Probability

A review, Weak Law of Large Numbers and Central Limit Theorem

Noara Razzak

May 9, 2020

What is Probability? Probability measures **how likely** something is to happen, from:

- **0** (never happens)
- to **1** (always happens) (or 0% to 100%)

Rolling a fair die:

- $P(\text{Even}) = \frac{3}{6} = 0.5$
- $P(\text{Prime}) = \frac{3}{6} = 0.5$

Basic Probability Formula

$$P(\text{Event}) = \frac{\text{Number of ways it can happen}}{\text{Total possible outcomes}}$$

Important rules

Certain Event (100%)

$$P(\text{Sun rises}) = 1$$

Impossible Event (0%)

$$P(\text{Rolling a 7 on standard die}) = 0$$

Complement Rule ("Not")

$$P(\text{Not A}) = 1 - P(A)$$

Conditional Probability ("If...Then")

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Rules of probability

Rule	Formula
Complement	$P(A^c) = 1 - P(A)$
Union	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Mutually Exclusive	$P(A \cap B) = 0$
Independent	$P(A \cap B) = P(A)P(B)$

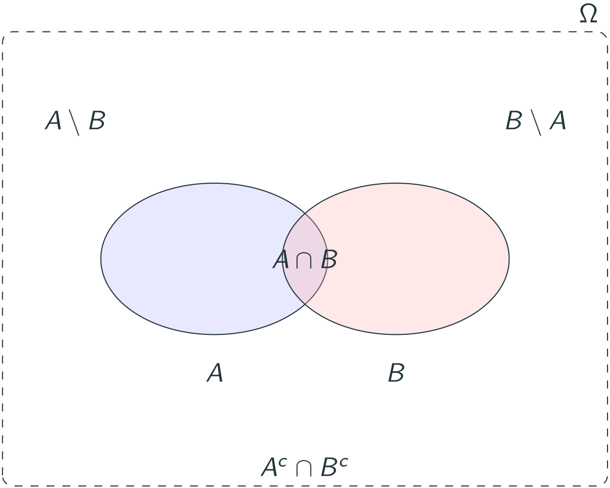
Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' Theorem originates from Conditional Probability

The probability of A *given* B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$


Ω
$A \setminus B$
$B \setminus A$
$A \cap B$
A
B
$A^c \cap B^c$

Real-world uses

- Medical diagnosis
- Spam filtering
- Risk assessment
- Machine learning (Naive Bayes)

Key Insight

- Conditional probability formalizes how we should rationally update beliefs given evidence.

Example 1

A family has two children. You know that at least one of them is a boy. What is the probability that both children are boys?

Example 1 - solution

Step 1: Sample Space The possible gender combinations for two children (older child first) are:

$$S = \{\text{BB}, \text{BG}, \text{GB}, \text{GG}\}$$

Step 2: Apply Condition We know *at least one* is a boy, so we eliminate GG:

$$S' = \{\text{BB}, \text{BG}, \text{GB}\}$$

Step 3: Count Favorable Outcomes Only BB has two boys:

$$\text{Favorable} = 1 \quad (\text{BB})$$

$$\text{Possible} = 3 \quad (\text{BB, BG, GB})$$

Step 4: Calculate Probability

$$P(\text{Both boys}) = \frac{\text{Favorable}}{\text{Possible}} = \frac{1}{3}$$

Example 2

A bag contains 4 red marbles, 5 blue marbles, and 6 green marbles. If 3 marbles are randomly selected without replacement, what is the probability that all three are blue?

- $\frac{1}{91}$
- $\frac{2}{91}$
- $\frac{5}{182}$
- $\frac{10}{273}$
- $\frac{25}{546}$

Example 2 - solution Step 1: Determine Total Marbles First, find the total number of marbles:

$$4 \text{ red} + 5 \text{ blue} + 6 \text{ green} = 15 \text{ marbles}$$

Step 2: Calculate Total Possible Outcomes Number of ways to choose any 3 marbles from 15:

$$\text{Total combinations} = \binom{15}{3} = \frac{15!}{3!(15-3)!} = 455$$

Step 3: Calculate Favorable Outcomes Number of ways to choose 3 blue marbles from 5 available:

$$\text{Favorable combinations} = \binom{5}{3} = \frac{5!}{3!(5-3)!} = 10$$

Step 4: Compute Probability

$$P(3 \text{ blue}) = \frac{\text{Favorable}}{\text{Total}} = \frac{10}{455} = \frac{2}{91}$$

– (Correct Answer)

Weak Law of Large Numbers (WLLN)

The **Weak Law of Large Numbers (WLLN)** is a fundamental theorem in probability theory that describes how the sample average of a sequence of independent and identically distributed (i.i.d.) random variables converges in probability to the expected value (i.e. population mean) as the sample size increases.

Statement of the Theorem

Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with finite mean $\mu = \mathbb{E}[X_i]$ and finite variance $\sigma^2 = \mathrm{Var}(X_i)$. Define the sample mean as:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

The WLLN states that for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\left(\left| \bar{X}_n - \mu \right| \geq \epsilon \right) = 0.$$

In other words, $\bar{X}_n$ **converges in probability** to μ, denoted as:

$$\bar{X}_n \xrightarrow{P} \mu.$$

Example

Suppose $X_1, X_2, \ldots$ are i.i.d. Bernoulli random variables with success probability $p = 0.5$. Then:

$$\mu = \mathbb{E}[X_i] = p = 0.5, \quad \sigma^2 = p(1-p) = 0.25.$$

The WLLN guarantees that for large n, the sample mean $\bar{X}_n$ will be very close to 0.5 with high probability. For instance, if $n = 10,000$ and $\epsilon = 0.01$:

$$\mathbb{P}\left(\left|\bar{X}_n - 0.5\right| \geq 0.01\right) \leq \frac{0.25}{10,000 \times 0.01^2} = 0.25.$$

As n increases, this probability decreases further.

Interpretation

- The WLLN justifies the intuitive idea that averaging many independent observations reduces noise.
- It is "weak" because it guarantees convergence in probability (not almost surely, which is the Strong Law of Large Numbers).
- Applications include:
  - Polling and surveys (sample proportions converge to population proportions).
  - Monte Carlo methods (approximating integrals via random sampling).

Comparison with Strong Law

Weak Law (WLLN)	Strong Law (SLLN)
$\bar{X}_n \xrightarrow{P} \mu$	$\bar{X}_n \xrightarrow{a.s.} \mu$
Convergence in probability	Almost sure convergence
Weaker guarantee	Stronger guarantee

Table 1: WLLN vs. SLLN

Central Limit Theorem

Two fundamental theorems in probability theory—the **Weak Law of Large Numbers (WLLN)** and the **Central Limit Theorem (CLT)**—describe the behavior of sample averages. While the WLLN establishes convergence, the CLT refines this by specifying the *distribution* of the sample mean around the population mean. This document explores their connection.

Central Limit Theorem (CLT)

Under the same conditions (i.i.d. with $\mu, \sigma^2 < \infty$), the CLT states:

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where $\xrightarrow{d}$ denotes convergence in distribution, and $\mathcal{N}(0, \sigma^2)$ is a normal distribution with mean 0 and variance σ^2.

Interpretation The CLT provides a *distributional approximation* for the sample mean:

$$\bar{X}_n \approx \mu + \frac{\sigma}{\sqrt{n}} Z, \quad Z \sim \mathcal{N}(0, 1).$$

This refines the WLLN by quantifying the rate of convergence and the shape of fluctuations.

Mathematical Link

The WLLN can be derived from the CLT:

$$\mathbb{P}\left(\left|\bar{X}_n - \mu\right| \geq \epsilon\right) = \mathbb{P}\left(\left|\sqrt{n}(\bar{X}_n - \mu)\right| \geq \epsilon\sqrt{n}\right) \to 0 \quad \text{as } n \to \infty,$$

because $\epsilon\sqrt{n} \to \infty$, and the CLT ensures $\sqrt{n}(\bar{X}_n - \mu)$ is stochastically bounded

Joint Implications in Statistical Inference -Confidence Intervals

- **WLLN**: Justifies using $\bar{X}_n$ as an estimator for μ.
- **CLT**: Provides the approximate distribution for constructing confidence intervals:

$$\mu \in \left[\bar{X}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

Connection Between WLLN and CLT

- **WLLN**: Ensures $\bar{X}_n$ converges to μ (consistency).
- **CLT**: Describes the *scaled deviations* $\sqrt{n}(\bar{X}_n - \mu)$ (asymptotic normality).

Intuition

- The WLLN shows that $\bar{X}_n - \mu \to 0$, but the CLT reveals that the **rate** is $\mathcal{O}(n^{-1/2})$:
$$\left|\bar{X}_n - \mu\right| \approx \frac{\sigma}{\sqrt{n}}|Z|.$$
- The CLT implies the WLLN (since convergence in distribution to a constant implies convergence in probability), but not vice versa.

Comparison Summary

Feature	WLLN	CLT
Convergence Type	Probability ($\xrightarrow{P}$)	Distribution ($\xrightarrow{d}$)
Rate	$\bar{X}_n - \mu \to 0$	$\sqrt{n}(\bar{X}_n - \mu) \sim \mathcal{N}(0, \sigma^2)$
Key Use	Consistency of estimators	Confidence intervals, hypothesis testing

Table 2: WLLN vs. CLT

Conclusion The WLLN and CLT are deeply connected: the CLT *extends* the WLLN by providing a distributional limit for the sample mean. Together, they underpin modern statistical inference, ensuring both consistency and asymptotic normality of estimators.

Next class we will cover confidence intervals.