5) \rightarrow$

$\mathbb{P}(X < 12) \rightarrow$

$\mathbb{P}(X \geq 0) \rightarrow$*
Continuity Correction

The binomial distribution is **discrete**, but the normal is **continuous**. Let’s correct for that (called a “*continuity correction*”)

🌟 Assign each value in the discrete range to a continuous interval

*Here the support of $X$ is \{..., −2, 5, 12, 19 ... \}*

\[\begin{align*}
\mathbb{P}(X = −2) & \rightarrow \mathbb{P}(−5.5 \leq X \leq 1.5) \\
\mathbb{P}(X \geq 5) & \rightarrow \mathbb{P}(X \geq 1.5) \\
\mathbb{P}(X < 12) & \rightarrow \mathbb{P}(X \leq 8.5) \\
\mathbb{P}(X \geq 0) & \rightarrow \mathbb{P}(X \geq 1.5)
\end{align*}\]
Continuity Correction

The binomial distribution is **discrete**, but the normal is **continuous**.

Let’s correct for that (called a “**continuity correction**”)

🌟 Assign each value in the discrete range to a continuous interval

*Here the support of $X$ is $\{..., -2, 5, 12, 19 \ldots \}$*

\[
\begin{align*}
\Pr(X = -2) &\rightarrow \Pr(-5.5 \leq X \leq 1.5) \\
\Pr(X \geq 5) &\rightarrow \Pr(X \geq 1.5) \\
\Pr(X < 12) &\rightarrow \Pr(X \leq 8.5) \\
\Pr(X \geq 0) &\rightarrow \Pr(X \geq 1.5)
\end{align*}
\]
Outline of CLT steps

1. **Setup the problem** (e.g., $X = \sum_{i=1}^{n} X_i$, $X_i$ are i.i.d., and we want $\mathbb{P}(X \leq k)$)
   
   Write event you are interested in, in terms of sum of random variables.

   ★ Apply *continuity correction here if RVs are discrete.*

2. **Apply CLT** (e.g., approx $X \sim N(n\mu, n\sigma^2)$ $\Rightarrow$ $\mathbb{P}(X \leq k) \approx \mathbb{P}(Y \leq k)$
   
   Approximate sum of RVs as normal with appropriate mean and variance
   
   *from here, we’re working with a normal distribution, which we’ve worked with before!*

3. **Compute probability approximation using Phi table**
   
   > **Standardize** ($Z = \frac{N-\mu}{\sigma}$) $\Rightarrow$ $\mathbb{P}(Y \leq k) = \mathbb{P}\left(\frac{Y-\mu}{\sigma} \leq \frac{k-\mu}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{k-\mu}{\sigma}\right)$
   
   > **Write in terms of** $\Phi(z) = \mathbb{P}(Z \leq z)$
   
   > **Look up in table**
Factory Widgets - CLT with continuity correction

Suppose you are managing a factory that produces widgets. Each widget produced is defective (independently) with probability 5%. Your factory will produce 1000 (possibly defective) widgets. What is the probability of producing at most 940 non-defective widgets?

1. **Setup the Problem:** $X$ is the number of non-defective widgets. $X = \sum_{i=1}^{1000} X_i$ where $X_i$ is 1 if the $i$'th widget is non-defective. We want to find $P(X \leq 940)$. Because $X$ is discrete, we use continuity correction: $P(X \leq 940) = P(X \leq 940.5)$

2. **Apply CLT.** $X$ is sum of i.i.d RVs each with $\mu = \mathbb{E}[X_i] = p = .95$ and $\text{Var}(X_i) = p(1-p) = .0475$, we can approximate $X$ with $Y \sim \mathcal{N}(1000 \cdot .95, 1000 \cdot .0475)$. So, $P(X \leq 940.5) \approx P(Y \leq 940.5)$

3. **Compute Probability.**

$P(Y \leq 940.5) = P \left( Z \leq \frac{940.5 - 1000 \cdot .95}{\sqrt{1000 \cdot .0475}} \right)$

$\approx \Phi(-1.38) = 1 - \Phi(1.38)$

$\approx 1 - .91621 = .08379.$

The exact probability is .08673. Still an approximation, but very close now! :D