Chapter 5: Large Random Samples

Mariona Segú

Assistant Professor
Thema, CY Cergy Paris Université

Fall 2023

Contents

Contents

1. Introduction: What Statistics Does

- **Statistics** aims to draw conclusions about a population using information from a sample.
- An **experiment** is a process that generates data —that is, a sample.
- A **sample** is a set of observed values of one or more random variables, obtained from one or several repetitions of the experiment.
- To make valid **inferences** about the population, we need to know how likely it is to observe a given sample.
- This, in turn, requires understanding the **probability distributions of the random variables** that generated the data.

1. Introduction: Probability and Approximations

- In practice, knowing the exact probability distribution is often difficult:
  - The calculations may be complex.
  - The number of variables involved may be large.
- When we have **large random samples**, we can use powerful **approximation results** to simplify inference:
  - Law of Large Numbers
  - Central Limit Theorem
- These approximations make statistical inference feasible, even when exact probability functions are unknown.

1. Introduction

Proportion of Heads

If you flip a fair coin, you know that the probability of heads is $\frac{1}{2}$. However, in a small number of flips, the observed proportion of heads will rarely be exactly $\frac{1}{2}$. For instance, if you flip the coin 10 times, it is unlikely to get exactly 5 heads; and if you flip it 100 times, it is even less likely to get exactly 50. We can compute

these probabilities using the binomial distribution with parameters n and $p = \frac{1}{2}$. If X denotes the number of heads in 10 independent flips, then:

$$P(X = 5) = \binom{10}{5} \left(\frac{1}{2}\right)^5 \left(1 - \frac{1}{2}\right)^5 = 0.2461.$$

1. Introduction

Proportion of Heads

If Y is the number of heads in 100 independent flips, we have

$$Pr(Y = 50) = \binom{100}{50} \left(\frac{1}{2}\right)^{50} \left(1 - \frac{1}{2}\right)^{50} = 0.0796.$$

Even though the probability of exactly $\frac{n}{2}$ heads in n flips is quite small, especially for large n, you still expect the proportion of heads to be close to $\frac{1}{2}$ if n is large.

1. Introduction

Proportion of Heads

For example, if $n = 10$, the proportion of heads is $\frac{X}{10}$. In this case, the probability that the proportion is within 0.1 of $\frac{1}{2}$ is

$$P\left(0.4 \le \frac{X}{10} \le 0.6\right) = P(4 \le X \le 6)$$

$$= \sum_{i=4}^{6} \binom{10}{i} \left(\frac{1}{2}\right)^{i} \left(1 - \frac{1}{2}\right)^{10-i}$$

1. Introduction

Proportion of Heads

For example, if $n = 10$, the proportion of heads is $\frac{X}{10}$. In this case, the probability that the proportion is within 0.1 of $\frac{1}{2}$ is

$$P\left(0.4 \leq \frac{X}{10} \leq 0.6\right) = P(4 \leq X \leq 6)$$

$$= \sum_{i=4}^{6} \binom{10}{i} \left(\frac{1}{2}\right)^{i} \left(1 - \frac{1}{2}\right)^{10-i}$$

$$= B(10, 0.5, x = 6) - B(10, 0.5, x = 3) = 0.828 - 0.172 = 0.6563$$

For $n = 100$, the **probability that the sample proportion is within 0.1 of** $\frac{1}{2}$ increases to approximately **0.965**.

1. Introduction

Queuing Time

Consider a queue serving customers, where the waiting time of the i-th customer is a random variable X_i. Suppose $X_1, X_2, \ldots$ are i.i.d. random variables following a uniform distribution on the interval $[0, 1]$. The mean waiting time is therefore 0.5.

Intuitively, the average waiting time over many customers should be close to this mean. However, for any finite sample size $n > 1$, the exact distribution of the sample average $\overline{X}$ is quite complicated. It is generally difficult to compute precisely the probability that the sample average is close to 0.5 when n is large.

So, even when each X_i has a simple uniform distribution, finding the exact distribution of $\bar{X}_n$ quickly becomes computationally cumbersome.

1. Introduction

In these cases, when sample is large we will be able to use:

- The **law of large numbers**: to show that the average of a large sample of i.i.d. random variables should be close to their mean.
- The **central limit theorem:** to approximate the probability distribution function of large random samples.

A Random Variable vs. An Observation

Definition

A **random variable (RV)** is a rule that assigns a numerical value to each possible outcome of a random experiment. It is a *mathematical object*.

- It represents uncertainty: it represents the outcome of a random process *before* we actually observe it.
- It has a **probability distribution** describing how likely each outcome is.

Definition

An **observation** (or **realization**) is the actual value taken by a random variable once the experiment has been performed.

- It is the **observed outcome** of a random variable.
- It is a fixed number, not random.

Contents

2. The Law of Large Numbers

- The average of a random sample of i.i.d. random variables is called the **sample mean**.
- The sample mean provides a summary of the sample, just as the **expected value** summarizes the information contained in a probability distribution of a random variable.
- In this section, we explore the connection between the sample mean and the expected value of the underlying random variables that make up the sample.

2. The Law of Large Numbers

- The **Law of Large Numbers (LLN)** states that:

  "*If an experiment is repeated independently many times and the results are averaged, the average will be close to the expected value.*"

- There are two main versions of the LLN:
  - The *Weak Law of Large Numbers (WLLN)*
  - The *Strong Law of Large Numbers (SLLN)*
- The distinction between the two is mostly theoretical —both express the idea that sample averages converge to expected values as the sample size grows.
- In what follows, we focus on the **Weak Law of Large Numbers**.
- But before that, let us define the *sample mean*.

2. The Law of Large Numbers

Definition. For i.i.d. random variables $X_1, X_2, \ldots, X_n$, the **sample mean**, denoted by $\overline{X}$, is defined as

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Since each X_i is a random variable, the sample mean $\overline{X}$ is also a random variable. In particular, its expected value is:

$$\begin{aligned} E[\overline{X}] &= \frac{E[X_1] + E[X_2] + \cdots + E[X_n]}{n} \\ &= \frac{nE[X]}{n} \qquad \text{(since } E[X_i] = E[X] \text{ for all } i\text{)} \\ &= E[X]. \end{aligned}$$

Note: All random variables $X_1, \ldots, X_n$ are drawn from the same probability distribution and share the same expected value.

2. The Law of Large Numbers

Let us now look at the **variance** of the sample mean $\overline{X}$.

Recall that for any random variable X,

$$\mathrm{Var}(X) = E\left[(X - E[X])^2\right].$$

Before we compute the variance of $\overline{X}$, let us first review how variance behaves under scaling. Specifically, we will find the variance of aX, where a is a constant.

2. The Law of Large Numbers

To find how variance behaves when a random variable is multiplied by a constant a:

$$
\begin{aligned}
\mathrm{Var}(aX) &= E\left[(aX - E[aX])^2\right] \\
&= E\left[(aX - aE[X])^2\right] \\
&= E\left[a^2(X - E[X])^2\right] \\
&= a^2 E\left[(X - E[X])^2\right] \\
&= a^2 \mathrm{Var}(X).
\end{aligned}
$$

Conclusion: Multiplying a random variable by a constant a scales its variance by a^2.

2. The Law of Large Numbers

The variance of the sample mean $\overline{X}$ is given by:

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

$$\mathrm{Var}(\overline{X}) = \frac{\mathrm{Var}(X_1 + X_2 + \cdots + X_n)}{n^2} \qquad \text{(since } \mathrm{Var}(aX) = a^2\mathrm{Var}(X$$

$$= \frac{\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)}{n^2}$$

(independent X_i: covariances are zero)

$$= \frac{n\,\mathrm{Var}(X)}{n^2} \qquad \text{(since all } X_i \text{ have the same}$$

$$= \frac{\mathrm{Var}(X)}{n}.$$

Note: The variance of the sample mean decreases with the sample size —as n grows, $\overline{X}$ becomes less variable and concentrates around $E[X]$.

2. The Law of Large Numbers

Let us now state and prove the **Weak Law of Large Numbers (WLLN)**.

Theorem (WLLN). Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with finite expected value $E[X_i] = \mu < \infty$. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} P(|\overline{X} - \mu| \geq \epsilon) = 0.$$

Interpretation: As the sample size n increases, the probability that the sample mean $\overline{X}$ deviates from the population mean μ by more than ϵ approaches zero.

In other words, the sample mean converges in probability to the expected value.

2. The Law of Large Numbers

Proof of the Weak Law of Large Numbers

To prove the WLLN, we will use two key results:

- *Markov's Inequality*
- *Tchebysheff's Inequality*

Markov's Inequality. Let X be a random variable that takes only nonnegative values. Then, for any $a > 0$,

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

Interpretation: Markov's inequality bounds the probability that a nonnegative random variable exceeds a multiple of its mean. For example:

$$P(X \geq 2E[X]) \leq \frac{1}{2}, \quad P(X \geq kE[X]) \leq \frac{1}{k}.$$

2. The Law of Large Numbers

Markov Example

Suppose the average grade on an upcoming Probability exam is $E[X] = 12$. What is the maximum possible proportion of students who can score at least 15?

$$P(X \geq 15) \leq \frac{E[X]}{15} = \frac{12}{15} = \frac{4}{5}.$$

So, at most 80% of students could possibly score this high. However, to achieve this average, we would need a very extreme distribution:

$$\frac{4}{5} \text{ of the class scoring exactly 15, and } \frac{1}{5} \text{ scoring 0.}$$

This illustrates how Markov's bound can be very loose in realistic cases.

2. The Law of Large Numbers

Markov Example 2

Consider a random variable X that takes the value 0 with probability $\frac{24}{25}$ and the value 5 with probability $\frac{1}{25}$.

$$E[X] = \frac{24}{25} \cdot 0 + \frac{1}{25} \cdot 5 = \frac{1}{5}.$$

Let's use Markov's inequality to find an upper bound on the probability that $X \geq 5$:

$$P(X \geq 5) \leq \frac{E[X]}{5} = \frac{1/5}{5} = \frac{1}{25}.$$

But this is exactly the true probability that $X = 5$! *In this case, Markov's inequality is exact —we say it is* ***tight****.*

2. The Law of Large Numbers

From Markov to Tchebysheff: Let $Y = (X - E[X])^2$. Then $E[Y] = \mathrm{Var}(X)$. By Markov's inequality, we have:

$$P(Y \geq a^2) \leq \frac{E[Y]}{a^2} = \frac{\mathrm{Var}(X)}{a^2}.$$

Notice that the event $Y = (X - E[X])^2 \geq a^2$ is equivalent to $|X - E[X]| \geq a$. Hence, we obtain:

$$P(|X - E[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

This result is known as 'Tchebysheff's inequality'.

2. The Law of Large Numbers

Tchebysheff's inequality:

$$P(|X - E[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

Tchebysheff's inequality provides a bound on the probability that a random variable deviates from its expected value. If we set $a = k\sigma$, where σ is the standard deviation, then:

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

Example: At least 75% of the probability mass lies within 2σ of the mean, and at least 89% within 3σ. 'Ch2'

2. The Law of Large Numbers

Proof of the Weak Law of Large Numbers (WLLN). Apply Tchebysheff's inequality to the sample mean $\overline{X}$, using $a = \epsilon$:

$$P(|\overline{X} - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}(\overline{X})}{\epsilon^2}.$$

Since $\mathrm{Var}(\overline{X}) = \frac{\mathrm{Var}(X)}{n}$, we obtain:

$$P(|\overline{X} - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}(X)}{n\epsilon^2},$$

which clearly goes to zero as $n \to \infty$. **Conclusion:**

$$\lim_{n\to\infty} P(|\overline{X} - \mu| \geq \epsilon) = 0.$$

In words: as the sample size increases, the sample mean becomes arbitrarily close to the population mean.

Contents

3. The Central Limit Theorem

Many phenomena observed in the real world can be described reasonably well by a normal probability distribution. In such cases, it is often assumed that the observable random variables in a random sample $X_1, X_2, \ldots, X_n$ are independent and identically distributed, each following a **normal distribution**. We have seen that the sample mean

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

has expectation $E[\overline{X}] = \mu$ and variance $\mathrm{Var}(\overline{X}) = \frac{\mathrm{Var}(X)}{n}$. Next, let us determine the **distribution** of $\overline{X}$.

3. The Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ are drawn from a normal distribution with mean μ and variance σ^2, then their sample mean

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

is also normally distributed, with

$$E[\overline{X}] = \mu, \qquad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}.$$

This property holds for *any* sample size n: when the underlying distribution is normal, the mean of the sample is itself normal.

3. The Central Limit Theorem

We can standardize the sample mean to express it in standard normal units:

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\,\frac{\overline{X} - \mu}{\sigma}.$$

The random variable Z follows a **standard normal distribution**.

Interpretation: when sampling from a normal population, we can measure how far the sample mean deviates from the population mean in units of its standard deviation.

3. The Central Limit Theorem

Bottling Machine Example

A bottling machine fills bottles with an average of μ cl of liquid. The amount dispensed per bottle follows a normal distribution with standard deviation $\sigma = 1$ cl.

A random sample of $n = 9$ filled bottles is selected from the production line. What is the probability that the sample mean will be within 0.3 cl of the true mean μ?

3. The Central Limit Theorem

Bottling Machine Example (continued)

Let $X_1, X_2, \ldots, X_9$ denote the amounts filled in each bottle. Since each $X_i \sim \mathcal{N}(\mu, \sigma^2)$ with $\sigma = 1$, we know that the sample mean

$$\overline{X} = \frac{1}{9} \sum_{i=1}^{9} X_i$$

is also normally distributed, with

$$E[\overline{X}] = \mu, \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} = \frac{1}{9}.$$

We are interested in:

$$P(|\overline{X} - \mu| \leq 0.3) = P[-0.3 \leq (\overline{X} - \mu) \leq 0.3].$$

3. The Central Limit Theorem

Bottling Machine Example (continued)

Standardizing with

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}},$$

we obtain:

$$\begin{aligned} P(|\overline{X} - \mu| \leq 0.3) &= P\left(\frac{-0.3}{1/\sqrt{9}} \leq Z \leq \frac{0.3}{1/\sqrt{9}}\right) \\ &= P(-0.9 \leq Z \leq 0.9). \end{aligned}$$

From the standard normal table:

$$P(-0.9 \leq Z \leq 0.9) = 1 - 2P(Z > 0.9) = 1 - 2(0.1841) = 0.6318.$$

Interpretation: There is about a 63% chance that the sample mean falls within 0.3 cl of the true mean.

3. The Central Limit Theorem

Bottling Machine Example

How many observations should be included in the sample if we want the sample mean $\overline{X}$ to be within 0.3 cl of the true mean μ with probability 0.95?

We want

$$P(|\overline{X} - \mu| \leq 0.3) = P[-0.3 \leq (\overline{X} - \mu) \leq 0.3] = 0.95.$$

3. The Central Limit Theorem

Bottling Machine Example (continued)

$$P(|\overline{X}-\mu|\leq 0.3) = P[-0.3 \leq (\overline{X}-\mu) \leq 0.3] = 0.95.$$

Dividing by the standard deviation of the sample mean, $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ (with $\sigma = 1$), we obtain:

$$P\left(-0.3\sqrt{n} \leq Z \leq 0.3\sqrt{n}\right) = 0.95,$$

where Z follows a standard normal distribution. We know that

$$P(-z_0 \leq Z \leq z_0) = 0.95 \quad \Rightarrow \quad z_0 = 1.96.$$

Therefore, $\quad 0.3\sqrt{n} = 1.96 \quad \Rightarrow \quad n = \left(\frac{1.96}{0.3}\right)^2 = 42.68.$

Hence, $n \approx 43$ observations are required.

3. The Central Limit Theorem

- When we sample from a normal population, the sample mean $\overline{X}$ follows a normal sampling distribution.
- But what happens if the underlying variables X_i are *not* normally distributed?
- Fortunately, even in that case, the distribution of $\overline{X}$ becomes approximately normal as the sample size n increases.
- This remarkable result is known as the **Central Limit Theorem (CLT)**.

3. The Central Limit Theorem

The Central Limit Theorem. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with $E[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Define the standardized sample mean as:

$$Z_n = \sqrt{n}\,\frac{\overline{X} - \mu}{\sigma}, \qquad \text{where} \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then, as $n \to \infty$, the distribution of Z_n converges to the standard normal distribution:

$$\lim_{n\to\infty} P(Z_n \leq x) = \Phi(x) \quad \text{for all } x \in \mathbb{R},$$

where $\Phi(x)$ is the cumulative distribution function (CDF) of the standard normal variable.

3. The Central Limit Theorem

Test Scores Example

Achievement test scores of all high school seniors in a state have a mean of $\mu = 60$ and a variance of $\sigma^2 = 64$. A random sample of $n = 100$ students from one large high school has a mean score of $\overline{X} = 58$. Is there evidence to suggest that this high school is performing below the state average? *Compute the probability that the sample mean is at most 58 when $n = 100$.*

3. The Central Limit Theorem

Test Scores Example

Let $\overline{X}$ denote the sample mean of $n = 100$ test scores from a population with $\mu = 60$ and $\sigma^2 = 64$ (so $\sigma = 8$). We want to approximate: $P(\overline{X} \leq 58)$. Since

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

is approximately standard normal, we have:

$$P(\overline{X} \leq 58) = P\left(Z \leq \frac{58 - 60}{8/\sqrt{100}}\right) = P(Z \leq -2.5).$$

From the standard normal table, $P(Z \leq -2.5) = 0.0062$.

Interpretation: The probability is very small, providing strong evidence that this high school's average score is below the state average.

The End

Tchebysheff's Theorem

- In certain scenarios, empirical rule may not provide useful approximations.
- Tchebysheff's theorem offers a lower bound for the probability of Y being within an interval $\mu \pm k\sigma$.
- **Tchebysheff's theorem**

$$P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$
$$P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- The theorem:
  - Is valid for any probability distribution.
  - Provides conservative estimates.
  - Doesn't contradict empirical rule (verify!).

Back