Introduction to Probability
Lectures 9: Central Limit Theorem
Mateja Jamnik, Thomas Sauerwald

University of Cambridge, Department of Computer Science and Technology
email: {mateja.jamnik, thomas.sauerwald}@cl.cam.ac.uk
Recap: Weak Law of Large Numbers

Central Limit Theorem

Illustrations

Examples

Bonus Material (non-examinable)
Weak Law of Large Numbers (4/4)

Weak Law of Large Numbers: For any $\epsilon > 0$,

$$\lim_{n \to \infty} P \left[ |\bar{X}_n - \mu| > \epsilon \right] = 0 \quad \Rightarrow \quad \exists N: \forall n \geq N: P \left[ |\bar{X}_n - \mu| > 0.2 \right] \leq 0.25$$

$$N = \frac{\sigma^2}{\delta \epsilon^2} = \frac{1}{0.25 \cdot 0.2^2} = 100$$

WLLN: probability for any $\bar{X}_n$ to be outside $[-0.2, 0.2]$ is at most 0.25 for any $n \geq 100$.

Central Limit Theorem will characterise the entire distribution of $\bar{X}_n$ for large $n$!
Outline

Recap: Weak Law of Large Numbers

Central Limit Theorem

Illustrations

Examples

Bonus Material (non-examinable)
Towards the CLT: Finding the Right Scaling

- Let $X_1, X_2, \ldots$ i.i.d. with $\mu = 0$ and finite $\sigma^2$

  **The Sum**
  - Let $\tilde{X}_n := \sum_{i=1}^{n} X_i$ (often denoted by $S_n$)
  - The variance is $V[\tilde{X}_n] = n\sigma^2 \to \infty$

  **The Sample Average (Sample Mean)**
  - Let $\bar{X}_n := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$
  - The variance is $V[\bar{X}_n] = \sigma^2 / n \to 0$

  **The “Proper” Scaling (Standardising)**
  - Let $Z_n := \frac{1}{\sqrt{n}\sigma} \cdot \sum_{i=1}^{n} X_i$
  - The variance is $V[Z_n] = 1$
Central Limit Theorem

Let $X_1, X_2, \ldots$ be any sequence of independent identically distributed random variables with finite expectation $\mu$ and finite variance $\sigma^2$. Let

$$Z_n := \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma} = \frac{1}{\sqrt{n} \cdot \sigma} \left( \sum_{i=1}^{n} X_i - n \cdot \mu \right)$$

Then for any number $a \in \mathbb{R}$, it holds that

$$\lim_{n \to \infty} F_{Z_n}(a) = \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} \, dx,$$

where $\Phi$ is the distribution function of the $\mathcal{N}(0, 1)$ distribution.

In words: the distribution of $Z_n$ always converges to the distribution function $\Phi$ of the standard normal distribution.
Comments on the CLT

- one of the most remarkable results in probability/statistics
- extremely powerful tool in applications: we may not know the actual distribution in real-world, and CLT says we don’t have to(!)
- applies also to sums of random variables which may be unbounded
- adding up independent noises in measurements leads to an error following the Normal distribution

- catch: the CLT only holds *approximately*, i.e., for large $n$

When is the approximation good?

- usually $n \geq 10$ or $n \geq 15$ is sufficient in practice
- approximation tends to be worse when threshold $a$ is far from 0, distribution of $X_i$’s asymmetric, bimodal or discrete
Outline

Recap: Weak Law of Large Numbers

Central Limit Theorem

Illustrations

Examples

Bonus Material (non-examinable)
Illustration of CLT (1/4)

\[ P \left[ \sum_{j=1}^{\infty} X_j = x \right] \]

- \( \mu = \frac{1}{3} \cdot (-1) + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 = 0 \)
- \( \sigma^2 = \frac{1}{3} \cdot (-1)^2 + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1^2 = \frac{2}{3} \)

By the CLT:

\[ Z_n = \frac{1}{\sqrt{n} \cdot \sigma} \cdot \left( \sum_{i=1}^{n} X_i - n \cdot \mu \right) \xrightarrow{n \to \infty} Z \sim \mathcal{N}(0, 1) \]

\[ \Rightarrow \sum_{i=1}^{n} X_i \approx \sqrt{n} \cdot \sigma Z \sim \mathcal{N}(0, n \cdot \sigma^2) \]
Illustration of CLT (2/4)

\[
\Pr \left[ \sum_{j=1}^{10} X_j = x \right]
\]

- \( \mu = 0.15 \cdot (-3) + 0.1 \cdot (-2) + 0.05 \cdot (-1) + 0.7 \cdot 1 = 0 \)
- \( \sigma^2 = 0.15 \cdot 9 + 0.1 \cdot 4 + 0.05 \cdot 1 + 0.7 \cdot 1 = 2.5 \)
Illustration of CLT (3/4) (example from Lecture 8)

\[ P \left[ \sum_{j=1}^{\infty} X_j \leq x \right] \]

- \( \mu = \frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot 1 = 0 \)
- \( \sigma^2 = \frac{1}{2} \cdot (-1)^2 + \frac{1}{2} \cdot 1^2 = 1 \)

Recall: CLT only guarantees convergence of the cumulative distribution!
Illustration of CLT (4/4) (example from Lecture 8 cntd.)

\[ P \left[ \sum_{j=1}^{n} X_j \leq x \right] \]

- \( \mu = \frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot 1 = 0 \)
- \( \sigma^2 = \frac{1}{2} \cdot (-1)^2 + \frac{1}{2} \cdot 1^2 = 1 \)
Illustration of CLT with Standardising

Fig. 14.2. Densities of standardized averages $Z_n$. Left column: from a gamma density; right column: from a bimodal density. Dotted line: $N(0, 1)$ probability density.

Source: Deeking et al., Modern Introduction to Statistics
Outline

Recap: Weak Law of Large Numbers

Central Limit Theorem

Illustrations

Examples

Bonus Material (non-examinable)
Recall: Standard Normal Table

Source: Ross, Probability 8th ed.

\[ Z \sim \mathcal{N}(0, 1) \]

\[ P \left[ Z \leq x \right] = \Phi(x) \]

**Question:** What if we need \( \Phi(x) \) for negative \( x \)?

Due to symmetry of density we have \( \Phi(x) = 1 - \Phi(-x) \).
Suppose you are attending a multiple-choice exam of 10 questions and you are completely unprepared. Each question has 4 choices, and you are going to pass the exam if you guess at least 6 correct answers. Use the normal approximation to estimate the probability of passing.

Let $X \sim Bin(10, 1/4)$. We are interested in $P [ X \geq 6 ]$.

Note $X := \sum_{i=1}^{n} X_i$, where each $X_i \sim Ber(p)$ and $n = 10$, $p = 1/4$.

$\Rightarrow \mu = 1/4$ and $\sigma^2 = p(1 - p) = 3/16$.

Applying the CLT yields:

$$P [ X \geq 6 ] = P \left[ \sum_{i=1}^{n} X_i \geq 6 \right] = P \left[ \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \geq \frac{6 - n\mu}{\sqrt{n}\sigma} \right]$$

A better approximation is obtained by $P \left[ \sum_{i=1}^{n} X_i \geq 5.5 \right] \approx \approx 0.0143$

True value is 0.0197. Error lies in the discretisation!
Approximation of the Binomial Distribution

- Let $X \sim Bin(50, 1/2)$
- Hence $\mu = 25, \sigma^2 = 50 \cdot 1/4 = 12.5$

How good is the approximation by the CLT?

- Let $Y \sim \mathcal{N}(25, 12.5)$
- $P[X \leq x] \approx P[Y \leq x] \Rightarrow$ reasonable approximation, but some error

![Graph showing the approximate distribution of X and Y]

$P[X \leq x]$
Approximation of the Binomial Distribution

- Let $X \sim Bin(50, 1/2)$
- Hence $\mu = 25, \sigma^2 = 50 \cdot 1/4 = 12.5$

How good is the approximation by the CLT?

- Let $Y \sim \mathcal{N}(25, 12.5)$
- $P[ X \leq x ] \approx P[ Y \leq x ] \leadsto$ reasonable approximation, but some error
- $P[ X \leq x ] \approx P[ Y \leq x + 0.5 ] \leadsto$ very tight approximation!
A “Reverse” Application of the CLT

Example 2

Suppose we are sequentially loading one container with packets, whose weights are i.i.d. exponential variables with parameter $\lambda = 1/2$. The container has a capacity of 100 weight units. How many packets can we load so that we meet the capacity threshold with at least .95 probability?

Answer

- We have $X_1, X_2, \ldots, X_n \sim \text{Exp}(1/2)$, where $n$ is unknown.
- Recall that $\mu = \sigma = 2$.
- By the CLT,

$$
P \left[ \sum_{i=1}^{n} X_i \geq 100 \right] = P \left[ \frac{\sum_{i=1}^{n} X_i - 2n}{2\sqrt{n}} \geq \frac{100 - 2n}{2\sqrt{n}} \right]
$$

$$
\approx 1 - \Phi \left( \frac{100 - 2n}{2\sqrt{n}} \right) \doteq 0.05.
$$

- Using a normal table (looking for value 0.95) yields: $\frac{100-2n}{2\sqrt{n}} = 1.645$.

⇒ Solving the quadratic gives $n \leq 39.6$.

- No continuity correction ($100 \sim 99.5$) here, as $\sum_{i=1}^{n} X_i$ is continuous.
A Sample of 100 Exponential Random Variables $Exp(1/2)$

$$\sum_{i=1}^{100} X_i = 214.662$$
Consider $n = 100$ independent coin flips. Estimate the probability that the number of heads is greater or equal than 75.

- **Markov:** $X = \sum_{i=1}^{100} X_i$, $X_i \in \{0, 1\}$ and $E[X] = 100 \cdot \frac{1}{2} = 50$.  
  \[
P[X \geq 3/2 \cdot E[X]] \leq 2/3 = 0.666.
\]

- **Chebyshev:** $V[X] = \sum_{i=1}^{100} V[X_i] = 100 \cdot (1/2)^2 = 25$.  
  \[
P[|X - \mu| \geq 25] \leq \frac{V[X]}{25^2} = \frac{1}{25} = 0.04.
\]

- **Central Limit Theorem:** First standardise: $Z_n = \frac{X - n \cdot 1/2}{\sqrt{n} \cdot 1/2}$  
  \[
P[X \geq 75] = P \left[ Z_n \geq \frac{75 - n \cdot 1/2}{\sqrt{n} \cdot 1/2} \right] \approx 1 - \Phi(5) = 0.0000002866\ldots
\]
  
  exact probability is 0.0000002818\ldots

- **Addendum:** Replacing 75 by 74.5:  
  - This leads to $1 - \Phi(4.9) = 0.000000479\ldots$
  - Issue: threshold too large ($P[X \geq a] \approx P[X = a]$) $\Rightarrow$ CLT less precise  
  - In this region, 75 gives a better approximation than 74.5, but for smaller values (e.g., $\leq 63$) the “.5-shift” gives significantly better results.

As $X$ is symmetric, we could deduce probability is at most 0.02.

CLT gives a much better result (but relies on i.i.d. assumption)
A Distribution whose Average does not converge

\( \text{Cau}(2, 1) \) distribution, Source: Deeking et al., Modern Introduction to Statistics

The Cauchy distribution has “too heavy” tails (no expectation), in particular the average does not converge.
Outline

Recap: Weak Law of Large Numbers

Central Limit Theorem

Illustrations

Examples

Bonus Material (non-examinable)
Towards a Proof of CLT: Moment Generating Functions

Moment-Generating Function

The moment-generating function of a random variable $X$ is

$$M_X(t) = E\left[e^{tX}\right], \quad \text{where} \ t \in \mathbb{R}.$$ 

Using power series of $e$ and differentiating shows that $M_X(t)$ encapsulates all moments of $X$, i.e., $E[X], E[X^2], \ldots$.

Lemma

1. If $X$ and $Y$ are two r.v.’s with $M_X(t) = M_Y(t)$ for all $t \in (-\delta, +\delta)$ for some $\delta > 0$, then the distributions $X$ and $Y$ are identical.
2. If $X$ and $Y$ are independent random variables, then

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t).$$

Proof of 2: (Proof of 1 is quite non-trivial!)

$$M_{X+Y}(t) = E\left[e^{t(X+Y)}\right] = E\left[e^{tX} \cdot e^{tY}\right] \stackrel{(!)}{=} E\left[e^{tX}\right] \cdot E\left[e^{tY}\right] = M_X(t)M_Y(t) \quad \square$$

If $X \sim \mathcal{N}(0, 1)$, then $M_X(t) = \frac{t^2}{2}$. 

Intro to Probability Bonus Material (non-examinable) 23
Proof Sketch of the Central Limit Theorem (1/2)

Proof Sketch:

- Assume w.l.o.g. that $\mu = 0$ and $\sigma = 1$ (if not, scale variables).
- We also assume that the moment generating function of $X_i$, $M(t) = \mathbb{E} \left[ e^{tX_i} \right]$ exists and is finite.
- The moment generating function of $X_i/\sqrt{n}$ is given by
  \[ \mathbb{E} \left[ e^{tX_i/\sqrt{n}} \right] = M(t/\sqrt{n}). \]
- Hence by the Lemma (second statement) from the previous slide,
  \[ \mathbb{E} \left[ \exp \left( \frac{t \sum_{i=1}^{n} X_i}{\sqrt{n}} \right) \right] = \left( M \left( \frac{t}{\sqrt{n}} \right) \right)^n. \]
- Now define $L(t) := \log(M(t))$.
- Differentiating (details omitted here, see book by Ross) shows $L(0) = 0$, $L'(0) = \mu = 0$ and $L''(0) = \mathbb{E} \left[ X^2 \right] = 1.$
Proof Sketch of the Central Limit Theorem (2/2)

Proof Sketch (cntd):
- To prove the theorem, we must show that

\[
\lim_{n \to \infty} \left( M \left( \frac{t}{\sqrt{n}} \right) \right)^n \to e^{t^2/2}
\]

- We take logarithms on both sides and obtain

\[
\lim_{n \to \infty} \frac{L(t/\sqrt{n})}{n^{-1}} = \lim_{n \to \infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}}
\]

\[
= \lim_{n \to \infty} \frac{-L'(t/\sqrt{n})t}{2n^{-1/2}}
\]

\[
= \lim_{n \to \infty} \frac{-L''(t/\sqrt{n})n^{3/2}t^2}{-2n^{-3/2}}
\]

\[
= \lim_{n \to \infty} \left[ -L''(t/\sqrt{n})n^{3/2} \cdot \frac{t^2}{2} \right]
\]

\[
= \frac{t^2}{2}.
\]

We proved that the MGF of \( Z_n \) converges to that one of \( \mathcal{N}(0, 1) \).

This is the moment generating function of \( N(0, 1) \).

Using L'Hopital’s rule.

Using L’Hopital’s rule (again)

We have \( L''(0) = 1! \)