Math refresher C: Fundamentals of mathematical statistics

Lecture 2

Tunga Kantarcı, Fall 2025

Fundamentals of mathematical statistics: Populations, parameters, and random sampling: Statistical inference

Statistical inference involves drawing conclusions about a population based on a sample extracted from it. For this we also formulate a model that incorporates unknown population parameters, which we estimate using the sample data.

Fundamentals of mathematical statistics: Populations, parameters, and random sampling: Statistical inference

The first step in statistical inference is to identify the population of interest.

The second step is to specify a model that describes the relationship of interest in that population. Models typically involve probability distributions or characteristics of these distributions, and they depend on unknown parameters. Parameters are constants that define the direction and strength of relationships among variables.

The next step is to collect data, so we draw a random sample from the population.

The final step is to estimate the parameters using the sample data.

Fundamentals of mathematical statistics: Populations, parameters, and random sampling: Statistical inference: Example

Suppose in the population of all working adults in the NL, we are interested in learning about the return to education, measured by the average percentage increase in earnings resulting from an additional year of education.

Obtaining information on earnings and education for the entire working population in the NL is impractical and costly. However, we can collect data from a subset of the population which we call a sample.

Fundamentals of mathematical statistics: Populations, parameters, and random sampling: Statistical inference: Example

The parameter of interest is the return to education in the population.

Using the collected sample data, and using some model, we may report that our best estimate of the return to another year of education is 7.5%. This is an example of a point estimate.

We may also report a range, such as "the return to education is between 5.6% and 9.4%". This is an example of an interval estimate.

Fundamentals of mathematical statistics: Populations, parameters, and random sampling: Statistical inference: Formal definition

Let's be more formal about statistical inference. Suppose we are interested in learning about a population characteristic. We represent this characteristic as a random variable, Y. Since Y is random, it follows a population distribution described by a PDF:

$$f_Y(y; \theta)$$

where

- y is a realized value, the data we observe, and
- θ is an unknown population parameter, like the mean of the PDF.

Different values of θ correspond to different population distributions.

Fundamentals of mathematical statistics: Populations, parameters, and random sampling: Statistical inference: Formal definition

In general statistical inference, we typically do not observe the form of the PDF, so we assume a form, such as the normal or exponential distribution. However, making an assumption is not required. In this course, we will hardly make a distributional assumption for

$$f_Y(y; \theta)$$

Our goal will be to learn about the parameter θ, focusing especially on the conditional mean and variance. If we are able to collect data from the population, we can use it to infer θ. That is, to estimate it.

Fundamentals of mathematical statistics: Populations, parameters, and random sampling: Statistical inference: Formal definition: Example

Suppose we are interested in estimating the average income of working adults in the NL. Let Y represent the income of a randomly selected individual from this population. We assume that Y follows a normal distribution with unknown mean μ and known standard deviation σ, so the PDF is:

$$f_Y(y;\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

μ is the unknown parameter we want to learn about. We collect a random sample of incomes from the population and use this data to estimate μ. For example, if we obtain a sample of 100 individuals and calculate the sample mean income to be €42,000, then €42,000 is our point estimate of μ.

Fundamentals of mathematical statistics: Populations, parameters, and random sampling: Random sampling

A set of random variables $\{Y_1, Y_2, \ldots, Y_n\}$, each with PDF $f(y;\theta)$, is called a random sample from the population described by $f(y;\theta)$. The random nature of $\{Y_1, \ldots, Y_n\}$ reflects the fact that many different outcomes are possible before the sampling is actually carried out.

When $\{Y_1, \ldots, Y_n\}$ is a random sample from $f(y;\theta)$, we also say that the Y_i are independent and identically distributed, or i.i.d., random variables from $f(y;\theta)$.

Random sampling ensures that the sample is representative of the population in a statistical sense, because each unit has an equal chance of being selected.

Fundamentals of mathematical statistics: Populations, parameters, and random sampling: Random Sampling: Example

If income is collected for a sample of $n = 100$ families in the NL, the observed incomes will typically differ from one sample of 100 families to another.

Once a sample is obtained, we have a set of observed values, say, $\{y_1, y_2, \ldots, y_n\}$, which constitute the data we work with.

If the sample is drawn in a random fashion, then it is representative of the population, and we are good to go to estimate the average income.

Fundamentals of mathematical statistics: Estimators and estimates

An estimator is a mathematical tool or formula used to estimate an unknown population parameter, such as θ, based on data from a sample. It provides a systematic way to compute an estimate using observed data. It is chosen before any data is collected.

Fundamentals of mathematical statistics: Estimators and estimates

An estimator W of a parameter θ can be expressed generally as a mathematical formula:

$$W = h(Y_1, Y_2, \ldots, Y_n)$$

for some known function h of the random variables $Y_1, Y_2, \ldots, Y_n$. The function h is chosen before any data is observed and remains fixed regardless of the actual sample outcome. Once the sample is collected, the estimator W is evaluated using the observed values $\{y_1, y_2, \ldots, y_n\}$, producing an estimate of the parameter θ.

Fundamentals of mathematical statistics: Estimators and estimates: Example

Let $\{Y_1, Y_2, \ldots, Y_n\}$ be a random sample from a population with mean μ. A natural estimator of μ is the average of the random sample:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

$\bar{Y}$ is called the random sample average. Unlike in Math Refresher A, where we defined the sample average of a set of numbers as a descriptive statistic, $\bar{Y}$ is now viewed as an estimator, a rule for estimating the unknown population mean μ.

Fundamentals of mathematical statistics: Estimators and estimates: Example

Given any outcome of the random variables $\{Y_1, Y_2, \ldots, Y_n\}$, we apply the same rule: we average the values. For actual data outcomes $\{y_1, y_2, \ldots, y_n\}$, the estimate is:

$$\bar{y} = \frac{1}{n}(y_1 + y_2 + \cdots + y_n)$$

This value $\bar{y}$ is called the sample mean, and it is the numerical estimate of the population mean μ based on the observed data.

Fundamentals of mathematical statistics: Estimators: Sampling distribution

It is important to realize that W is a random variable since h is a function of random sample data. So W has a distribution. The distribution of an estimator is called its sampling distribution. That is, there is a likelihood of various outcomes of W across different random samples.

This is abstract, so let's make it concrete.

Fundamentals of mathematical statistics: Estimators: Sampling distribution

Let $\{Y_1, Y_2, \ldots, Y_n\}$ be a random sample from a population that follows a normal distribution, and assume that we know its mean and standard deviation: let them be $\mu = 5$ and $\sigma = 2$. As before, we consider the random sample average as an estimator of μ:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

In a computer simulation, we can draw random samples of size $n = 30$ for a total of 1,000 samples, computing $\bar{Y}$ from each sample. This yields 1,000 sample means, each as an estimate of the population mean. Since Y_i is random, $\bar{Y}$ is random.

Fundamentals of mathematical statistics: Estimators: Sampling distribution

If we repeatedly draw samples from the population and compute the sample mean for each, we obtain a collection of estimates. These estimates vary from sample to sample because the estimator is a random variable. If we plot these estimates, we visualize the sampling distribution of the estimator. This distribution reflects the variability of the estimator across different samples.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Note on the definition of population

In theory, the population is the full distribution of the variable of interest, such as the wage distribution in the Netherlands. The sampling distribution of an estimator is defined over all possible samples of size n drawn from that population.

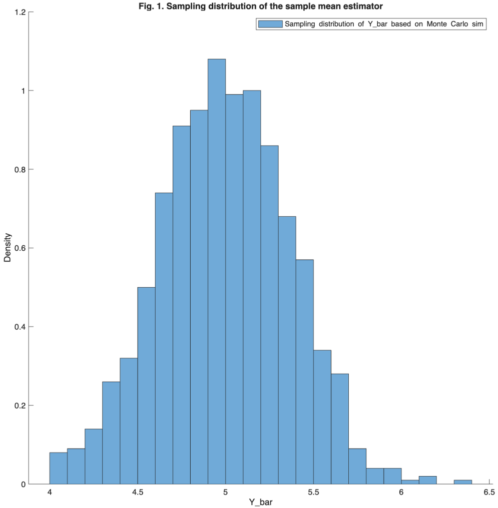
Fundamentals of mathematical statistics: Estimators: Sampling distribution: Note on the definition of population

In a simulation, however, we do not use the real population. Instead, we specify a data generating process, such as

$$Y \sim N(0, \sigma^2)$$

and the entire probability distribution implied by this process serves as the population in the simulation. Each simulated dataset of size n is then a sample from this artificial population. By repeating the sampling many times, for instance 1,000 repetitions, we obtain an empirical approximation of the theoretical sampling distribution of the estimator. As the number of repetitions increases, this empirical distribution converges more closely to the true sampling distribution, though it never literally exhausts the entire population.

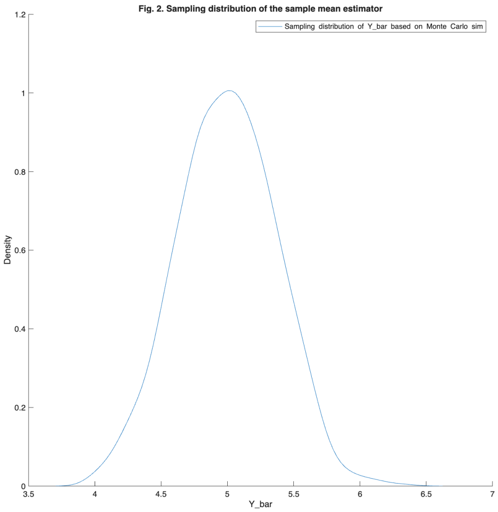
Fundamentals of mathematical statistics: Estimators: Sampling distribution



Fundamentals of mathematical statistics: Estimators: Sampling distribution

It is easier to work with the kernel smoothed version of the distribution. A kernel smoothed version of a distribution means we replace the bumpy histogram with a smooth curve that estimates the shape of the data, like drawing a gentle line through noisy dots to see the overall pattern more clearly. It is not important to understand this.

Fundamentals of mathematical statistics: Estimators: Sampling distribution



Fundamentals of mathematical statistics: Estimators: Sampling distribution

What do we use the sampling distribution for?

It is possible to have different estimators, so we need some criteria for choosing among estimators. In mathematical statistics, we study the sampling distributions of estimators to understand their behavior and evaluate their performance. We choose an estimator that behaves the best.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Unbiasedness as a finite sample property

An estimator W of θ is an unbiased estimator if

$$\mathrm{E}(W) = \theta$$

This means that the expected value of the sampling distribution of W should be equal to the population parameter it is supposed to be estimating. And what does this mean?

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Unbiasedness as a finite sample property

Imagine a thought experiment: we repeatedly draw random samples of Y from the population and compute an estimate each time using the same estimator. This generates a collection of estimates, and their distribution is called the sampling distribution of the estimator. If the estimator is unbiased, then the average of these estimates, over many samples, equals the true parameter θ:

$$\mathsf{E}(W) = \theta$$

This is what we mean by unbiasedness, on average, the estimator gets it right.

Would you want an estimator that systematically misses the target? No.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Unbiasedness as a finite sample property

This thought experiment is abstract because, in most applications, we just have one random sample to work with. So theoretical econometricians simulate repeated sampling on computers, just like we did.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Unbiasedness as a finite sample property

If W is a biased estimator of θ, its bias is defined as

$$\mathrm{E}(W) = \theta + \mathrm{Bias}(W)$$

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Unbiasedness as a finite sample property: Example

Let $\{Y_1, Y_2, \ldots, Y_n\}$ be a random sample from a population that follows a normal distribution, and assume that we know its mean and standard deviation: let them be $\mu = 5$ and $\sigma = 2$. As before, we consider the sample average as an estimator of μ:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

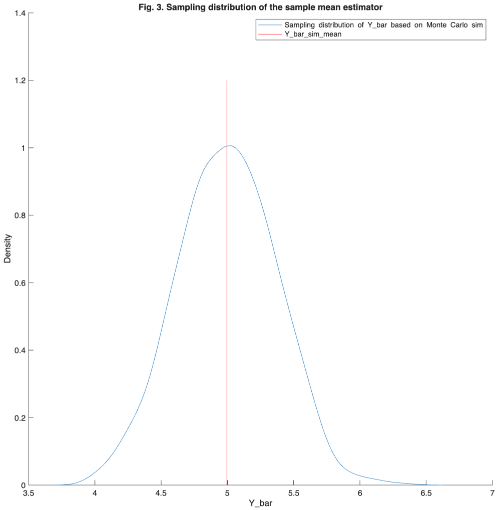
Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Unbiasedness as a finite sample property

We can demonstrate unbiasedness of $\bar{Y}$ using simulation. In a computer simulation, we can draw random samples of size $n = 30$ for a total of 1,000 samples, computing $\bar{Y}$ from each sample. This yields 1,000 sample means, each as an estimate of the population mean.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Unbiasedness as a finite sample property: Example

If we plot these estimates, we obtain the sampling distribution of the estimator. Let's plot it, now with its mean marked. Is it surprising that the average of these 1,000 sample means is very close to the true population mean $\mu = 5$? This outcome reflects the unbiasedness of $\bar{Y}$ as an estimator of μ.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Unbiasedness as a finite sample property: Example



Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Unbiasedness as a finite sample property

We can also demonstrate unbiasedness of $\bar{Y}$ mathematically.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Unbiasedness as a finite sample property: Example

Assume that each Y_i is drawn from a population with mean μ, which implies:

$$\mathrm{E}(Y_i) = \mu \quad \text{for all } i = 1, 2, \ldots, n$$

We can show that the sample average $\bar{Y}$ is an unbiased estimator of the population mean μ. Using the properties of expected values,

$$\mathrm{E}(\bar{Y}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(Y_i) = \frac{1}{n}\cdot n\mu = \mu$$

Thus, $\bar{Y}$ has expectation equal to μ, and is therefore an unbiased estimator of the population mean.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Variance of it: Efficiency as a finite sample property

Sometimes, there is more than one unbiased estimator for the same parameter. For instance, in the example, if we had only a single observation instead of n observations, we would obtain a different estimator, one that uses fewer data points yet remains unbiased. This highlights the need for an additional criterion to select among multiple unbiased estimators.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Variance of it: Efficiency as a finite sample property

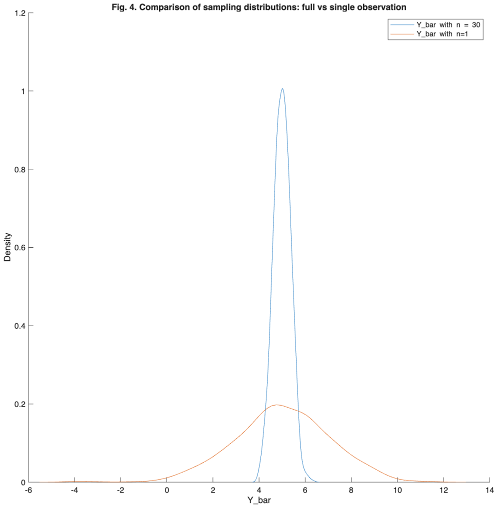
We considered the mean of an estimator's sampling distribution as a criterion for evaluating its quality.

The variance of the sampling distribution can also serve as a criterion. When faced with multiple unbiased estimators, we prefer the one with the smallest variance, as it tends to produce estimates with less uncertainty, on average.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Variance of it: Efficiency as a finite sample property: Example

Consider the same computer simulation described above. We now compare two different estimators of the population mean μ. One estimator is the sample average based on $n = 30$ observations, and the other is based on a single observation, $n = 1$. For each estimator, we generate its sampling distribution by repeatedly drawing 1,000 samples and computing the corresponding estimate. The resulting distributions are overlaid in the plot. Notice how the variance of the estimator that uses only one observation is significantly larger than that of the estimator based on 30 observations.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Variance of it: Efficiency as a finite sample property: Example



Fundamentals of mathematical statistics: Estimators: Sampling distribution: Variance of it: Efficiency as a finite sample property: Example

We can also prove the smaller variance of $\bar{Y}$ when n is large mathematically.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Variance of it: Efficiency as a finite sample property: Example

$$\mathrm{Var}(\bar{Y}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(Y_i) = \frac{1}{n^2}\cdot n\cdot\sigma^2 = \frac{\sigma^2}{n}$$

If we use only a single observation, we obtain a different estimator with a larger variance. This provides a formal criterion for choosing among estimators: In this case, we prefer the estimator that uses n observations, as it yields a smaller variance and therefore less uncertainty in the estimate.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Variance of it: Efficiency as a finite sample property: Example

An estimator that has a smaller variance is said to be more efficient.

Fundamentals of mathematical statistics: Estimator: Sampling distribution: Consistency as a large sample property

Any estimator should produce an estimate that approaches the true population parameter as the sample size increases. This makes intuitive sense: as the sample becomes larger, so more representative of the population, the estimate should naturally converge toward the actual population parameter. That is, the sampling distribution of the estimator should collapse around the parameter as the sample size gets large.

Fundamentals of mathematical statistics: Estimator: Sampling distribution: Consistency as a large sample property

Let $\{Y_1, Y_2, \ldots, Y_n\}$ be a random sample from a population, and let W_n be an estimator of μ based on this sample. Then, W_n is a consistent estimator of μ if for every $\varepsilon > 0$,

$$P\left(|W_n - \mu| > \varepsilon\right) \to 0 \quad \text{as} \quad n \to \infty.$$

If W_n is not consistent for μ, then we say it is inconsistent.

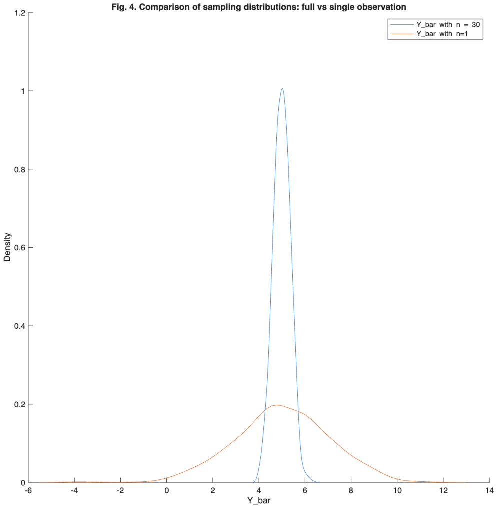
Fundamentals of mathematical statistics: Estimator: Sampling distribution: Consistency as a large sample property

This large-sample property follows from a Law of Large Numbers, which ensures that many estimators converge to the true parameter as the sample size grows.

Fundamentals of mathematical statistics: Estimator: Sampling distribution: Consistency as a large sample property: Example

Consider the sample average estimator, $\bar{Y}$, described above. In a similar computer simulation, let the sample size n increase from 1 to 30, and observe how the sampling distribution of this estimator evolves. As n grows, the distribution becomes more concentrated around the true population mean, illustrating the estimator's consistency.

Fundamentals of mathematical statistics: Estimator: Sampling distribution: Consistency as a large sample property: Example



Fundamentals of mathematical statistics: Estimator: Sampling distribution: Asymptotic normality as a large sample property

Consistency is a property of point estimators. Although it tells us that the distribution of the estimator collapses around the parameter as the sample size grows, it tells us nothing about the shape of that distribution for a given sample size. For constructing interval estimators and testing hypotheses, we need a way to approximate the distribution of our estimators. Most estimators have distributions that are well approximated by a normal distribution in large samples, which motivates the following definition.

Fundamentals of mathematical statistics: Estimator: Sampling distribution: Asymptotic normality as a large sample property

Let $\{Z_1, Z_2, \ldots, Z_n\}$ be a sequence of random variables. What does this represent?

Z_n denotes the standardized version of the estimator $\bar{Y}_n$ based on a sample of size n:

$$Z_n = \frac{\sqrt{n}\,(\bar{Y}_n - \mu)}{\sigma}$$

Each Z_n has its own sampling distribution.

Asymptotic normality describes how these sampling distributions behave as n becomes large.

Fundamentals of mathematical statistics: Estimator: Sampling distribution: Asymptotic normality as a large sample property

Here we use the sample mean $\bar{Y}_n$ as an example estimator. But the same idea applies to many estimators that can be written, or approximated, as averages.

Fundamentals of mathematical statistics: Estimator: Sampling distribution: Asymptotic normality as a large sample property

Let $\Phi(z)$ denote the standard normal CDF. Let z represent a real number. If

$$P(Z_n \leq z) \;\rightarrow\; \Phi(z)$$

as

$$n \rightarrow \infty$$

Z_n is said to have an asymptotic standard normal distribution. We often write this as

$$Z_n \;\stackrel{a}{\sim}\; N(0,1)$$

where the tilde with the superscript a denotes asymptotically distributed as.

Fundamentals of mathematical statistics: Estimator: Sampling distribution: Asymptotic normality as a large sample property

Asymptotic normality means that the

CDF of Z_n

gets closer and closer to the

CDF of the standard normal distribution

as the sample size n becomes large.

Fundamentals of mathematical statistics: Estimator: Sampling distribution: Asymptotic normality as a large sample property

This large-sample result follows from a Central Limit Theorem, which ensures that many estimators behaving like averages become approximately normal as the sample size grows.

Fundamentals of mathematical statistics: General approaches to parameter estimation

There are several methods for estimating a population parameter. In this course, we focus on the least squares method. Alternative estimation techniques are beyond the scope of this course and are therefore not discussed.

Fundamentals of mathematical statistics: General approaches to parameter estimation: The method of least squares

The least squares method is an approach to parameter estimation that seeks to minimize the sum of squared differences between observed data and model predictions. Let $Y_1, Y_2, \ldots, Y_n$ be observed data, and $f_i(\theta)$ represent the model prediction for observation i, depending on parameter θ. The least squares estimator $\hat{\theta}$ minimizes the objective function:

$$S(\theta) = \sum_{i=1}^{n} (Y_i - f_i(\theta))^2$$

We will discuss the least squares method in the context of regression analysis, along with the underlying intuition, later in this course.

Fundamentals of mathematical statistics: Hypothesis testing

Hypothesis testing will be discussed in the context of regression analysis later in this course.