MTL108
Central Limit Theorem (CLT)

Rahul Singh

Many problems in probability and statistics involve repeated measurements, averages, or sums of random phenomena: flipping coins, repeated trials in an experiment, measurements of manufacturing parts, etc. Two fundamental results are:

- The **Law of Large Numbers (LLN)** says that averages converge (in a precise sense) to the theoretical mean: empirical averages are reliable estimates of the population mean when sample size is large.

Example 1 (Bernoulli trials)**.** For a fair coin, $X_i \sim \text{Bernoulli}(0.5)$, $\mu = 0.5$. The probability that the proportion of heads is within 0.1 of 0.5 increases with n:

n	$P(0.4 \leq \overline{X}_n \leq 0.6)$
10	0.65625
50	0.88108
100	0.96480
500	0.99999
1000	1

Example 2 (Beyond IID)**.** Let $X_1, X_2, \ldots$ are independent random variables such that $\mathbb{E}(X_i) = 1$ and $\text{Var}(X_i) = 1 + 1/i$ for $i \in \mathbb{N}$. Does WLLN hold?

- The **Central Limit Theorem (CLT)** says that sums (or properly normalized averages) of many independent random variables behave approximately like a normal (Gaussian) random variable, regardless of the original distribution (under mild conditions).

These results justify many everyday statistical procedures: point estimation, confidence intervals, hypothesis tests, Monte Carlo error estimates, control charts, and many more.

Theorem 1 (IID case)**.** *Let $X_1, X_2, \ldots$ be IID random variables with $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2 \in (0, \infty)$. Furthermore, suppose that the MGF for the X_i, $M_{X_i}(t)$ exists in a neighborhood of $t = 0$ and define*

$$S_n := \sum_{i=1}^{n} X_i.$$

Then as $n \to \infty$,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1),$$

i.e. for each real x,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) = \Phi(x),$$

where Φ is the standard normal CDF.

Remark 1 (Interpretation)**.** In the above theorem, for the sample mean $\bar{X}_n = S_n/n$,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

The CLT says that, after centering and scaling, the sum (or average) behaves approximately normal for large n. This is why normal approximations appear so often even when the underlying data are not normal.

Remark 2. The standard classical CLT assumes that the moment generating function (MGF) $M_X(t) = \mathbb{E}[e^{tX}]$ exists in a neighborhood of $t = 0$. This assumption simplifies the exposition (precise CLT needs only finite variance; characteristic functions give the general proof).

Proof. First, we define a new sequence of IID random variables Y_i by standardizing each X_i,

$$Y_i = \frac{X_i - \mu}{\sigma}.$$

Each Y_i has a mean of $\mathbb{E}[Y_i] = 0$ and a variance of $Var(Y_i) = 1$. We can also express T_n in terms of these new variables:

$$T_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i.$$

Let $M_Y(t) = \mathbb{E}[e^{tY_i}]$ be the MGF of a single standardized variable Y_i. We find the first and second derivatives of $M_Y(t)$ evaluated at $t = 0$,

- $M_Y(0) = \mathbb{E}[e^{0Y_i}] = \mathbb{E}[1] = 1$
- $M_Y'(0) = \mathbb{E}[Y_i] = 0$
- $M_Y''(0) = \mathbb{E}[Y_i^2] = Var(Y_i) + (\mathbb{E}[Y_i])^2 = 1 + 0^2 = 1$

Using a Taylor series expansion of $M_Y(t)$ around $t = 0$, we get:

$$M_Y(t) = M_Y(0) + M_Y'(0)t + \frac{M_Y''(0)}{2!}t^2 + o(t^2)$$

Substituting the values above gives:

$$M_Y(t) = 1 + (0)t + \frac{1}{2}t^2 + o(t^2) = 1 + \frac{t^2}{2} + o(t^2)$$

where $o(t^2)$ denotes a function such that $\lim_{t\to 0} \frac{o(t^2)}{t^2} = 0$.

The MGF of T_n is $M_{T_n}(t) = \mathbb{E}[e^{tT_n}]$. Due to the IID property of the Y_i variables, the MGF of their sum is the product of their individual MGFs:

$$M_{T_n}(t) = E\left[e^{t\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i\right)}\right] = \left[M_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

We now take the limit of $M_{T_n}(t)$ as $n \to \infty$. We substitute the Taylor expansion from the previous step:

$$\lim_{n\to\infty} M_{T_n}(t) = \lim_{n\to\infty} \left[M_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n = \lim_{n\to\infty} \left[1 + \frac{(t/\sqrt{n})^2}{2} + o\left(\left(\frac{t}{\sqrt{n}}\right)^2\right)\right]^n$$

$$= \lim_{n\to\infty} \left[1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n$$

This is a standard limit form $\lim_{n\to\infty}(1 + a_n/n)^n = e^{\lim a_n}$. In our case, $a_n = \frac{t^2}{2} + n \cdot o(\frac{t^2}{n})$. As $n \to \infty$, the term $n \cdot o(\frac{t^2}{n}) \to 0$. Therefore, we have:

$$\lim_{n\to\infty} M_{T_n}(t) = \exp\left\{\lim_{n\to\infty} n \left[\frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]\right\} = \exp\left\{\frac{t^2}{2}\right\}$$

The resulting limiting MGF, $e^{t^2/2}$, is the MGF of a standard normal distribution, $N(0, 1)$. Since MGFs uniquely determine a distribution, the convergence of the MGF of T_n to the MGF of a standard normal distribution implies that T_n converges in distribution to a standard normal distribution. □

Example 3 (Sum of coin flips)**.** Let $X_i \sim \text{Bernoulli}(p)$. Then $\mu = p$, $\sigma^2 = p(1-p)$. For $S_n = \sum X_i$, CLT gives, for large n,

$$\mathbb{P}\left(S_n \le k\right) \approx \Phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right).$$

With $n = 100$ and $p = 0.5$, to approximate $\mathbb{P}(S_{100} \le 60)$,

$$\frac{60 - 50}{\sqrt{25}} = \frac{10}{5} = 2 \quad \Rightarrow \quad \mathbb{P}(S_{100} \le 60) \approx \Phi(2) \approx 0.9772.$$

(Continuity correction is used for better accuracy if n is not very large.)

Example 4 (Exponential data average)**.** Let $X_i \sim \text{Exp}(1)$, $\mu = 1$, $\sigma^2 = 1$. For $n = 30$, approximate $\mathbb{P}(\bar{X}_{30} \le 1.2)$:

$$Z \approx \frac{\bar{X}_{30} - 1}{1/\sqrt{30}} = \sqrt{30}(\bar{X}_{30} - 1).$$

So

$$\mathbb{P}(\bar{X}_{30} \le 1.2) \approx \Phi\left(\sqrt{30} \cdot 0.2\right) = \Phi(1.0954) \approx 0.863.$$

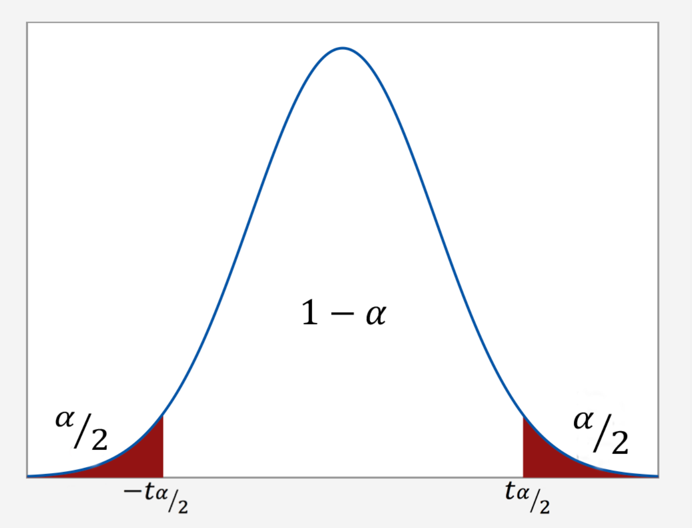
Applications

Application 1 (Confidence intervals)**.** By CLT, for large n,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \text{ approximately follows } \mathcal{N}(0,1)\,.$$

Let $Z \sim \mathcal{N}(0,1)$ and $t_{\alpha/2} > 0$ be a fixed number such that

$$\mathbb{P}(Z > t_{\alpha/2}) = \alpha/2 \quad \textit{and} \quad \mathbb{P}(Z < -t_{\alpha/2}) = \alpha/2$$



So,

$$\mathbb{P}(-t_{\alpha/2} < Z < t_{\alpha/2}) = 1-\alpha \implies \mathbb{P}(|Z| < t_{\alpha/2}) = 1-\alpha.$$

Consequently,

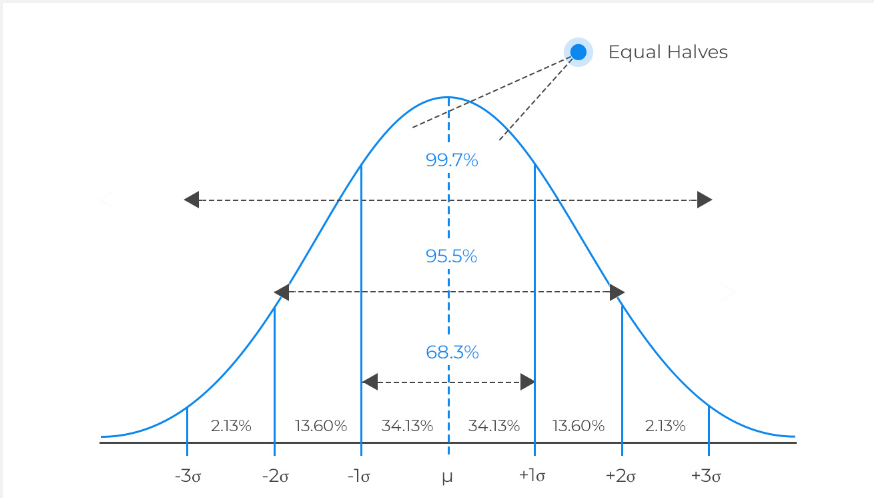
$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right| < t_{\alpha/2}\right) \approx 1-\alpha \\
\implies &\mathbb{P}\left(-t_{\alpha/2} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < t_{\alpha/2}\right) \approx 1-\alpha \\
\implies &\mathbb{P}\left(-t_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}_n - \mu < t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \approx 1-\alpha \\
\implies &\mathbb{P}\left(\bar{X}_n - t_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \approx 1-\alpha \\
\implies &\mathbb{P}\left(\mu \in \left(\bar{X}_n - t_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)\right) \approx 1-\alpha
\end{aligned}$$

In the above expression, the interval $\left(\bar{X}_n - t_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ is a random interval and μ is a fixed parameter. In a real application, μ is unknown and its possible values are of interest; for simplicity, assume that σ is known. This random interval $\left(\bar{X}_n - t_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

contains μ with approximate probability $1-\alpha$, and known as $(1-\alpha)\times 100\%$ confidence interval of unknown parameter μ. Hence a (approximate) $100\times(1-\alpha)\%$ confidence interval for μ is

$$\left(\bar{X}_n - t_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

We have the following information about the probability associated with $\mathcal{N}(\mu,\sigma)$ random variable,



Thus,

$$\mathbb{P}\left(\mu \in \left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}, \bar{X}_n + \frac{\sigma}{\sqrt{n}}\right)\right) \approx 0.683,$$

$$\mathbb{P}\left(\mu \in \left(\bar{X}_n - 2\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 2\frac{\sigma}{\sqrt{n}}\right)\right) \approx 0.955,$$

$$\mathbb{P}\left(\mu \in \left(\bar{X}_n - 3\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 3\frac{\sigma}{\sqrt{n}}\right)\right) \approx 0.997.$$

Application 2 (Hypothesis testing, will not be asked in mid-sem)**.** Many test statistics are (after normalization) sums or averages; CLT gives approximate null distributions enabling p-values.

Application 3 (Monte Carlo integration)**.** If X_i are IID samples and g a function, the Monte Carlo estimator $\frac{1}{n}\sum g(X_i)$ has variance σ_g^2/n, where $\sigma_g^2 = \mathrm{Var}(g(X))$. By CLT,

$$\frac{1}{n}\sum g(X_i) \text{ is approximately distributed as } \mathcal{N}\Big(\mathbb{E}[g(X)],\ \sigma_g^2/n\Big).$$

approximately normal. Therefore, a (approximate) $100\times(1-\alpha)\%$ confidence interval for $\mathbb{E}[g(X)]$ is

$$\left(\frac{1}{n}\sum g(X_i) - \frac{\sigma_g}{\sqrt{n}}\,t_{\alpha/2},\ \frac{1}{n}\sum g(X_i) + \frac{\sigma_g}{\sqrt{n}}\,t_{\alpha/2}\right).$$

So, we have an idea of how much the approximation can vary.

Application 4 (Simple Linear Regression, will not be asked in mid-sem)**.** Simple linear regression is a statistical method used to model the linear relationship between a dependent variable (Y) and a single independent variable (X). The model is defined by the following equation:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

where:

- Y is the dependent (or response) variable.
- X is the independent (or predictor) variable.
- β_0 is the y-intercept, representing the expected value of Y when $X = 0$.
- β_1 is the slope coefficient, representing the change in the expected value of Y for a one-unit change in X.
- ϵ is the random error term, which captures all other factors influencing Y that are not explained by X.

The coefficients β_0 and β_1 are typically estimated from sample data using the Ordinary Least Squares (OLS) method, which minimizes the sum of the squared error terms. The estimated regression line is written as:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X.$$

The primary assumption for performing inference on regression coefficients is that their sampling distributions are normal. While classical regression theory requires the error term (ϵ) to be normally distributed, the CLT provides a more robust justification for large sample sizes. Even if the error distribution is non-normal, the CLT guarantees that the sampling distributions of the OLS estimators ($\hat{\beta}_0$ and $\hat{\beta}_1$) will approach a normal distribution as the sample size increases.
We can construct a confidence interval for the true slope β_1 using the formula:

$$\hat{\beta}_1 \pm t_{\alpha/2} \times \sqrt{\mathrm{Var}(\hat{\beta}_1)}$$

This interval provides a range of plausible values for the true population parameter.

Remark 3 (When CLT is slow)**.** If distributions are highly skewed or n small, normal approximation may be poor.

Remark 4 (When CLT fails)**.** If variance is infinite (heavy tails), classical CLT does not apply.

Properties of Normal Distribution

Definition 1. A random variable X is said to have a normal distribution with mean μ and variance σ^2 (denoted $X \sim N(\mu, \sigma^2)$) if its pdf is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

The cdf is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt.$$

There is no closed form for F, but it is tabulated and available in software (e.g., R's `pnorm`).

Standard Normal Distribution

The standard normal distribution is $Z \sim N(0, 1)$ with pdf

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2},$$

and cdf $\Phi(z) = P(Z \le z)$.

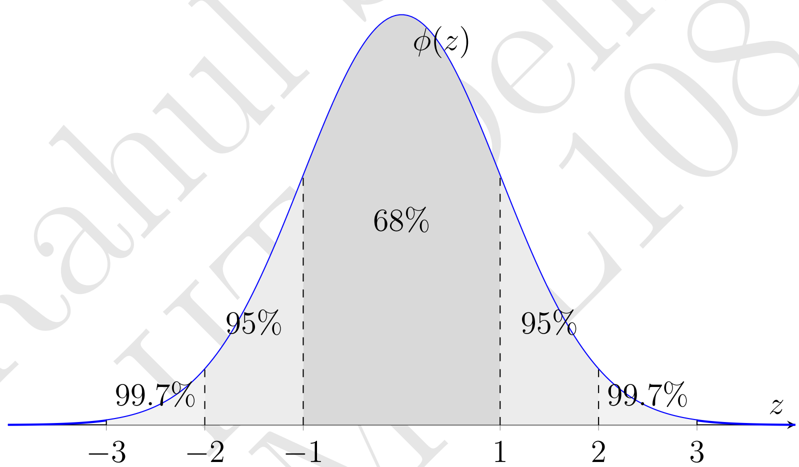


Figure 1: Standard normal pdf with shaded areas showing the 68%, 95%, and 99.7% rules. The darkest region (inner) contains 68% of the probability; the combined middle region (dark + medium) contains 95%; the entire shaded area (all three shades) contains 99.7%.

1. **Symmetry** The pdf $\phi(z)$ is symmetric about 0: $\phi(-z) = \phi(z)$. For the cdf, this gives

   $$\Phi(-z) = 1 - \Phi(z).$$

   This is crucial for finding left-tail probabilities from right-tail tables.

2. **Linear Transformation** If $X \sim N(\mu, \sigma^2)$, then

   $$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

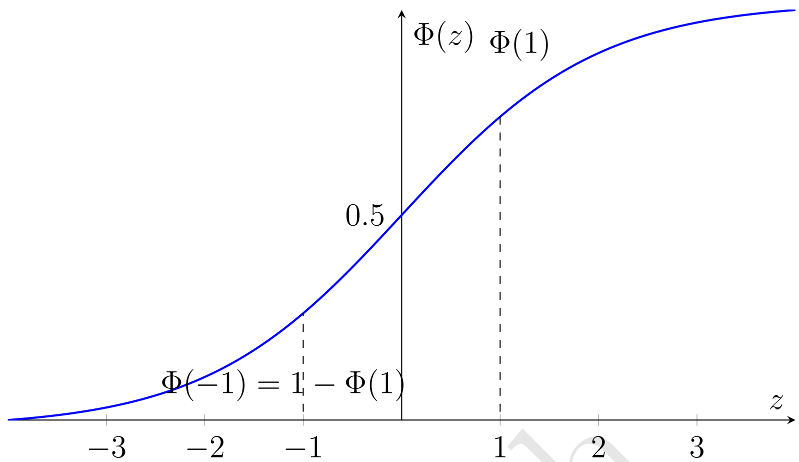


Figure 2: Approximate standard normal cdf (using logistic function) illustrating symmetry: $\Phi(-z) = 1 - \Phi(z)$. The true normal cdf has the same symmetry and a very similar shape.

Conversely, if $Z \sim N(0,1)$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.

This standardization is the core tool for computing any normal probability:

$$P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

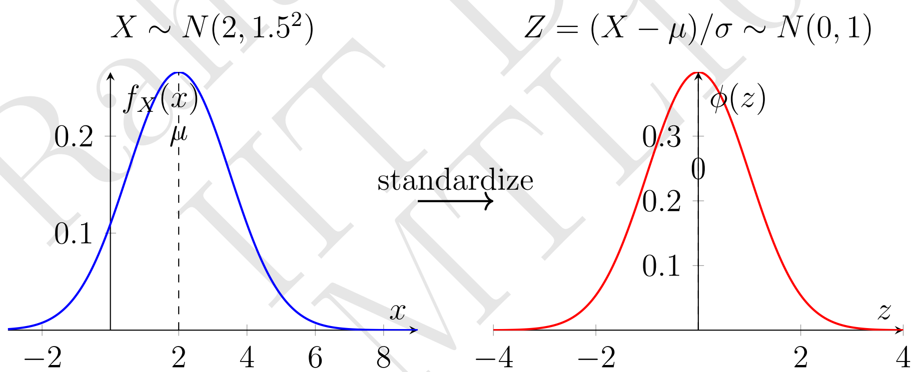


Figure 3: Standardization: shifting and scaling turns any normal distribution into the standard normal.

3. **Percentiles and the 68-95-99.7 Rule** For a standard normal:

$$P(|Z| < 1) \approx 0.6827,$$
$$P(|Z| < 2) \approx 0.9545,$$
$$P(|Z| < 3) \approx 0.9973.$$

More precisely, the 95% interval is $z = \pm 1.96$:

$$P(|Z| < 1.96) \approx 0.95.$$

These numbers are easy to remember and often sufficient for rough estimates.

4. **Tail Probabilities** From symmetry, we get:

$$P(Z > z) = 1 - \Phi(z) = \Phi(-z).$$

Common tail values:

z	$P(Z > z)$
1	0.1587
1.645	0.05
1.96	0.025
2	0.0228
2.326	0.01
2.576	0.005
3	0.00135

Remark 5. Traditionally, one uses a standard normal table that gives $\Phi(z)$ for $z \geq 0$. For negative z, use $\Phi(-z) = 1 - \Phi(z)$.

Application in CLT Problems

Suppose $X_1, \dots, X_n$ are IID with mean μ and variance σ^2. The CLT says:

$$\overline{X}_n \overset{\text{approx}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right), \qquad S_n = \sum_{i=1}^{n} X_i \overset{\text{approx}}{\sim} N(n\mu, n\sigma^2).$$

Example 5. Let S be the sum of 100 IID random variables with $\mu = 5$, $\sigma = 2$. Approximate $P(S > 520)$ and $P(490 \leq S \leq 510)$ using CLT.
Solution: By CLT, S approximately follows $N(100 \cdot 5,\ 100 \cdot 4) = N(500, 20^2)$. Then

$$P(S > 520) = P\left(\frac{S - 500}{20} > \frac{520 - 500}{20}\right) \approx P(Z > 1) = 0.1587.$$

Next,

$$P(490 \leq S \leq 510) \approx P\left(\frac{490 - 500}{20} \leq Z \leq \frac{510 - 500}{20}\right) = P(-0.5 \leq Z \leq 0.5).$$

From symmetry and tables, $P(-0.5 \leq Z \leq 0.5) = 2\Phi(0.5) - 1$. Using $\Phi(0.5) \approx 0.6915$, we get 0.383.

Continuity Correction in CLT for discrete distribution

When we approximate a **discrete** random variable (such as a Binomial or Poisson) by a **continuous** normal distribution using the Central Limit Theorem (CLT), we apply a **continuity correction**.

Why is it needed? A discrete random variable takes values at isolated points (e.g., $0, 1, 2, \dots$), whereas the normal distribution is continuous. Thus, probabilities like $P(X = k)$ or $P(a \leq X \leq b)$ for discrete X correspond to intervals of width 1 on the continuous scale. **To better approximate discrete probabilities using a normal distribution, we shift the boundaries by ± 0.5.**

General Rule (Binomial Example): Let $X \sim \text{Bin}(n, p)$ with

$$\mu = np, \qquad \sigma = \sqrt{np(1-p)}.$$

Using the CLT,

$$Z = \frac{X-\mu}{\sigma} \approx N(0,1).$$

To approximate probabilities:

$$P(X \le k) \approx P\left(Z \le \frac{k+0.5-\mu}{\sigma}\right),$$

$$P(X \ge k) \approx P\left(Z \ge \frac{k-0.5-\mu}{\sigma}\right),$$

$$P(a \le X \le b) \approx P\left(\frac{a-0.5-\mu}{\sigma} \le Z \le \frac{b+0.5-\mu}{\sigma}\right).$$

More clearly,

$$P(X=k) \quad \text{is approximated by} \quad P(k-0.5 < Y < k+0.5), \text{ where } Y \sim N(\mu, \sigma^2).$$

Remark: Continuity correction improves accuracy, especially for moderate sample sizes.

Problems and Solutions

In the following text, $\approx$ refers to approximately follows.

1. An archer hits the target with probability 3/4. What is the chance of him hitting the target at least 885 and at most 930 times in 1200 shots? What is the chance of him hitting the target at least 870 times in 1200 shots?

   Solution: Let X be the number of hits in 1200 shots. Then $X \sim \text{Binomial}(1200, 3/4)$. Mean $\mu = 1200 \times 0.75 = 900$, variance $\sigma^2 = 1200 \times 0.75 \times 0.25 = 225$, so $\sigma = 15$. By the CLT, $X \approx N(900, 15^2)$.

   $$\begin{aligned} P(885 \le X \le 930) &\approx P\left(\frac{885-900}{15} \le Z \le \frac{930-900}{15}\right) = P(-1 \le Z \le 2) \\ &= \Phi(2) - \Phi(-1) = 0.9772 - 0.1587 = 0.8185. \end{aligned}$$

   For the second part,

   $$P(X \ge 870) \approx P\left(Z \ge \frac{870-900}{15}\right) = P(Z \ge -2) = \Phi(2) = 0.9772.$$

   Thus the chances are approximately 0.8185 and 0.9772.

2. Let the probability of a newborn being male is 0.515. Find the probability that among 10000 newborns there are more boys than girls.

Solution: Let X be the number of boys. Then $X \sim \text{Binomial}(10000, 0.515)$. Mean $\mu = 10000 \times 0.515 = 5150$, variance $\sigma^2 = 10000 \times 0.515 \times 0.485 = 2497.75$, $\sigma \approx 49.98$. More boys than girls means $X > 5000$. Using the CLT,

$$P(X > 5000) \approx P\left(Z > \frac{5000 - 5150}{49.98}\right) = P(Z > -3.00) = \Phi(3) \approx 0.9987.$$

(If we use continuity correction, $X > 5000.5$ gives $Z \approx -3.01$, still about 0.9987.)

3. How many times should a fair coin be tossed so that, with probability 0.95, the frequency of heads falls within 0.1 of 1/2?

   Solution: Let n be the number of tosses, X the number of heads, $\hat{p} = X/n$. For a fair coin, $\mu = 0.5$, $\sigma = \sqrt{0.5 \times 0.5} = 0.5$. By the CLT, $\hat{p} \approx N(0.5, 0.25/n)$. We want

   $$P(|\hat{p} - 0.5| < 0.1) = 0.95.$$

   Standardizing,

   $$P\left(|Z| < \frac{0.1}{0.5/\sqrt{n}}\right) = P\left(|Z| < 0.2\sqrt{n}\right) = 0.95.$$

   For a standard normal, $P(|Z| < 1.96) \approx 0.95$. Thus $0.2\sqrt{n} = 1.96$ gives $\sqrt{n} = 9.8$, $n = 96.04$. Since n must be an integer, we take $n = 97$ (or 96; the probability will be slightly less than 0.95 for 96). So about 97 tosses.

4. The probability that a light bulb would last at least 1000 hours is 1/3. Estimate the probability that at least 580 out of 1800 bulbs would last at least 1000 hours.

   Solution: Let X be the number of bulbs lasting at least 1000 hours. Then $X \sim \text{Binomial}(1800, 1/3)$. Mean $\mu = 1800 \times 1/3 = 600$, variance $\sigma^2 = 1800 \times (1/3) \times (2/3) = 400$, $\sigma = 20$. By CLT,

   $$P(X \geq 580) \approx P\left(Z \geq \frac{580 - 600}{20}\right) = P(Z \geq -1) = \Phi(1) \approx 0.8413.$$

   (Using continuity correction, $P(X \geq 579.5)$ gives $Z \approx -1.025$, probability about 0.847.)

5. Let $\chi_n^2 = X_1^2 + \cdots + X_n^2$, where X_i are IID $\mathcal{N}(0,1)$. Argue that

   $$\lim_{n\to\infty} P\left(\frac{\chi_n^2 - n}{\sqrt{2n}} \leq x\right) = \Phi(x).$$

   Solution: Each $Y_i = X_i^2$ is IID with mean $E[Y_i] = 1$ (since $\text{Var}(X_i) = 1$ and $E[X_i^2] = 1$) and variance $\text{Var}(Y_i) = E[X_i^4] - (E[X_i^2])^2 = 3 - 1 = 2$. Then $\chi_n^2 = \sum_{i=1}^n Y_i$. By the CLT,

   $$\frac{\sum Y_i - n \cdot 1}{\sqrt{n} \cdot \sqrt{2}} = \frac{\chi_n^2 - n}{\sqrt{2n}} \xrightarrow{d} N(0,1),$$

   so the limit of its cdf is $\Phi(x)$.

6. Let Y_n be a Poisson random variable with parameter n. Argue that

$$\lim_{n\to\infty} P\left(\frac{Y_n - n}{\sqrt{n}} \leq x\right) = \Phi(x).$$

Solution: A Poisson(n) random variable can be represented as the sum of n IID Poisson(1) variables: $Y_n = \sum_{i=1}^{n} Z_i$ with $Z_i \sim$ Poisson(1). Each Z_i has mean 1 and variance 1. By the CLT,

$$\frac{\sum Z_i - n \cdot 1}{\sqrt{n} \cdot 1} = \frac{Y_n - n}{\sqrt{n}} \xrightarrow{d} N(0,1),$$

hence the limit is $\Phi(x)$.

7. Let $X_1, \ldots, X_{1200}$ be IID random variables, uniformly distributed in $[0,1]$. Check that with a probability of about 0.98, their sum is in the interval $(576.7, 623.3)$.

Solution: For $U[0,1]$, mean $\mu = 0.5$, variance $\sigma^2 = 1/12$. The sum $S = \sum_{i=1}^{1200} X_i$ has mean $1200 \times 0.5 = 600$ and variance $1200 \times 1/12 = 100$, so standard deviation $\sigma_S = 10$. The interval $(576.7, 623.3)$ is 600 ± 23.3. Then

$$23.3/10 = 2.33, \quad P(576.7 < S < 623.3) \approx P(|Z| < 2.33) = 2\Phi(2.33) - 1.$$

From normal tables, $\Phi(2.33) \approx 0.9901$, so the probability is about 0.9802, i.e., approximately 0.98.

8. Estimate $P(\chi^2_{100} > 127.6)$, where χ^2_{100} is a chi-square random variable with 100 degrees of freedom.

Solution: For χ^2_{100}, mean $\mu = 100$, variance $\sigma^2 = 2 \times 100 = 200$, so $\sigma = \sqrt{200} \approx 14.142$. Using the normal approximation (from Problem 5),

$$P(\chi^2_{100} > 127.6) \approx P\left(Z > \frac{127.6 - 100}{14.142}\right) = P(Z > 1.952) \approx 1 - \Phi(1.95).$$

$\Phi(1.95) \approx 0.9744$, so the probability is about 0.0256.

9. A drunkard executes a random walk in the following way: Each minute he takes a step north or south, with probability 1/2 each, and his successive step directions are independent. His step length is 50 cm. Use the central limit theorem to approximate the probability distribution of his location after 1 hour. Where is he most likely to be?

Solution: After 60 minutes, let N be the number of north steps. Then $N \sim$ Binomial$(60, 1/2)$. Each north step gives displacement $+0.5$ m, south step -0.5 m. Net displacement $D = 0.5(2N - 60) = N - 30$ (in meters). Mean $E[D] = 30 - 30 = 0$, variance $\text{Var}(D) = \text{Var}(N) = 60 \times 0.5 \times 0.5 = 15$, so standard deviation $\sqrt{15} \approx 3.87$ m. By the CLT, D is approximately normally distributed with mean 0 and variance 15. The most likely location is at the mean, i.e., 0 meters from the start.

10. Answer Problem 9 under the assumption that the drunkard has some idea of where he wants to go, so that he steps north with probability 2/3 and south with probability 1/3.

Solution: Now $N \sim$ Binomial$(60, 2/3)$. Mean $E[N] = 60 \times 2/3 = 40$, variance $\text{Var}(N) = 60 \times (2/3) \times (1/3) = 40/3 \approx 13.33$, standard deviation $\sqrt{40/3} \approx 3.65$ m. Displacement $D = N - 30$ has mean $40 - 30 = 10$ m, variance $40/3$. Thus D is approximately $N(10, 40/3)$. He is most likely to be near 10 meters north of the starting point.

Conclusions

The CLT is the most famous and widely used result involving convergence in distribution.

- Let $\{X_1, X_2, \dots\}$ be a sequence of IID. random variables with mean μ and variance σ^2.
- The CLT states that the standardized sample mean converges in distribution to a standard normal distribution:
$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$
- This is an extremely powerful result because it tells us that, regardless of the original distribution of the X_i's, the distribution of their average (properly normalized) will approach a normal distribution as the sample size grows.

References

[1] Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability.* Chapman and Hall/CRC.

Disclaimer