MTL108
Convergence of RVs and Law of Large Numbers

Rahul Singh

Markov's and Chebyshev's Inequalities

Markov's and Chebyshev's Inequalities provide bounds on the probability that a random variable. What makes it so powerful is that it applies to *any* probability distribution, provided that the appropriate moments exist.

Theorem 1 (Markov's inequality)**.** *Let Y be a non-negative random variable with $\mathbb{E}[Y] < \infty$. For any constant $a > 0$,*

$$\mathbb{P}(Y \geq a) \leq \frac{\mathbb{E}[Y]}{a}.$$

Proof. Assume Y is a continuous random variable with probability density function $f_Y(y)$. The expectation of Y is given by

$$\mathbb{E}[Y] = \int_0^\infty y f_Y(y)\, dy$$

We can split this integral into two parts, one from 0 to a and one from a to ∞, i.e.,

$$\mathbb{E}[Y] = \int_0^a y f_Y(y)\, dy + \int_a^\infty y f_Y(y)\, dy$$

Since Y is non-negative, the first integral is non-negative. Consequently,

$$\mathbb{E}[Y] \geq \int_a^\infty y f_Y(y)\, dy$$

For all values of y in the second integral (i.e., $y \geq a$), we know that $y \geq a$. Thus, replacing y with a, we get

$$\mathbb{E}[Y] \geq \int_a^\infty a f_Y(y)\, dy = a \int_a^\infty f_Y(y)\, dy$$

The integral on the right, $\int_a^\infty f_Y(y)\, dy$, is exactly the definition of the probability that $Y \geq a$. So we obtain

$$\mathbb{E}[Y] \geq a\, \mathbb{P}(Y \geq a) \quad \Rightarrow \quad \mathbb{P}(Y \geq a) \leq \frac{\mathbb{E}[Y]}{a}.$$

The proof for a discrete random variable is analogous, replacing integrals with summations. □

Theorem 2 (Chebyshev's Inequality)**.** *Let X be a random variable with finite mean $\mu = \mathbb{E}[X]$ and finite variance $\sigma^2 = \mathrm{Var}(X)$. Then for any real number $k > 0$,*

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

For any $\epsilon > 0$ *and by setting* $\epsilon = k\sigma$,

$$\mathbb{P}(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

This means that the probability of a random variable being more than k standard deviations away from its mean is at most $1/k^2$. The inequality gives us a "worst-case scenario" bound for any distribution, which is why it's less precise for well-behaved distributions (like the Normal) compared to the Empirical Rule, but far more general.

This can also be restated as

$$\mathbb{P}(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad \text{and} \quad \mathbb{P}(|X - \mu| \leq \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}.$$

This inequality provides a lower bound on the probability that a random variable falls within a certain number of standard deviations from its mean, regardless of the distribution's shape, provided the mean and variance exist.

Proof. To prove Chebyshev's inequality, we apply Markov's inequality to the non-negative random variable $Y = (X - \mu)^2$.

Define a new random variable, $Y = (X - \mu)^2$. Because it's a squared quantity, Y is always non-negative. The expectation of Y is the variance of X, that is,

$$\mathbb{E}[Y] = \mathbb{E}[(X - \mu)^2] = \sigma^2$$

We are interested in the event

$$\{|X - \mu| \geq k\sigma\} \ = \ \{(X - \mu)^2 \geq (k\sigma)^2\}.$$

Now, using Markov's inequality on Y with the constant $a = (k\sigma)^2$, we get

$$\begin{aligned}
&\mathbb{P}(Y \geq (k\sigma)^2) \leq \frac{\mathbb{E}[Y]}{(k\sigma)^2} \\
\Rightarrow &\mathbb{P}((X - \mu)^2 \geq (k\sigma)^2) \leq \frac{\sigma^2}{(k\sigma)^2} \\
\Rightarrow &\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2}, \quad \text{as } \{(X - \mu)^2 \geq (k\sigma)^2\} \ = \ \{|X - \mu| \geq k\sigma\} \\
\Rightarrow &\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.
\end{aligned}$$

□

Example 1 (Test Scores)**.** Suppose a class takes a test where the average score is 75 ($\mu = 75$) and the standard deviation is 5 ($\sigma = 5$). We can use Chebyshev's inequality to find a lower bound on the percentage of scores that fall between 65 and 85.

- The interval is from 65 to 85.
- The distance from the mean is $85 - 75 = 10$.

- We want to find how many standard deviations this distance is: $k = \frac{10}{\sigma} = \frac{10}{5} = 2$.
- The probability that a score is outside this range is bounded by $\mathbb{P}(|X - 75| \geq 10) \leq \frac{1}{k^2} = \frac{1}{2^2} = \frac{1}{4} = 0.25$.
- So, at most 25% of the scores fall outside the range of 65-85.
- This means at least $1 - 0.25 = 0.75$ or 75% of the scores must be within the range.

Example 2 (Public Health Monitoring)**.** A public health agency monitors monthly influenza cases in a metropolitan area. Based on current incidences pattern, the mean number of cases is $\mu = 500$, with a standard deviation of $\sigma = 50$. An official wants to find the maximum probability of a major outbreak, defined as more than 750 cases.

- Parameters: $\mu = 500$, $\sigma = 50$.
- Event of interest: $\{X > 750\}$. This implies a deviation from the mean of $750 - 500 = 250$.
- The inequality: $\mathbb{P}(X > 750) \leq \mathbb{P}(|X - 500| \geq 250)$.
- We need $k\sigma = 250$.
$$k(50) = 250 \implies k = 5$$
- Apply the inequality:
$$\mathbb{P}(|X - 500| \geq 250) \leq \frac{1}{5^2} = \frac{1}{25} = 0.04$$

Conclusion: The maximum probability of a major outbreak (more than 750 cases) is no more than 4%. This is a reliable, conservative upper bound that holds true regardless of the distribution of influenza cases.
Remark: If the probability increases rapidly with time, then this gives a warning situation. This may suggest that current policies are insufficient.

Example 3 (Phishing Email Detection)**.** To apply Chebyshev's inequality to phishing detection, we can follow these steps:

1. Gather Baseline Data: Collect a large dataset of emails that are known to be legitimate.
2. Select a Feature: Choose a feature for analysis that is likely to differ between legitimate and phishing emails, such as the number of hyperlinks.
3. Calculate Mean and Standard Deviation: For the selected feature, calculate the mean (μ) and standard deviation (σ) from the baseline dataset.
4. Set a Threshold: For a new, incoming email, calculate how many standard deviations (k) its feature value is from the mean.

5. Evaluate Probability: Use Chebyshev's inequality to determine the maximum probability of a legitimate email having that feature value. If this probability is below a pre-defined threshold, flag the email as suspicious.

Link Count Analysis

Let's consider a practical example for an email system.

Establishing the Baseline

After analyzing millions of legitimate emails, our system has determined the following statistics for the number of links per email:

- Mean (μ): 2.0 links
- Standard Deviation (σ): 1.0 link

Analyzing a New Email

A new email arrives with an observed value (x) of 7 links. We need to determine if this email is an anomaly.

Step 1: Calculate k

We calculate how many standard deviations k the observed value is from the mean:

$$k = \frac{|x-\mu|}{\sigma} = \frac{|7-2.0|}{1.0} = \frac{5.0}{1.0} = 5$$

The new email's link count is 5 standard deviations away from the mean of legitimate emails.

Step 2: Apply Chebyshev's Inequality

Next, we find the upper bound for the probability of a legitimate email having a link count that deviates by at least 5 standard deviations:

$$\mathbb{P}(|X-2| \geq 5(1)) \leq \frac{1}{5^2} = \frac{1}{25} = 0.04$$

This result tells us that the probability of a legitimate email having a link count of 7 or more is at most 4%.

Conclusion: The very low probability (4%) of a legitimate email having 7 or more links suggests that the incoming email is a significant statistical outlier. An automated detection system could be configured to flag such emails as suspicious and send them to a quarantine folder for further review. This method offers a robust, general-purpose approach to anomaly detection without making strong assumptions about the data's underlying distribution.

Convergence of Random Variables

We often deal with sequences of random variables. It is natural to ask whether such a sequence approaches a specific limit. Unlike standard calculus, where a sequence of numbers converges, a sequence of random variables can converge in several different ways. The two most common types are convergence in probability and convergence in distribution.

Convergence in Probability

Convergence in probability describes how a sequence of random variables behaves as the number of variables increases. Intuitively, it means that the probability of the sequence deviating from a limit becomes vanishingly small.

Definition 1 (Degenerate random variable)**.** A random variable X is said to degenerate at a if

$$\mathbb{P}(X = a) = 1.$$

Here, the distribution function of X is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ 1 & \text{if } x \geq a. \end{cases}$$

Remark 1. A constant can be viewed as a degenerate random variable.

Definition 2. A sequence of random variables $\{X_n\}$ converges in probability to a random variable X (denoted $X_n \xrightarrow{p} X$) if, for every $\epsilon > 0$, the following holds:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0$$

This means that as n gets larger, the probability mass of X_n becomes more and more concentrated around X.

Example 4 (Estimating the Mean)**.** Let $\{X_1, X_2, \dots\}$ be a sequence of independent and identically distributed (IID.) random variables with mean μ and variance σ^2. Let us denote the sample mean by

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

Then,

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

We can use Chebyshev's inequality to prove this:

- We know
$$\mathbb{E}[\bar{X}_n] = \mu, \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$
- By Chebyshev's inequality, for any $\epsilon > 0$:
$$\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2/n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

- As $n \to \infty$, we can see that $\frac{\sigma^2}{n\epsilon^2} \to 0$.
- Thus, $\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) = 0$, which is the definition of convergence in probability. Therefore,
$$\bar{X}_n \xrightarrow{p} \mu.$$

Convergence in Distribution

Convergence in distribution is a weaker form of convergence than convergence in probability. It only requires that the cumulative distribution function (CDF) of the sequence of random variables converges to the CDF of the limiting random variable.
(For any random variable Y, $F_Y(x) = \mathbb{P}(Y \leq x)$.)

Definition 3. A sequence of random variables $\{X_n\}$ converges in distribution to a random variable X (denoted $X_n \xrightarrow{d} X$) if, for every point x at which the CDF of X, $F_X(x)$, is continuous,

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

The random variables themselves do not need to be related, only their distributions need to converge.

Example 5 (Maximum of Uniforms)**.** Let $\{X_1, X_2, \dots\}$ be a sequence of IID $Uniform(0,1)$ random variables. Let $X_{(n)} = \max\{X_1, \dots, X_n\}$. We want to examine the convergence in distribution of $X_{(n)}$.

- The CDF of X_i is $F_X(x) = x$ for $x \in (0,1)$.
- For $x \in (0,1)$, the CDF of $X_{(n)}$ is
$$F_{X_{(n)}}(x) = \mathbb{P}(X_{(n)} \leq x) = \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) = x^n.$$
- As $n \to \infty$, the limit of x^n for $x \in (0,1)$ is:
$$\lim_{n\to\infty} F_{X_{(n)}}(x) = \lim_{n\to\infty} x^n = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$
- This is the CDF of a random variable that is equal to 1 with probability 1 (a degenerate distribution). So, $X_{(n)} \xrightarrow{d} 1$.
- In this case, we also have $X_{(n)} \xrightarrow{p} 1$ because the limit is a constant.

Relationships

Theorem 3. *If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$.*

Proof. We denote CDF of a random variable Y by $F_Y(a) = \mathbb{P}(Y \leq a)$. Observe that for any $\epsilon > 0$, using the total probability law, we have

$$\mathbb{P}(X_n \leq x) = \mathbb{P}(X_n \leq x, |X_n - X| \leq \epsilon) + \mathbb{P}(X_n \leq x, |X_n - X| > \epsilon).$$

Since $\{X_n \leq x, |X_n - X| > \epsilon\} \subseteq \{|X_n - X| > \epsilon\}$, so the second term

$$\mathbb{P}(X_n \leq x, |X_n - X| > \epsilon) \leq \mathbb{P}(|X_n - X| > \epsilon)$$

As $n \to \infty$, the assumption $X_n \xrightarrow{P} X$ implies $\mathbb{P}(|X_n - X| > \epsilon) \to 0$, consequently

$$\mathbb{P}(X_n \leq x, |X_n - X| > \epsilon) \leq \mathbb{P}(|X_n - X| > \epsilon) \to 0.$$

Next, the first term

$$\mathbb{P}(X_n \leq x, |X_n - X| \leq \epsilon) \leq \mathbb{P}(X \leq x + \epsilon),$$

since if $X_n \leq x$ and $|X_n - X| \leq \epsilon$, then $X \leq x + \epsilon$.

Using similar arguments, we obtain

$$\mathbb{P}(X_n \leq x) \geq \mathbb{P}(X \leq x - \epsilon) - \mathbb{P}(|X_n - X| > \epsilon).$$

Thus,

$$\mathbb{P}(X \leq x - \epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \leq F_{X_n}(x) \leq \mathbb{P}(X \leq x + \epsilon) + \mathbb{P}(|X_n - X| > \epsilon).$$

As $n \to \infty$, the terms with $\mathbb{P}(|X_n - X| > \epsilon) \to 0$, so

$$F_X(x - \epsilon) \leq \liminf_{n\to\infty} F_{X_n}(x) \leq \limsup_{n\to\infty} F_{X_n}(x) \leq F_X(x + \epsilon).$$

Since x is a continuity point, let $\epsilon \to 0^+$, then $F_X(x - \epsilon) \to F_X(x)$ and $F_X(x + \epsilon) \to F_X(x)$, so by squeeze theorem,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \implies \quad X_n \xrightarrow{d} X.$$

□

Theorem 4. *Convergence in distribution does not imply convergence in probability in general.*

Proof. We prove this using a counterexample. Let $X \sim \textit{Bernoulli}(1/2)$. Suppose $Y = 1 - X$, and $X_n = X$ for $n = 1, 2, 3, \ldots$. Then,

- $X_n \overset{d}{=} X$ as $X_n = X$ for $n = 1, 2, 3, \ldots$.
- $Y \overset{d}{=} X$ as $\mathbb{P}(Y = 1) = \mathbb{P}(X = 0) = 1/2$ and $\mathbb{P}(Y = 0) = \mathbb{P}(X = 1) = 1/2$.
- Observe that, for any $x \in \mathbb{R}$, and for all $n = 1, 2, 3, \ldots$,

$$\begin{aligned} F_{X_n}(x) &= \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x), \quad \text{since} \quad X_n = X \\ &= \mathbb{P}(Y \leq x), \quad \text{since} \quad Y \overset{d}{=} X. \end{aligned}$$

Therefore, for any $x \in \mathbb{R}$,

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

$$\implies X_n \xrightarrow{d} Y.$$

Next, we want to show X_n does not converge in probability to Y. Observe that

$$|X_n - Y| = |X - (1 - X)| = |X - 1 + X| = |2X - 1|$$

so,

- if $X = 0$, $|X_n - Y| = |2X - 1| = |2 \times 0 - 1| = 1$,
- if $X = 1$, $|X_n - Y| = |2X - 1| = |2 \times 1 - 1| = 1$.

Consequently,

$$\mathbb{P}(|X_n - Y| > 0.5) = \mathbb{P}(X \in \{0, 1\}) = 1.$$

Therefore, for $\epsilon = 0.5$, as $n \to \infty$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - Y| > 0.5) = 1 \not\to 0.$$

Thus, by definition X_n does not converge in probability to Y. □

Theorem 5. *If $X_n \xrightarrow{d} c$, where c is a constant, then $X_n \xrightarrow{p} c$.*

Proof. To show $X_n \xrightarrow{p} c$, we need to prove that for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - c| > \epsilon) = 0.$$

The probability can be expressed as

$$\mathbb{P}(|X_n - c| > \epsilon) = \mathbb{P}(X_n > c + \epsilon) + \mathbb{P}(X_n \leq c - \epsilon) = [1 - F_{X_n}(c + \epsilon)] + F_{X_n}(c - \epsilon).$$

Since $X_n \xrightarrow{d} c$, $\lim_{n\to\infty} F_{X_n}(y) = F_c(y)$ for all continuity points y of CDF of degenerate random at c, say F_c.

Observe that the continuity points of F_c are all $\{y \in \mathbb{R} : \ y \neq c\}$. For $y = c - \epsilon < c$ (since $\epsilon > 0$),

$$\lim_{n\to\infty} F_{X_n}(c - \epsilon) = F_c(c - \epsilon) = 0.$$

For $y = c + \epsilon > c$, $\lim_{n\to\infty} F_{X_n}(c + \epsilon) = F_c(c + \epsilon) = 1$, consequently, $\lim_{n\to\infty}[1 - F_{X_n}(c + \epsilon)] = 0$. Therefore, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - c| > \epsilon) = 0 + 0 = 0 \quad \implies X_n \xrightarrow{p} c.$$

□

In summary,

- Convergence in Probability $\implies$ Convergence in Distribution.
- Convergence in Distribution $\not\Longrightarrow$ Convergence in Probability.
- If $X_n \xrightarrow{d} c$ (a constant), then $X_n \xrightarrow{p} c$.

Weak Law of Large Numbers (WLLN)

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. The **Weak Law of Large Numbers (WLLN)** is a fundamental result that tells us what happens to such sums or averages when the number of observations becomes large. Precisely, the **WLLN** states that averages converge to the theoretical mean; i.e., empirical averages are reliable estimates of the population mean when the sample size is large, under very general assumptions.

Theorem 6 (IID case)**.** *Let* $X_1, X_2, \ldots$ *be IID random variables with finite mean* $\mu = \mathbb{E}[X_i]$ *and finite variance* $\sigma^2 = \mathrm{Var}(X_i)$*. Define the sample average*

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

Then

$$\overline{X}_n \xrightarrow{p} \mu,$$

i.e. for every $\varepsilon > 0$*,*

$$\lim_{n\to\infty} \mathbb{P}\big(|\overline{X}_n - \mu| > \varepsilon\big) = 0.$$

Proof. Observe that by linearity,

$$\mathbb{E}[\overline{X}_n] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \mu.$$

Because the X_i are independent,

$$\mathrm{Var}(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Now apply Chebyshev's inequality, we have

$$\mathbb{P}\big(|\overline{X}_n - \mu| > \varepsilon\big) \le \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

The right-hand side tends to 0 as $n \to \infty$, proving convergence in probability. □

Remark 2. This is called the "weak" law because it asserts convergence in probability. There is also the Strong Law of Large Numbers (SLLN), which states that $\overline{X}_n \to \mu$ almost surely under the same finite expectation condition (Kolmogorov's SLLN requires $\mathbb{E}|X_1| < \infty$). The SLLN is stronger, almost sure convergence implies convergence in probability.

Example 6. Let X_i be independent Bernoulli(p). Then $\mu = p$, $\sigma^2 = p(1-p)$. The sample average $\bar{X}_n$ is the proportion of heads. WLLN implies

$$\bar{X}_n \xrightarrow{p} p.$$

Concretely, for $\varepsilon = 0.05$, applying Chebyshev's inequality, we have

$$\mathbb{P}(|\bar{X}_n - p| > 0.05) \le \frac{p(1-p)}{n(0.05)^2},$$

so to make the bound < 0.01 we need $n \gtrsim \frac{p(1-p)}{0.01 \cdot 0.05^2}$ (plugging worst-case $p = 0.5$ gives $n \approx 2000$ from the bound; CLT gives a much tighter estimate).

Example 7 (Approximating integrals)**.** The Weak Law of Large Numbers provides a foundation for estimating integrals through simulation. For a definite integral

$$I = \int_a^b g(x)\,dx,$$

we can express it as an expected value. If we let $X \sim$ Uniform(a, b), its PDF is

$$f(x) = \frac{1}{b-a} \quad for \quad x \in (a, b).$$

The expected value of $g(X)$ is given by:

$$E[g(X)] = \int_a^b g(x)f(x)\,dx = \int_a^b g(x)\frac{1}{b-a}\,dx$$

Rearranging this equation gives us the integral we wish to approximate:

$$\int_a^b g(x)\,dx = (b-a)E[g(X)]$$

The WLLN states that if we draw a large number of IID samples $X_1, X_2, \ldots, X_n$ from Uniform(a, b), the sample average of $g(X_i)$ will converge in probability to $E[g(X)]$. That is,

$$E[g(X)] \approx \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

Combining these, we get the Monte Carlo approximation for the integral,

$$\int_a^b g(x)\,dx \approx (b-a)\frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

Approximating $\int_0^1 x^2\,dx$

We will approximate the value of the integral $I = \int_0^1 x^2\,dx$. The exact value is $\frac{1}{3}$.

- The function is $g(x) = x^2$.
- The interval is $(a, b) = (0, 1)$, so $(b - a) = 1$.

We draw a small sample of $n = 5$ random numbers from a uniform distribution on $(0, 1)$,

$$\begin{aligned} X_1 &= 0.12 \\ X_2 &= 0.84 \\ X_3 &= 0.53 \\ X_4 &= 0.29 \\ X_5 &= 0.91 \end{aligned}$$

Next, we evaluate the function $g(x) = x^2$ for each of these random numbers,

$$\begin{aligned} g(X_1) &= (0.12)^2 = 0.0144 \\ g(X_2) &= (0.84)^2 = 0.7056 \\ g(X_3) &= (0.53)^2 = 0.2809 \\ g(X_4) &= (0.29)^2 = 0.0841 \\ g(X_5) &= (0.91)^2 = 0.8281 \end{aligned}$$

We then compute the average of these function values,

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) = \frac{1}{5}(0.0144 + 0.7056 + 0.2809 + 0.0841 + 0.8281) = \frac{1.9131}{5} \approx 0.3826$$

Since $(b - a) = 1$, the approximation of the integral is simply the average of the function values,

$$I \approx 0.3826$$

This result is reasonably close to the true value of $1/3 \approx 0.3333$, especially considering the small sample size.

Remark: The accuracy of the approximation improves as the number of random samples, n, increases.

References

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

Disclaimer

This lecture note is prepared solely for teaching and academic purposes. Some parts of the material, including definitions, examples, and explanations, have been adapted or reproduced from the references. These notes are not intended for commercial distribution or publication, and all rights remain with the respective copyright holders.