18.600: Lecture 30
Central limit theorem

Scott Sheffield

MIT
Central limit theorem

Proving the central limit theorem
Central limit theorem

Proving the central limit theorem
Recall: DeMoivre-Laplace limit theorem

- Let $X_i$ be an i.i.d. sequence of random variables. Write $S_n = \sum_{i=1}^{n} X_n$.

- Suppose each $X_i$ is 1 with probability $p$ and 0 with probability $q = 1 - p$.

- DeMoivre-Laplace limit theorem: $\lim_{n \to \infty} P\{a \leq S_n - np \leq b \sqrt{npq}\} \to \Phi(b) - \Phi(a)$.

- Here $\Phi(b) - \Phi(a) = P\{a \leq Z \leq b\}$ when $Z$ is a standard normal random variable.

- $S_n - np \sqrt{npq}$ describes “number of standard deviations that $S_n$ is above or below its mean”.

- Question: Does a similar statement hold if the $X_i$ are i.i.d. but have some other probability distribution?

- Central limit theorem: Yes, if they have finite variance.
Recall: DeMoivre-Laplace limit theorem

Let $X_i$ be an i.i.d. sequence of random variables. Write $S_n = \sum_{i=1}^{n} X_i$.

Suppose each $X_i$ is 1 with probability $p$ and 0 with probability $q = 1 - p$.

DeMoivre-Laplace limit theorem: \[
\lim_{n \to \infty} P\{a \leq S_n - np \sqrt{npq} \leq b\} \to \Phi(b) - \Phi(a).
\]

Here $\Phi(b) - \Phi(a) = P\{a \leq Z \leq b\}$ when $Z$ is a standard normal random variable.

$S_n - np \sqrt{npq}$ describes "number of standard deviations that $S_n$ is above or below its mean".

Question: Does a similar statement hold if the $X_i$ are i.i.d. but have some other probability distribution?

Central limit theorem: Yes, if they have finite variance.
Recall: DeMoivre-Laplace limit theorem

Let $X_i$ be an i.i.d. sequence of random variables. Write $S_n = \sum_{i=1}^{n} X_i$.

Suppose each $X_i$ is 1 with probability $p$ and 0 with probability $q = 1 - p$.

DeMoivre-Laplace limit theorem:

$$\lim_{n \to \infty} P\{a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\} \to \Phi(b) - \Phi(a).$$
Recall: DeMoivre-Laplace limit theorem

Let $X_i$ be an i.i.d. sequence of random variables. Write $S_n = \sum_{i=1}^{n} X_i$.

Suppose each $X_i$ is 1 with probability $p$ and 0 with probability $q = 1 - p$.

DeMoivre-Laplace limit theorem:

$$\lim_{n \to \infty} P\{a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\} \to \Phi(b) - \Phi(a).$$

Here $\Phi(b) - \Phi(a) = P\{a \leq Z \leq b\}$ when $Z$ is a standard normal random variable.
Recall: DeMoivre-Laplace limit theorem

- Let $X_i$ be an i.i.d. sequence of random variables. Write $S_n = \sum_{i=1}^{n} X_i$.
- Suppose each $X_i$ is 1 with probability $p$ and 0 with probability $q = 1 - p$.
- **DeMoivre-Laplace limit theorem:**

$$\lim_{n \to \infty} \mathbb{P}\{a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\} \to \Phi(b) - \Phi(a).$$

- Here $\Phi(b) - \Phi(a) = P\{a \leq Z \leq b\}$ when $Z$ is a standard normal random variable.
- $\frac{S_n - np}{\sqrt{npq}}$ describes “number of standard deviations that $S_n$ is above or below its mean”.

Question: Does a similar statement hold if the $X_i$ are i.i.d. but have some other probability distribution?

Central limit theorem: Yes, if they have finite variance.
Recall: DeMoivre-Laplace limit theorem

- Let $X_i$ be an i.i.d. sequence of random variables. Write $S_n = \sum_{i=1}^{n} X_n$.
- Suppose each $X_i$ is $1$ with probability $p$ and $0$ with probability $q = 1 - p$.
- **DeMoivre-Laplace limit theorem:**
  
  $$\lim_{n \to \infty} P\{a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\} \to \Phi(b) - \Phi(a).$$

- Here $\Phi(b) - \Phi(a) = P\{a \leq Z \leq b\}$ when $Z$ is a standard normal random variable.
- $\frac{S_n - np}{\sqrt{npq}}$ describes “number of standard deviations that $S_n$ is above or below its mean”.
- Question: Does a similar statement hold if the $X_i$ are i.i.d. but have some other probability distribution?
Recall: DeMoivre-Laplace limit theorem

- Let \( X_i \) be an i.i.d. sequence of random variables. Write \( S_n = \sum_{i=1}^{n} X_i \).
- Suppose each \( X_i \) is 1 with probability \( p \) and 0 with probability \( q = 1 - p \).
- **DeMoivre-Laplace limit theorem:**
  \[
  \lim_{n \to \infty} P\{ a \leq \frac{S_n - np}{\sqrt{npq}} \leq b \} \to \Phi(b) - \Phi(a).
  \]
- Here \( \Phi(b) - \Phi(a) = P\{ a \leq Z \leq b \} \) when \( Z \) is a standard normal random variable.
- \( \frac{S_n - np}{\sqrt{npq}} \) describes “number of standard deviations that \( S_n \) is above or below its mean”.
- Question: Does a similar statement hold if the \( X_i \) are i.i.d. but have some other probability distribution?
- **Central limit theorem:** Yes, if they have finite variance.
Example

- Say we roll $10^6$ ordinary dice independently of each other.
Example

Say we roll $10^6$ ordinary dice independently of each other.

Let $X_i$ be the number on the $i$th die. Let $X = \sum_{i=1}^{10^6} X_i$ be the total of the numbers rolled.

What is $E[X]$?

$10^6 \cdot (7/2)$

What is $\text{Var}[X]$?

$10^6 \cdot (35/12)$

How about $\text{SD}[X] = \sqrt{\text{Var}[X]}$?

$1000 \sqrt{35/12}$

What is the probability that $X$ is less than a standard deviations above its mean?

Central limit theorem: should be about $1/\sqrt{2\pi} \int_{-\infty}^{a} e^{-x^2/2} dx$. 
Say we roll $10^6$ ordinary dice independently of each other.

Let $X_i$ be the number on the $i$th die. Let $X = \sum_{i=1}^{10^6} X_i$ be the total of the numbers rolled.

What is $E[X]$?
Example

- Say we roll $10^6$ ordinary dice independently of each other.
- Let $X_i$ be the number on the $i$th die. Let $X = \sum_{i=1}^{10^6} X_i$ be the total of the numbers rolled.
- What is $E[X]$?
- $10^6 \cdot (7/2)$
Example

- Say we roll $10^6$ ordinary dice independently of each other.
- Let $X_i$ be the number on the $i$th die. Let $X = \sum_{i=1}^{10^6} X_i$ be the total of the numbers rolled.
- What is $E[X]$?
  - $10^6 \cdot (7/2)$
- What is $\text{Var}[X]$?
  - $1000 \sqrt{ \frac{35}{12} }$
Example

Say we roll $10^6$ ordinary dice independently of each other.

Let $X_i$ be the number on the $i$th die. Let $X = \sum_{i=1}^{10^6} X_i$ be the total of the numbers rolled.

What is $E[X]$?

$10^6 \cdot (7/2)$

What is $\text{Var}[X]$?

$10^6 \cdot (35/12)$
Example

Say we roll $10^6$ ordinary dice independently of each other.

Let $X_i$ be the number on the $i$th die. Let $X = \sum_{i=1}^{10^6} X_i$ be the total of the numbers rolled.

What is $E[X]$?

$10^6 \cdot (7/2)$

What is $\text{Var}[X]$?

$10^6 \cdot (35/12)$

How about $\text{SD}[X] = \sqrt{\text{Var}[X]}$?
Say we roll $10^6$ ordinary dice independently of each other.

Let $X_i$ be the number on the $i$th die. Let $X = \sum_{i=1}^{10^6} X_i$ be the total of the numbers rolled.

What is $E[X]$?

$10^6 \cdot (7/2)$

What is $\text{Var}[X]$?

$10^6 \cdot (35/12)$

How about $\text{SD}[X] = \sqrt{\text{Var}[X]}$?

$1000 \sqrt{35/12}$
Example

- Say we roll $10^6$ ordinary dice independently of each other.
- Let $X_i$ be the number on the $i$th die. Let $X = \sum_{i=1}^{10^6} X_i$ be the total of the numbers rolled.
- What is $E[X]$?
  - $10^6 \cdot (7/2)$
- What is $\text{Var}[X]$?
  - $10^6 \cdot (35/12)$
- How about $\text{SD}[X] = \sqrt{\text{Var}[X]}$?
  - $1000 \sqrt{35/12}$
- What is the probability that $X$ is less than a standard deviations above its mean?

**Central limit theorem:** should be about $\sqrt{\frac{2}{\pi}} \int_{-\infty}^{a} e^{-x^2/2} \, dx$. 
Say we roll $10^6$ ordinary dice independently of each other. Let $X_i$ be the number on the $i$th die. Let $X = \sum_{i=1}^{10^6} X_i$ be the total of the numbers rolled.

What is $E[X]$?

$10^6 \cdot (7/2)$

What is $\text{Var}[X]$?

$10^6 \cdot (35/12)$

How about $\text{SD}[X] = \sqrt{\text{Var}[X]}$?

$1000\sqrt{35/12}$

What is the probability that $X$ is less than a standard deviations above its mean?

Central limit theorem: should be about \( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} \, dx \).
Example

- Suppose earthquakes in some region are a Poisson point process with rate $\lambda$ equal to 1 per year.
Example

- Suppose earthquakes in some region are a Poisson point process with rate $\lambda$ equal to 1 per year.
- Let $X$ be the number of earthquakes that occur over a ten-thousand year period. Should be a Poisson random variable with rate 10000.
Example

- Suppose earthquakes in some region are a Poisson point process with rate $\lambda$ equal to 1 per year.
- Let $X$ be the number of earthquakes that occur over a ten-thousand year period. Should be a Poisson random variable with rate 10000.
- What is $E[X]$?
Example

- Suppose earthquakes in some region are a Poisson point process with rate $\lambda$ equal to 1 per year.
- Let $X$ be the number of earthquakes that occur over a ten-thousand year period. Should be a Poisson random variable with rate 10000.
- What is $E[X]$?
- 10000
Example

- Suppose earthquakes in some region are a Poisson point process with rate $\lambda$ equal to 1 per year.
- Let $X$ be the number of earthquakes that occur over a ten-thousand year period. Should be a Poisson random variable with rate 10000.
- What is $E[X]$?
  - 10000
- What is $\text{Var}[X]$?
Example

- Suppose earthquakes in some region are a Poisson point process with rate $\lambda$ equal to 1 per year.
- Let $X$ be the number of earthquakes that occur over a ten-thousand year period. Should be a Poisson random variable with rate 10000.
- What is $E[X]$?
  - 10000
- What is $\text{Var}[X]$?
  - 10000
Example

- Suppose earthquakes in some region are a Poisson point process with rate $\lambda$ equal to 1 per year.
- Let $X$ be the number of earthquakes that occur over a ten-thousand year period. Should be a Poisson random variable with rate 10000.
- What is $E[X]$?
  - 10000
- What is $\text{Var}[X]$?
  - 10000
- How about $\text{SD}[X]$?
Example

- Suppose earthquakes in some region are a Poisson point process with rate $\lambda$ equal to 1 per year.
- Let $X$ be the number of earthquakes that occur over a ten-thousand year period. Should be a Poisson random variable with rate 10000.
- What is $E[X]$?
  - 10000
- What is $\text{Var}[X]$?
  - 10000
- How about $\text{SD}[X]$?
  - 100

Central limit theorem: should be about $\frac{1}{\sqrt{2\pi}} \int_{a}^{-\infty} e^{-x^2/2} \, dx$. 
Example

- Suppose earthquakes in some region are a Poisson point process with rate $\lambda$ equal to 1 per year.
- Let $X$ be the number of earthquakes that occur over a ten-thousand year period. Should be a Poisson random variable with rate 10000.
- What is $E[X]$?
  - 10000
- What is $\text{Var}[X]$?
  - 10000
- How about $\text{SD}[X]$?
  - 100
- What is the probability that $X$ is less than a standard deviations above its mean?
  - Central limit theorem: should be about $1 / \sqrt{2\pi}$
Example

- Suppose earthquakes in some region are a Poisson point process with rate $\lambda$ equal to 1 per year.
- Let $X$ be the number of earthquakes that occur over a ten-thousand year period. Should be a Poisson random variable with rate 10000.
- What is $E[X]$?
  - 10000
- What is $\text{Var}[X]$?
  - 10000
- How about $\text{SD}[X]$?
  - 100
- What is the probability that $X$ is less than a standard deviations above its mean?
  - Central limit theorem: should be about $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} \, dx$. 
Let $X_i$ be an i.i.d. sequence of random variables with finite mean $\mu$ and variance $\sigma^2$.
Let $X_i$ be an i.i.d. sequence of random variables with finite mean $\mu$ and variance $\sigma^2$.

Write $S_n = \sum_{i=1}^{n} X_i$. So $E[S_n] = n\mu$ and $\text{Var}[S_n] = n\sigma^2$ and $\text{SD}[S_n] = \sigma\sqrt{n}$.
Let $X_i$ be an i.i.d. sequence of random variables with finite mean $\mu$ and variance $\sigma^2$.

Write $S_n = \sum_{i=1}^{n} X_i$. So $E[S_n] = n\mu$ and $\text{Var}[S_n] = n\sigma^2$ and $\text{SD}[S_n] = \sigma \sqrt{n}$.

Write $B_n = \frac{X_1+X_2+\ldots+X_n-n\mu}{\sigma \sqrt{n}}$. Then $B_n$ is the difference between $S_n$ and its expectation, measured in standard deviation units.
Let $X_i$ be an i.i.d. sequence of random variables with finite mean $\mu$ and variance $\sigma^2$.

Write $S_n = \sum_{i=1}^{n} X_i$. So $E[S_n] = n\mu$ and $\text{Var}[S_n] = n\sigma^2$ and $\text{SD}[S_n] = \sigma\sqrt{n}$.

Write $B_n = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$. Then $B_n$ is the difference between $S_n$ and its expectation, measured in standard deviation units.

Central limit theorem:

$$\lim_{n \to \infty} P\{a \leq B_n \leq b\} \to \Phi(b) - \Phi(a).$$
Central limit theorem

Proving the central limit theorem
Outline

Central limit theorem

Proving the central limit theorem
Recall: characteristic functions

- Let $X$ be a random variable.
Recall: characteristic functions

- Let $X$ be a random variable.
- The **characteristic function** of $X$ is defined by
  \[ \phi(t) = \phi_X(t) := E[e^{itX}] \]. Like $M(t)$ except with $i$ thrown in.

  ![Characteristic function formula](image-url)
Let $X$ be a random variable.

The **characteristic function** of $X$ is defined by

$$
\phi(t) = \phi_X(t) := E[e^{itX}].
$$

Like $M(t)$ except with $i$ thrown in.

Recall that by definition $e^{it} = \cos(t) + i \sin(t)$. 
Recall: characteristic functions

- Let $X$ be a random variable.
- The **characteristic function** of $X$ is defined by
  $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with $i$ thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
Recall: characteristic functions

- Let $X$ be a random variable.
- The **characteristic function** of $X$ is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with $i$ thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X \phi_Y$, just as $M_{X+Y} = M_X M_Y$, if $X$ and $Y$ are independent.
Recall: characteristic functions

- Let $X$ be a random variable.
- The **characteristic function** of $X$ is defined by
  \[ \phi(t) = \phi_X(t) := E[e^{itX}] \]. Like $M(t)$ except with $i$ thrown in.
- Recall that by definition $e^{it} = \cos(t) + i \sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X \phi_Y$, just as $M_{X+Y} = M_X M_Y$, if $X$ and $Y$ are independent.
- And $\phi_{aX}(t) = \phi_X(at)$ just as $M_{aX}(t) = M_X(at)$. 
Recall: characteristic functions

- Let $X$ be a random variable.
- The **characteristic function** of $X$ is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with $i$ thrown in.
- Recall that by definition $e^{it} = \cos(t) + i \sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X \phi_Y$, just as $M_{X+Y} = M_X M_Y$, if $X$ and $Y$ are independent.
- And $\phi_{aX}(t) = \phi_X(at)$ just as $M_{aX}(t) = M_X(at)$.
- And if $X$ has an $m$th moment then $E[X^m] = im\phi_X^{(m)}(0)$. 
Recall: characteristic functions

- Let $X$ be a random variable.
- The **characteristic function** of $X$ is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with $i$ thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X \phi_Y$, just as $M_{X+Y} = M_X M_Y$, if $X$ and $Y$ are independent.
- And $\phi_{aX}(t) = \phi_X(at)$ just as $M_{aX}(t) = M_X(at)$.
- And if $X$ has an $m$th moment then $E[X^m] = i^m \phi_X^{(m)}(0)$.
- Characteristic functions are well defined at all $t$ for all random variables $X$. 
Let $X$ be a random variable and $X_n$, a sequence of random variables.
Let $X$ be a random variable and $X_n$ a sequence of random variables.

Say $X_n$ converge in distribution or converge in law to $X$ if
$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$
at all $x \in \mathbb{R}$ at which $F_X$ is continuous.
Let $X$ be a random variable and $X_n$ a sequence of random variables.

Say $X_n$ converge in distribution or converge in law to $X$ if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at all $x \in \mathbb{R}$ at which $F_X$ is continuous.

Recall: the weak law of large numbers can be rephrased as the statement that

$$A_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

converges in law to $\mu$ (i.e., to the random variable that is equal to $\mu$ with probability one) as $n \to \infty$. 

The central limit theorem can be rephrased as the statement that

$$B_n = X_1 + X_2 + \ldots + X_n - \frac{n \mu}{\sigma \sqrt{n}}$$

converges in law to a standard normal random variable as $n \to \infty$. 

Let $X$ be a random variable and $X_n$ a sequence of random variables.

Say $X_n$ converge in distribution or converge in law to $X$ if $\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which $F_X$ is continuous.

Recall: the weak law of large numbers can be rephrased as the statement that $A_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$ converges in law to $\mu$ (i.e., to the random variable that is equal to $\mu$ with probability one) as $n \to \infty$.

The central limit theorem can be rephrased as the statement that $B_n = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma \sqrt{n}}$ converges in law to a standard normal random variable as $n \to \infty$. 
Lévy’s continuity theorem (see Wikipedia): if

$$\lim_{n \to \infty} \phi_{X_n}(t) = \phi_X(t)$$

for all $t$, then $X_n$ converge in law to $X$. 
Lévy’s continuity theorem (see Wikipedia): if

$$\lim_{n \to \infty} \phi_{X_n}(t) = \phi_X(t)$$

for all $t$, then $X_n$ converge in law to $X$.

By this theorem, we can prove the central limit theorem by showing $\lim_{n \to \infty} \phi_{B_n}(t) = e^{-t^2/2}$ for all $t$. 
Lévy’s continuity theorem (see Wikipedia): if

\[ \lim_{n \to \infty} \phi_{X_n}(t) = \phi_X(t) \]

for all \( t \), then \( X_n \) converge in law to \( X \).

By this theorem, we can prove the central limit theorem by showing \( \lim_{n \to \infty} \phi_{B_n}(t) = e^{-t^2/2} \) for all \( t \).

Moment generating function continuity theorem: if moment generating functions \( M_{X_n}(t) \) are defined for all \( t \) and \( n \) and \( \lim_{n \to \infty} M_{X_n}(t) = M_X(t) \) for all \( t \), then \( X_n \) converge in law to \( X \).
Lévy’s continuity theorem (see Wikipedia): if

$$\lim_{n \to \infty} \phi_{X_n}(t) = \phi_X(t)$$

for all $t$, then $X_n$ converge in law to $X$.

By this theorem, we can prove the central limit theorem by showing $\lim_{n \to \infty} \phi_{B_n}(t) = e^{-t^2/2}$ for all $t$.

Moment generating function continuity theorem: if moment generating functions $M_{X_n}(t)$ are defined for all $t$ and $n$ and $\lim_{n \to \infty} M_{X_n}(t) = M_X(t)$ for all $t$, then $X_n$ converge in law to $X$.

By this theorem, we can prove the central limit theorem by showing $\lim_{n \to \infty} M_{B_n}(t) = e^{t^2/2}$ for all $t$. 
Write \( Y = \frac{X - \mu}{\sigma} \). Then \( Y \) has mean zero and variance 1.
Proof of central limit theorem with moment generating functions

- Write $Y = \frac{X - \mu}{\sigma}$. Then $Y$ has mean zero and variance 1.
- Write $M_Y(t) = E[e^{tY}]$ and $g(t) = \log M_Y(t)$. So $M_Y(t) = e^{g(t)}$. 

- $\text{Chain rule: } M_Y'(0) = g'(0) e^{g(0)} = g'(0)$ and $M_Y''(0) = g''(0) e^{g(0)} + g'(0)^2 e^{g(0)} = g''(0) = 1$.

- So $g$ is a nice function with $g(0) = g'(0) = 0$ and $g''(0) = 1$.
- Taylor expansion: $g(t) = t^2 / 2 + o(t^2)$ for $t$ near zero.

- Now $B_n$ is $\sqrt{n}$ times the sum of $n$ independent copies of $Y$.
- So $M_{B_n}(t) = \left( M_Y(t/\sqrt{n}) \right)^n = e^{ng(t/\sqrt{n})}$.

- But $e^{ng(t/\sqrt{n})} \approx e^{nt^2/2}$ in sense that LHS tends to $e^{t^2/2}$ as $n$ tends to infinity.
Proof of central limit theorem with moment generating functions

- Write $Y = \frac{X-\mu}{\sigma}$. Then $Y$ has mean zero and variance 1.
- Write $M_Y(t) = E[e^{tY}]$ and $g(t) = \log M_Y(t)$. So $M_Y(t) = e^{g(t)}$.
- We know $g(0) = 0$. Also $M_Y'(0) = E[Y] = 0$ and $M_Y''(0) = E[Y^2] = \text{Var}[Y] = 1$. 

Chain rule:

- $M_Y'(0) = g'(0)e^{g(0)} = g'(0)$ and $M_Y''(0) = g''(0)e^{g(0)} + g'(0)^2 e^{g(0)} = g''(0)$.

Taylor expansion:

- $g(t) = t^2/2 + o(t^2)$ for $t$ near zero.

Now $B_n$ is $1/\sqrt{n}$ times the sum of $n$ independent copies of $Y$.

- So $M_{B_n}(t) = (M_Y(t/\sqrt{n}))^n = e^{ng(t/\sqrt{n})}$.

But $e^{ng(t/\sqrt{n})} \approx e^{n(t/\sqrt{n})^2/2} = e^{t^2/2}$, in sense that LHS tends to $e^{t^2/2}$ as $n$ tends to infinity.
Proof of central limit theorem with moment generating functions

- Write $Y = \frac{X - \mu}{\sigma}$. Then $Y$ has mean zero and variance 1.
- Write $M_Y(t) = E[e^{tY}]$ and $g(t) = \log M_Y(t)$. So $M_Y(t) = e^{g(t)}$.
- We know $g(0) = 0$. Also $M'_Y(0) = E[Y] = 0$ and $M''_Y(0) = E[Y^2] = \text{Var}[Y] = 1$.
- Chain rule: $M'_Y(0) = g'(0)e^{g(0)} = g'(0) = 0$ and $M''_Y(0) = g''(0)e^{g(0)} + g'(0)^2e^{g(0)} = g''(0) = 1$. 

Now $B_n$ is $\sqrt{n}$ times the sum of $n$ independent copies of $Y$. 

So $M_{B_n}(t) = (M_Y(t/\sqrt{n}))^n = e^{ng(t/\sqrt{n})}$.

But $e^{ng(t/\sqrt{n})} \approx e^{n(g(t/\sqrt{n})^2/2} = e^{t^2/2}$, in sense that LHS tends to $e^{t^2/2}$ as $n$ tends to infinity.
Proof of central limit theorem with moment generating functions

- Write $Y = \frac{X - \mu}{\sigma}$. Then $Y$ has mean zero and variance 1.
- Write $M_Y(t) = E[e^{tY}]$ and $g(t) = \log M_Y(t)$. So $M_Y(t) = e^{g(t)}$.
- We know $g(0) = 0$. Also $M'_Y(0) = E[Y] = 0$ and $M''_Y(0) = E[Y^2] = \text{Var}[Y] = 1$.
- Chain rule: $M'_Y(0) = g'(0)e^{g(0)} = g'(0) = 0$ and $M''_Y(0) = g''(0)e^{g(0)} + g'(0)^2e^{g(0)} = g''(0) = 1$.
- So $g$ is a nice function with $g(0) = g'(0) = 0$ and $g''(0) = 1$. Taylor expansion: $g(t) = t^2/2 + o(t^2)$ for $t$ near zero.
Proof of central limit theorem with moment generating functions

- Write $Y = \frac{X - \mu}{\sigma}$. Then $Y$ has mean zero and variance 1.
- Write $M_Y(t) = E[e^{tY}]$ and $g(t) = \log M_Y(t)$. So $M_Y(t) = e^{g(t)}$.
- We know $g(0) = 0$. Also $M_Y'(0) = E[Y] = 0$ and $M_Y''(0) = E[Y^2] = \text{Var}[Y] = 1$.
- Chain rule: $M_Y'(0) = g'(0)e^{g(0)} = g'(0) = 0$ and $M_Y''(0) = g''(0)e^{g(0)} + g'(0)^2e^{g(0)} = g''(0) = 1$.
- So $g$ is a nice function with $g(0) = g'(0) = 0$ and $g''(0) = 1$. Taylor expansion: $g(t) = t^2/2 + o(t^2)$ for $t$ near zero.
- Now $B_n$ is $\frac{1}{\sqrt{n}}$ times the sum of $n$ independent copies of $Y$. 
Proof of central limit theorem with moment generating functions

- Write $Y = \frac{X - \mu}{\sigma}$. Then $Y$ has mean zero and variance 1.
- Write $M_Y(t) = E[e^{tY}]$ and $g(t) = \log M_Y(t)$. So $M_Y(t) = e^{g(t)}$.
- We know $g(0) = 0$. Also $M'_Y(0) = E[Y] = 0$ and $M''_Y(0) = E[Y^2] = \text{Var}[Y] = 1$.
- Chain rule: $M'_Y(0) = g'(0)e^{g(0)} = g'(0) = 0$ and $M''_Y(0) = g''(0)e^{g(0)} + g'(0)^2e^{g(0)} = g''(0) = 1$.
- So $g$ is a nice function with $g(0) = g'(0) = 0$ and $g''(0) = 1$.
- Taylor expansion: $g(t) = t^2/2 + o(t^2)$ for $t$ near zero.
- Now $B_n$ is $\frac{1}{\sqrt{n}}$ times the sum of $n$ independent copies of $Y$.
- So $M_{B_n}(t) = (M_Y(t/\sqrt{n}))^n = e^{ng(\frac{t}{\sqrt{n}})}$. 
Proof of central limit theorem with moment generating functions

- Write $Y = \frac{X - \mu}{\sigma}$. Then $Y$ has mean zero and variance 1.
- Write $M_Y(t) = E[e^{tY}]$ and $g(t) = \log M_Y(t)$. So $M_Y(t) = e^{g(t)}$.
- We know $g(0) = 0$. Also $M'_Y(0) = E[Y] = 0$ and $M''_Y(0) = E[Y^2] = \text{Var}[Y] = 1$.
- Chain rule: $M'_Y(0) = g'(0)e^{g(0)} = g'(0) = 0$ and $M''_Y(0) = g''(0)e^{g(0)} + g'(0)^2e^{g(0)} = g''(0) = 1$.
- So $g$ is a nice function with $g(0) = g'(0) = 0$ and $g''(0) = 1$.
- Taylor expansion: $g(t) = t^2/2 + o(t^2)$ for $t$ near zero.
- Now $B_n$ is $\frac{1}{\sqrt{n}}$ times the sum of $n$ independent copies of $Y$.
- So $M_{B_n}(t) = (M_Y(t/\sqrt{n}))^n = e^{ng(\frac{t}{\sqrt{n}})}$.
- But $e^{ng(\frac{t}{\sqrt{n}})} \approx e^{n(\frac{t}{\sqrt{n}})^2/2} = e^{t^2/2}$, in sense that LHS tends to $e^{t^2/2}$ as $n$ tends to infinity.
Proof of central limit theorem with characteristic functions

- Moment generating function proof only applies if the moment generating function of $X$ exists.
Proof of central limit theorem with characteristic functions

- Moment generating function proof only applies if the moment generating function of $X$ exists.
- But the proof can be repeated almost verbatim using characteristic functions instead of moment generating functions.
Proof of central limit theorem with characteristic functions

- Moment generating function proof only applies if the moment generating function of $X$ exists.
- But the proof can be repeated almost verbatim using characteristic functions instead of moment generating functions.
- Then it applies for any $X$ with finite variance.
Write \( \phi_Y(t) = e^{g(t)} \) and \( g(t) = \log \phi_Y(t) \). So \( \phi_Y(t) = e^{g(t)} \).

We know \( g(0) = 0 \). Also \( \phi_Y'(0) = i E[Y] = 0 \) and \( \phi_Y''(0) = i 2 E[Y^2] = -\text{Var}[Y] = -1 \).

Chain rule: \( \phi_Y'(0) = g'(0) e^{g(0)} = g'(0) = 0 \) and \( \phi_Y''(0) = g''(0) e^{g(0)} + g'(0)^2 e^{g(0)} = g''(0) = -1 \).

So \( g \) is a nice function with \( g(0) = g'(0) = 0 \) and \( g''(0) = -1 \). Taylor expansion: \( g(t) = -t^2/2 + o(t^2) \) for \( t \) near zero.

Now \( B_n \) is \( 1/\sqrt{n} \) times the sum of \( n \) independent copies of \( Y \).

So \( \phi_{B_n}(t) = \left( \phi_Y(t/\sqrt{n}) \right)^n = e^{ng(t/\sqrt{n})} \).

But \( e^{ng(t/\sqrt{n})} \approx e^{-n(t/\sqrt{n})^2/2} = e^{-t^2/2} \), in sense that LHS tends to \( e^{-t^2/2} \) as \( n \) tends to infinity.
Write $\phi_Y(t) = E[e^{itY}]$ and $g(t) = \log \phi_Y(t)$. So $\phi_Y(t) = e^{g(t)}$. 
Write $\phi_Y(t) = E[e^{itY}]$ and $g(t) = \log \phi_Y(t)$. So $\phi_Y(t) = e^{g(t)}$.

We know $g(0) = 0$. Also $\phi'_Y(0) = iE[Y] = 0$ and $\phi''_Y(0) = i^2E[Y^2] = -\text{Var}[Y] = -1$. 
Write $\phi_Y(t) = E[e^{itY}]$ and $g(t) = \log \phi_Y(t)$. So $\phi_Y(t) = e^{g(t)}$.

We know $g(0) = 0$. Also $\phi'_Y(0) = iE[Y] = 0$ and $\phi''_Y(0) = i^2 E[Y^2] = -\text{Var}[Y] = -1$.

Chain rule: $\phi'_Y(0) = g'(0)e^{g(0)} = g'(0) = 0$ and $\phi''_Y(0) = g''(0)e^{g(0)} + g'(0)^2 e^{g(0)} = g''(0) = -1$. 

So $g$ is a nice function with $g(0) = g'(0) = 0$ and $g''(0) = -1$. Taylor expansion: $g(t) = -\frac{t^2}{2} + o(t^2)$ for $t$ near zero.

Now $B_n$ is $1/\sqrt{n}$ times the sum of $n$ independent copies of $Y$. So $\phi_{B_n}(t) = \left(\phi_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n = e^{ng\left(t\sqrt{n}\right)}$.

But $e^{ng\left(t\sqrt{n}\right)} \approx e^{-n\left(t\sqrt{n}\right)^2/2} = e^{-t^2/2}$, in sense that LHS tends to $e^{-t^2/2}$ as $n$ tends to infinity.
Write $\phi_Y(t) = E[e^{itY}]$ and $g(t) = \log \phi_Y(t)$. So $\phi_Y(t) = e^{g(t)}$.

We know $g(0) = 0$. Also $\phi'_Y(0) = iE[Y] = 0$ and $\phi''_Y(0) = i^2E[Y^2] = -\text{Var}[Y] = -1$.

Chain rule: $\phi'_Y(0) = g'(0)e^{g(0)} = g'(0) = 0$ and $\phi''_Y(0) = g''(0)e^{g(0)} + g'(0)^2e^{g(0)} = g''(0) = -1$.

So $g$ is a nice function with $g(0) = g'(0) = 0$ and $g''(0) = -1$. Taylor expansion: $g(t) = -t^2/2 + o(t^2)$ for $t$ near zero.
Write \( \phi_Y(t) = E[e^{itY}] \) and \( g(t) = \log \phi_Y(t) \). So \( \phi_Y(t) = e^{g(t)} \).

We know \( g(0) = 0 \). Also \( \phi'_Y(0) = iE[Y] = 0 \) and \( \phi''_Y(0) = i^2E[Y^2] = -\text{Var}[Y] = -1 \).

Chain rule: \( \phi'_Y(0) = g'(0)e^{g(0)} = g'(0) = 0 \) and \( \phi''_Y(0) = g''(0)e^{g(0)} + g'(0)^2e^{g(0)} = g''(0) = -1 \).

So \( g \) is a nice function with \( g(0) = g'(0) = 0 \) and \( g''(0) = -1 \). Taylor expansion: \( g(t) = -t^2/2 + o(t^2) \) for \( t \) near zero.

Now \( B_n \) is \( \frac{1}{\sqrt{n}} \) times the sum of \( n \) independent copies of \( Y \).
Write $\phi_Y(t) = E[e^{itY}]$ and $g(t) = \log \phi_Y(t)$. So $\phi_Y(t) = e^{g(t)}$.

We know $g(0) = 0$. Also $\phi'_Y(0) = iE[Y] = 0$ and $\phi''_Y(0) = i^2 E[Y^2] = -\text{Var}[Y] = -1$.

Chain rule: $\phi'_Y(0) = g'(0)e^{g(0)} = g'(0) = 0$ and $\phi''_Y(0) = g''(0)e^{g(0)} + g'(0)^2 e^{g(0)} = g''(0) = -1$.

So $g$ is a nice function with $g(0) = g'(0) = 0$ and $g''(0) = -1$. Taylor expansion: $g(t) = -t^2/2 + o(t^2)$ for $t$ near zero.

Now $B_n$ is $\frac{1}{\sqrt{n}}$ times the sum of $n$ independent copies of $Y$.

So $\phi_{B_n}(t) = (\phi_Y(t/\sqrt{n}))^n = e^{ng(\frac{t}{\sqrt{n}})}$. 
Write $\phi_Y(t) = E[e^{itY}]$ and $g(t) = \log \phi_Y(t)$. So $\phi_Y(t) = e^{g(t)}$.

We know $g(0) = 0$. Also $\phi'_Y(0) = iE[Y] = 0$ and $\phi''_Y(0) = i^2 E[Y^2] = -\text{Var}[Y] = -1$.

Chain rule: $\phi'_Y(0) = g'(0)e^{g(0)} = g'(0) = 0$ and $\phi''_Y(0) = g''(0)e^{g(0)} + g'(0)^2 e^{g(0)} = g''(0) = -1$.

So $g$ is a nice function with $g(0) = g'(0) = 0$ and $g''(0) = -1$. Taylor expansion: $g(t) = -t^2/2 + o(t^2)$ for $t$ near zero.

Now $B_n$ is $\frac{1}{\sqrt{n}}$ times the sum of $n$ independent copies of $Y$.

So $\phi_{B_n}(t) = (\phi_Y(t/\sqrt{n}))^n = e^{ng\left(\frac{t}{\sqrt{n}}\right)}$.

But $e^{ng\left(\frac{t}{\sqrt{n}}\right)} \approx e^{-n\left(\frac{t}{\sqrt{n}}\right)^2/2} = e^{-t^2/2}$, in sense that LHS tends to $e^{-t^2/2}$ as $n$ tends to infinity.
Perspective

The central limit theorem is actually fairly robust. Variants of the theorem still apply if you allow the $X_i$ not to be identically distributed, or not to be completely independent.
The central limit theorem is actually fairly robust. Variants of the theorem still apply if you allow the $X_i$ not to be identically distributed, or not to be completely independent.

We won’t formulate these variants precisely in this course.
The central limit theorem is actually fairly robust. Variants of the theorem still apply if you allow the $X_i$ not to be identically distributed, or not to be completely independent.

We won’t formulate these variants precisely in this course.

But, roughly speaking, if you have a lot of little random terms that are “mostly independent” — and no single term contributes more than a “small fraction” of the total sum — then the total sum should be “approximately” normal.

Example: if height is determined by lots of little mostly independent factors, then people’s heights should be normally distributed.

Not quite true... certain factors by themselves can cause a person to be a whole lot shorter or taller. Also, individual factors not really independent of each other.

Kind of true for homogenous population, ignoring outliers.
The central limit theorem is actually fairly robust. Variants of the theorem still apply if you allow the $X_i$ not to be identically distributed, or not to be completely independent.

We won’t formulate these variants precisely in this course.

But, roughly speaking, if you have a lot of little random terms that are “mostly independent” — and no single term contributes more than a “small fraction” of the total sum — then the total sum should be “approximately” normal.

Example: if height is determined by lots of little mostly independent factors, then people’s heights should be normally distributed.
The central limit theorem is actually fairly robust. Variants of the theorem still apply if you allow the $X_i$ not to be identically distributed, or not to be completely independent.

We won’t formulate these variants precisely in this course.

But, roughly speaking, if you have a lot of little random terms that are “mostly independent” — and no single term contributes more than a “small fraction” of the total sum — then the total sum should be “approximately” normal.

Example: if height is determined by lots of little mostly independent factors, then people’s heights should be normally distributed.

Not quite true... certain factors by themselves can cause a person to be a whole lot shorter or taller. Also, individual factors not really independent of each other.
The central limit theorem is actually fairly robust. Variants of the theorem still apply if you allow the $X_i$ not to be identically distributed, or not to be completely independent.

We won’t formulate these variants precisely in this course.

But, roughly speaking, if you have a lot of little random terms that are “mostly independent” — and no single term contributes more than a “small fraction” of the total sum — then the total sum should be “approximately” normal.

Example: if height is determined by lots of little mostly independent factors, then people’s heights should be normally distributed.

Not quite true... certain factors by themselves can cause a person to be a whole lot shorter or taller. Also, individual factors not really independent of each other.

*Kind of* true for homogenous population, ignoring outliers.