Math refresher C: Fundamentals of mathematical statistics

Econometrics for minor Finance, Lecture 3

Tunga Kantarcı, Fall 2025

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

Statistical inference involves drawing conclusions about a population based on a sample taken from that population.

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

In the population of all working adults in the NL, we, as labor economists, 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 would be 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

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 within the population.

The next step is to collect data by drawing a random sample from the 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 final step is to estimate the parameters using the sample data.

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

In the labor economics example, the parameter of interest is the return to education in the population. Using the collected sample data, 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. Alternatively, we may 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

Let's be more formal about this statistical inference. Suppose we are interested in learning about a population characteristic. We represent this characteristic as a random variable Y, which follows a population distribution described by a PDF, $f_Y(y; \theta)$, where θ is an unknown parameter. The form of the PDF is typically assumed to be known, except for the value of θ. Different values of θ correspond to different population distributions, so our goal is to learn about the true value of θ. If we can obtain a random sample from the population, we can use it to infer information about θ.

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

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

Here, 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. We might also want to estimate an interval, such as "the average income is between €40,500 and €43,500," which would be an interval estimate of μ.

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

If $Y_1, Y_2, \ldots, Y_n$ are independent random variables with a common PDF, $f(y;\theta)$, then $Y_1, Y_2, \ldots, Y_n$ are said to form a random sample from the population represented by $f(y;\theta)$.

The random nature of $Y_1, Y_2, \ldots, Y_n$ 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 the density $f(y;\theta)$, 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 is chosen before any data is collected. It provides a systematic way to compute an estimate using observed data.

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 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 $Y_1, \dots, Y_n$, we apply the same rule: we average the values. For actual data outcomes $\{y_1, y_2, \dots, y_n\}$, 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

Note 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.

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 sensible 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. This is what we discuss in the remainder of this lecture.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it

An estimator W of θ is an unbiased estimator if

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

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

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it

In a thought experiment, if we could indefinitely draw random samples on Y from the population, compute an estimate each time using the estimator, and then average these estimates, we should obtain θ, if the estimator is unbiased. Who wants an estimator that does not do this? Nobody.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it

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.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it

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

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

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Example

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

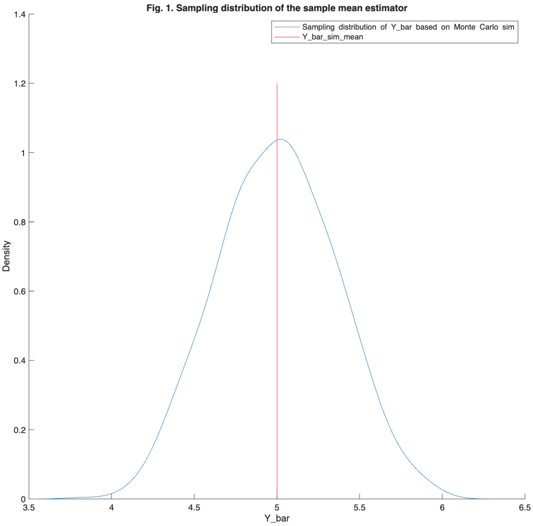
$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

In a computer simulation, we repeatedly draw random samples of size $n = 30$ for a total of 1,000 iterations, computing $\bar{Y}$ for each sample. This yields 1,000 estimates of the population mean.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Example

The resulting sampling distribution of these estimates is shown below, with its mean indicated. 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 the sample average as an estimator of μ.

Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Example



Fundamentals of mathematical statistics: Estimators: Sampling distribution: Mean of it: Example

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

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

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

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

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: Mean of it

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

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: Example

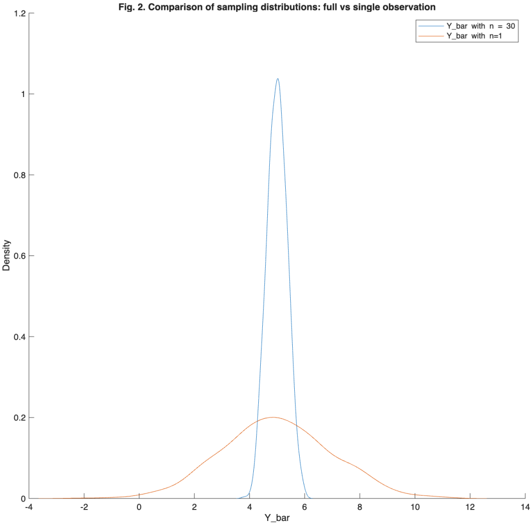
$$\mathrm{Var}(\bar{Y}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(Y_i) = \frac{1}{n^2}\cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

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: 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 below. 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: Example



Fundamentals of mathematical statistics: Estimators: Sampling distribution: Variance of it

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

Fundamentals of mathematical statistics: Estimator: Large sample properties

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 and more representative of the population, the estimate should naturally converge toward the actual 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: Large sample properties: Consistency

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 $\varepsilon > 0$,

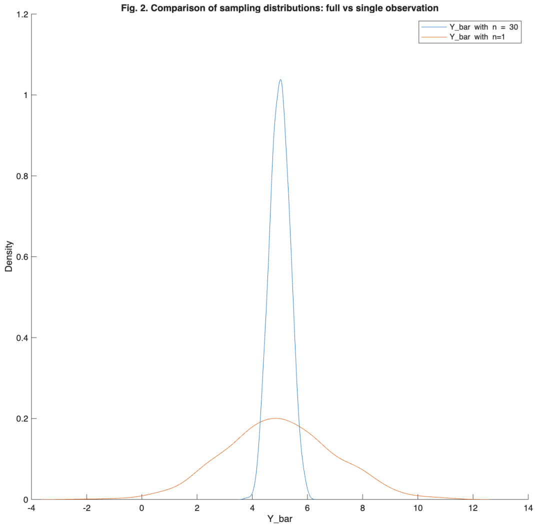
$$P\left(|W_n - \mu| > \varepsilon\right) \to 0 \quad \text{as} \quad n \to \infty.$$

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

Fundamentals of mathematical statistics: Estimator: Large sample properties: Consistency: 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: Large sample properties: Consistency: Example



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.