Chapter 5: Large Random Samples

Mariona Segú

Assistant Professor
Thema, CY Cergy Paris Université

Fall 2023

Contents

Contents

1. Introduction

- Statistics = Draws conclusions about populations using samples.
- Making inference about the population requires knowledge of the probability of occurrence of the observed sample,
- In turn, this requires knowledge of the **probability distributions of the random variables** that generated the sample.
- Knowing the probability function can be hard in some cases (complicated calculation, too many variables...)
- When we have **large random samples**, we can introduce a number of useful approximations to facilitate calculation.
- See some examples:

1. Introduction

Proportion of Heads

If you draw a fair coin, you feel confident that the head probability is $\frac{1}{2}$. However, if you flip the coin 10 times, you would not expect to see exactly 5 heads. If you flip it 100 times, you would be even less likely to see exactly 50 heads. Indeed, we can calculate the probabilities of each of these two results using binomial distribution with parameters n and $\frac{1}{2}$. So, if X is the number of heads in 10 independent flips, we know that

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

1. Introduction

Proportion of Heads

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

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

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

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

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

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

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

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

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

For $n = 100$, the proportion of heads in n tosses is 0.965.

1. Introduction

Queuing Time

A queue is serving customers, and the ith customer waits a random time X_i to be served. Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables having the uniform distribution on the interval $[0, 1]$. The mean waiting time is 0.5.

Intuition suggests that the average of a large number of waiting times should be close to the mean waiting time. But the distribution of the average of $X_1, \ldots, X_n$ is rather complicated for every $n > 1$. It may not be possible to calculate precisely the probability that the sample average is close to 0.5 for large samples.

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.

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 summarizes a random sample, like the mean of a probability distribution summarizes the information in the distribution.
- Here, we will see the link between the sample mean and the expected value of the individual random variables that comprise the random sample.

2. The Law of Large Numbers

- The law of large numbers: "If you repeat an experiment independently a large number of times and average the result, what you obtain should be close to the expected value".
- There are two main versions of the law of large numbers:
  - The *weak* Law of Large Numbers (WLLN)
  - The *strong* Laws of Large Numbers.
  - The difference between them is mostly theoretical.
- Here we will focus only on the WLLN,
- But before... 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

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

Since the X_i's are random variables, the sample mean, $\overline{X}$, is also a random variable. In particular, we have

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

Note: All RV are all drawn from the same probability distribution with the same expected value

2. The Law of Large Numbers

Let's now look at the variance of $\overline{X}$.
We know that $\text{Var}(X) = E\left[(X - E[X])^2\right]$.
Let's compute the variance of $\text{Var}(aX)$

2. The Law of Large Numbers

Let's now look at the variance of $\overline{X}$.
We know that $\text{Var}(X) = E\left[(X - E[X])^2\right]$.
Let's compute the variance of $\text{Var}(aX)$

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

2. The Law of Large Numbers

The variance of $\overline{X}$ is given by

$$\begin{aligned}
\text{Var}(\overline{X}) &= \frac{\text{Var}(X_1 + X_2 + \ldots + X_n)}{n^2} \quad (\text{since } \text{Var}(aX) = a^2\text{Var}(X)) \\
&= \frac{\text{Var}(X_1) + \text{Var}(X_2) + \ldots + \text{Var}(X_n)}{n^2} \\
&\quad (\text{since the } X_i\text{'s are independent, all covariance terms are zero}) \\
&= \frac{n\text{Var}(X)}{n^2} \quad (\text{since } \text{Var}(X_i) = \text{Var}(X)) \\
&= \frac{\text{Var}(X)}{n}
\end{aligned}$$

Note: same result as Ch4 part 8

2. The Law of Large Numbers

Now let us state and prove the weak law of large numbers (WLLN).
The weak law of large numbers (WLLN): Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with a finite expected value $E[X_i] = \mu < \infty$. Then, for any $\epsilon > 0$,

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

In words: the distance between the sample mean and the population mean tends to zero when n tends to infinity.

2. The Law of Large Numbers

Proof of the weak law of large numbers (WLLN).
We need the *Markov inequality* and the *Tchebysheff inequality*.
Markov's Inequality: Let X be a random variable that takes only nonnegative values. Then for any positive real number a,

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

Markov's inequality tells us that the probability that X is twice as large as its expected value is at most $\frac{1}{2}$, which we can see by setting $a = 2E(X)$. More generally, the probability that a random variable is at least k times its expected value is at most $\frac{1}{k}$.

2. The Law of Large Numbers

Markov Example

Suppose that the average grade on the upcoming Probability exam is exam is 12. What is the maximum 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 can possibly score this high. But in order to achieve this average, we would need $\frac{4}{5}$ of the class to score exactly 15 and the remaining $\frac{1}{5}$ to score a 0...

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 $\frac{1}{25}$. Then

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

Let's use Markov's inequality to find a bound on the probability that X is at least 5:

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

But this is exactly the probability that $X = 5$! Here, Markov's inequality is exact; we say that Markov's inequality is *tight*.

2. The Law of Large Numbers

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

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

Notice that the event $Y = (X - E(X))^2 \geq a^2$ is the same as $|X - E(X)| \geq a$, so we conclude that

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

2. The Law of Large Numbers

We have obtained **Tchebysheff's inequality**:

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

Tchebysheff's inequality gives a bound on the probability that X is far from its expected value. If we set $a = k\sigma$, where σ is the standard deviation, then the inequality takes the form

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

▸ Ch2

2. The Law of Large Numbers

Proof of the weak law of large numbers (WLLN).
Let's use Tchebysheff's inequality for the sample mean $\overline{X}$ with $a = \epsilon$

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

which goes to zero as $n \to \infty$.
Since $\text{Var}(\overline{X}) = \frac{\text{Var}(X)}{n}$

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

WLLN in words: the distance between the sample mean and the population mean tends to zero when n tends to infinity.

Contents

3. The Central Limit Theorem

Many phenomena observed in the real world can be modeled adequately by a normal probability distribution. Thus, in many applied problems, it is reasonable to assume that the observable random variables in a random sample, $X_1, X_2, \ldots, X_n$, are independent with the same normal density function.

We have established that the statistic

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

have $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{Var}(\overline{X}) = \frac{\mathrm{Var}(X)}{n}$.
Let's now conclude about its distribution.

3. The Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then,

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

is normally distributed with mean $\mu_{\overline{X}} = \mu$ and variance $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$.

3. The Central Limit Theorem

It follows that

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

has a standard normal distribution.

3. The Central Limit Theorem

Bottling Machine

A bottling machine discharges an average of μ cl per bottle. The amount of liquid dispensed by the machine is normally distributed with $\sigma = 1$cl. A sample of $n = 9$ filled bottles is randomly selected from the output of the machine on a given day. Find the probability that the sample mean will be within 0.3 cl of the true mean μ for the chosen machine setting.

3. The Central Limit Theorem

Bottling Machine

If $X_1, X_2, \ldots, X_9$ denote the observations, then we know that the X_i's are normally distributed with mean μ and variance $\sigma^2 = 1$ for $i = 1, 2, \ldots, 9$.
Therefore, $\overline{X}$ possesses a normal sampling distribution with mean $\mu_{\overline{X}} = \mu$ and variance $\sigma^2_{\overline{X}} = \sigma^2/n = 1/9$. We want to find

$$P(|\overline{X} - \mu| \leq 0.3) = P[-0.3 \leq (\overline{X} - \mu) \leq 0.3]$$
$$= P\left(\frac{-0.3}{\sigma/\sqrt{n}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.3}{\sigma/\sqrt{n}}\right).$$

3. The Central Limit Theorem

Bottling Machine

Because $Z = \frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ has a standard normal distribution, it follows that

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

Using Table 4, Appendix 3, we find

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

3. The Central Limit Theorem

Bottling Machine

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

3. The Central Limit Theorem

Bottling Machine

Now we want

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

Dividing each term by $\sigma_{\overline{X}} = \sigma/\sqrt{n}$ (recall that $\sigma = 1$), we have

$$P\left(\frac{-0.3}{\sigma/\sqrt{n}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.3}{\sigma/\sqrt{n}}\right) = P\left(-0.3\sqrt{n} \leq Z \leq 0.3\sqrt{n}\right) = 0.95.$$

We look for

$$P(-z_0 \leq Z \leq z_0) = 0.95. \quad A(z_0) = \frac{1 - 0.95}{2} = 0.025 \quad z_0 = 1.96$$

It must follow that $0.3\sqrt{n} = 1.96$ or, equivalently, $n = \left(\frac{1.96}{0.3}\right)^2 = 42.68$.
So $n = 43$,

3. The Central Limit Theorem

- If we sample from a normal population, we know that $\overline{X}$ has a normal sampling distribution.
- But what can we say about the sampling distribution of $\overline{X}$ if the variables X_i are not normally distributed?
- Fortunately, $\overline{X}$ will have a sampling distribution that is approximately normal if the sample size is large.
- The formal statement of this result is called **the central limit theorem.**

3. The Central Limit Theorem

Central Limit Theorem: Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with $E(X_i) = \mu$ and $V(X_i) = \sigma^2 < \infty$. Define

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

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then the distribution function of Z_n converges to the standard normal distribution function as $n \to \infty$. That is,

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

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

3. The Central Limit Theorem

Test scores

Achievement test scores of all high school seniors in a state have a mean of 60 and a variance of 64. A random sample of $n = 100$ students from one large high school had a mean score of 58. Is there evidence to suggest that this high school is inferior? (Calculate the probability that the sample mean is at most 58 when $n = 100$.)

3. The Central Limit Theorem

Test scores

Solution: Let $\overline{X}$ denote the mean of a random sample of $n = 100$ scores from a population with $\mu = 60$ and $\sigma^2 = 64$. We want to approximate $P(\overline{X} \leq 58)$. We know that $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ has a distribution that can be approximated by a standard normal distribution. Hence, using Table 4, Appendix 3, we have

$$P(\overline{X} \leq 58) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{58-60}{0.8}\right) \approx P(Z \leq -2.5) = 0.0062.$$

It is very low... The evidence suggests that the average score for this high school is lower than the overall average of $\mu = 60$.

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