Statistical inference

by Marco Taboga, PhD

Statistical inference is the act of using observed data to infer unknown properties and characteristics of the probability distribution from which the data have been extracted.

Example

In the simplest possible case of statistical inference, we observe the realizations

$$x_1, \ldots, x_n$$

of some independent random variables

$$X_1, \ldots, X_n$$

all having the same distribution.

We then use the observed realizations to infer some characteristics of the distribution.

Example The lifetime of a certain type of electronic device is a random variable X, whose probability distribution is unknown. Suppose that we independently observe the lifetimes of 10 components. Denote these realizations by x_1, x_2, ..., x_{10}. We are interested in the expected value of X, which is an unknown characteristic of its distribution. We use the sample mean

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i$$

as our estimate (best guess) of the expected value. In this simple example, the sample x_1, x_2, ..., x_{10} is used to make a statistical inference about a characteristic (the expected value) of the distribution that generated the sample (the probability distribution of X).

Fundamental elements

The previous example shows that three fundamental elements are required to make a statistical inference:

- a sample (the observed data);
- a probability distribution that generates the data;
- a characteristic of the distribution about which inferences are drawn.

In the next sections we define these three fundamental elements in a mathematically meaningful way.

The sample

In the above example of statistical inference, X_1, ..., X_n are independent random variables. However, more complicated cases are possible. For instance,

1. X_1, ..., X_n are not independent;
2. X_1, ..., X_n are random vectors having a common probability distribution;
3. X_1, ..., X_n do not have a common probability distribution.

Is there a definition of sample that generalizes all of the above special cases?

Definition

The definition is extremely simple.

Definition A **sample** ξ is the realization of a random vector Ξ.

As we will see in the following examples, ξ is a vector that collects the observed data.

The vector ξ is considered a realization of a random vector Ξ.

The object of the statistical inferences is the probability distribution of Ξ.

Examples

The following examples show how this general definition accommodates the special cases mentioned above.

Note that, from now on, in order to be more precise, we will use the term distribution function instead of generically speaking of probability distribution.

Example We observe the realizations x_1, ..., x_n of some independent random variables X_1, ..., X_n having a common distribution function $F_X(x)$. The sample is the n-dimensional vector

$$\xi = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}$$

which is a realization of the random vector

$$\Xi = \begin{bmatrix} X_1 & \ldots & X_n \end{bmatrix}$$

Since the observations are independent, the joint distribution function of the vector Ξ is equal to the product of the marginal distributions of its entries:

$$F_\Xi(\xi) = F_X(x_1) \cdot \ldots \cdot F_X(x_n)$$

The distribution function of Ξ, denoted by $F_\Xi(\xi)$, is the unknown distribution function that constitutes the object of inference.

Example If we take the previous example and drop the assumption of independence, the sample ξ and the vector Ξ are still defined in the same way, but the joint distribution function $F_\Xi(\xi)$ can no longer be written as the product of the distribution functions of X_1, ..., X_n.

In the next example the single observations are no longer scalars, but they are vectors.

Example If X_1, ..., X_n are independent K-dimensional random vectors having a common joint distribution function $F_X(x)$, then the sample ξ and the vector Ξ are nK-dimensional. We can still write the joint distribution function of Ξ as

$$F_\Xi(\xi) = F_X(x_1) \cdot \ldots \cdot F_X(x_n)$$

In the following example we relax the assumption that all the observations come from a unique distribution.

Example If the K-dimensional random vectors X_1, ..., X_n have different joint distribution functions $F_{X_1}(x_1)$, ..., $F_{X_n}(x_n)$, then ξ and Ξ are defined as before, but the joint distribution function of Ξ is

$$F_\Xi(\xi) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n)$$

why do we have sample here?

Sample size

When the sample is made of the realizations x_1, ..., x_n of n random variables (or vectors), then we say that the sample size is n.

An individual realization x_i is referred to as an observation from the sample.

Statistical model

We now shift our attention to the probability distribution that generates the sample, which is another one of the fundamental elements of a statistical inference problem.

In the previous section we have defined a sample ξ as a realization of a random vector Ξ having joint distribution function $F_\Xi(\xi)$.

The sample ξ is used to infer some characteristics of $F_\Xi(\xi)$ that are not fully known by the statistician.

The properties and the characteristics of $F_\Xi(\xi)$ that are already known (or are assumed to be known) before observing the sample are called a model for Ξ.

we make assumptions about the population before sampling.

Definition

In mathematical terms, a model for Ξ is a set of joint distribution functions to which $F_\Xi(\xi)$ is assumed to belong.

Definition Let the sample ξ be a realization of an l-dimensional random vector Ξ having joint distribution function $F_\Xi(\xi)$. Let Ψ be the set of all l-dimensional joint distribution functions:

$$\Psi = \left\{ F : \mathbb{R}^l \rightarrow \mathbb{R}_+ \textit{ such that F is a joint distribution function} \right\}$$

A subset $\Phi \subset \Psi$ is called a **statistical model** (or a model specification or, simply, a model) for Ξ.

In this definition,

- the set Ψ is a large set containing all the possible data-generating distributions;
- the set Φ is a smaller subset of data-generating distributions on which we focus our attention.

The smaller set Φ is called a statistical model.

Examples

The following examples are a continuation of the examples made in the previous section.

Example Suppose that our sample is made of the realizations x_1, ..., x_n of the random variables X_1, ..., X_n. Assume that the n random variables are mutually independent and that they have a common distribution function $F_X(x)$. The sample is the n-dimensional vector

$$\xi = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix} ✓$$

Ψ is the set of all possible distribution functions of the random vector $\Xi = \begin{bmatrix} X_1 & \ldots & X_n \end{bmatrix}$. Recalling the definition of marginal distribution function and the characterization of mutual independence, we can define the statistical model Φ as follows:

$$\Phi = \left\{ F \in \Psi : \textit{ all the marginals of F are equal and F is equal to the product of its marginals} \right\} ✓$$

(independence) we just multiply the marginals

Example Take the example above and drop the assumption that the random variables are mutually independent. The statistical model Φ is now:

$$\Phi = \left\{ F \in \Psi : \textit{ all the marginals of F are equal} \right\} ✓$$

Mis-specified models

If $F_\Xi \in \Phi$, the model is said to be correctly specified (or well-specified).

Otherwise, if $F_\Xi \notin \Phi$ the model is said to be mis-specified. ✓

Parametric model

but how do the parameters here saying anything about pop? (like mean?) should it come from the cdf instead of γ (the correspondence)

A model Φ for Ξ is called a parametric model if the joint distribution functions belonging to Φ are put into correspondence with a set Θ of real vectors.

Definition Let Φ be a model for Ξ. Let $\Theta \subseteq \mathbb{R}^p$ be a set of p-dimensional real vectors. Let $\gamma(\theta)$ be a correspondence that associates a subset of Φ to each $\theta \in \Theta$. The triple (Φ, Θ, γ) is a **parametric model** if and only if

$$\Phi = \bigcup_{\theta \in \Theta} \gamma(\theta)$$

model

real vector

Θ → parameter space

correspondence b/w subset of model to real vectors

The set Θ is called parameter space. A vector $\theta \in \Theta$ is called a parameter.

Therefore, in a parametric model every element of Φ is put into correspondence with at least one parameter θ.

Parametric families

When $\gamma(\theta)$ associates to each parameter a unique joint distribution function (i.e., when $\gamma(\theta)$ is a function) the parametric model is called a parametric family.

Definition Let (Φ, Θ, γ) be a parametric model. If γ is a function from Θ to Φ, then the parametric model is called a **parametric family**. In this case, the joint distribution function associated to a parameter θ is denoted by $F_\Xi(\xi; \theta)$.

Here is a classical example of a parametric family.

This is how the correspondence f'n works.

Example Suppose that Ξ is assumed to have a multivariate normal distribution. Then, the model Φ is the set of all multivariate normal distributions, which are completely described by two parameters (the mean vector μ and the covariance matrix Σ). Each parameter $\theta = (\mu, \Sigma)$ is associated to a unique distribution function in the set Φ. Therefore, we have a parametric family.

When each distribution function is associated with only one parameter, the parametric family is said to be identifiable.

Definition Let (Φ, Θ, γ) be a parametric family. If γ is one-to-one (i.e., each distribution function F is associated with only one parameter), then the parametric family is said to be **identifiable**.

The set of multivariate normal distributions in the previous example is also an identifiable parametric family because each distribution is associated to a unique parameter.

Inferences about the data-generating distribution

A **statistical inference** is a statement about the unknown distribution function $F_\Xi(\xi)$, based on the observed sample ξ and the statistical model Φ.

→ Our assumed set of distributions for the population.

Types of statistical inference

The following are common kinds of statistical inferences.

1. Hypothesis testing: we make an hypothesis about some feature of the distribution $F_\Xi(\xi)$ and we use the data to decide whether to reject or do not reject the hypothesis; ✓
2. Point estimation: we use the data to estimate the value of a parameter of the data-generating distribution $F_\Xi(\xi)$;
3. Bayesian inference: we use the observed sample ξ to update prior probabilities assigned to the possible data-generating distributions.

Interesting!

Model restrictions

Often, we make statistical inferences about model restrictions.

Given a subset of the original model $\Phi_R \subset \Phi$, a model restriction can be either an inclusion restriction:

$$F_\Xi \in \Phi_R$$

or an exclusion restriction:

$$F_\Xi \notin \Phi_R$$

Decision theory

The choice of the statement (the statistical inference) to make based on the observed data can often be formalized as a decision problem where:

1. making a statistical inference is regarded as an action;
2. each action can have different consequences, depending on which distribution function $F_\Xi(\xi)$ is the true one;
3. a preference ordering over possible consequences needs to be elicited;
4. an optimal course of action needs to be taken, coherently with elicited preferences.

There are several different ways of formalizing such a decision problem. The branch of statistics that analyzes these decision problems is called statistical decision theory.

Keep reading

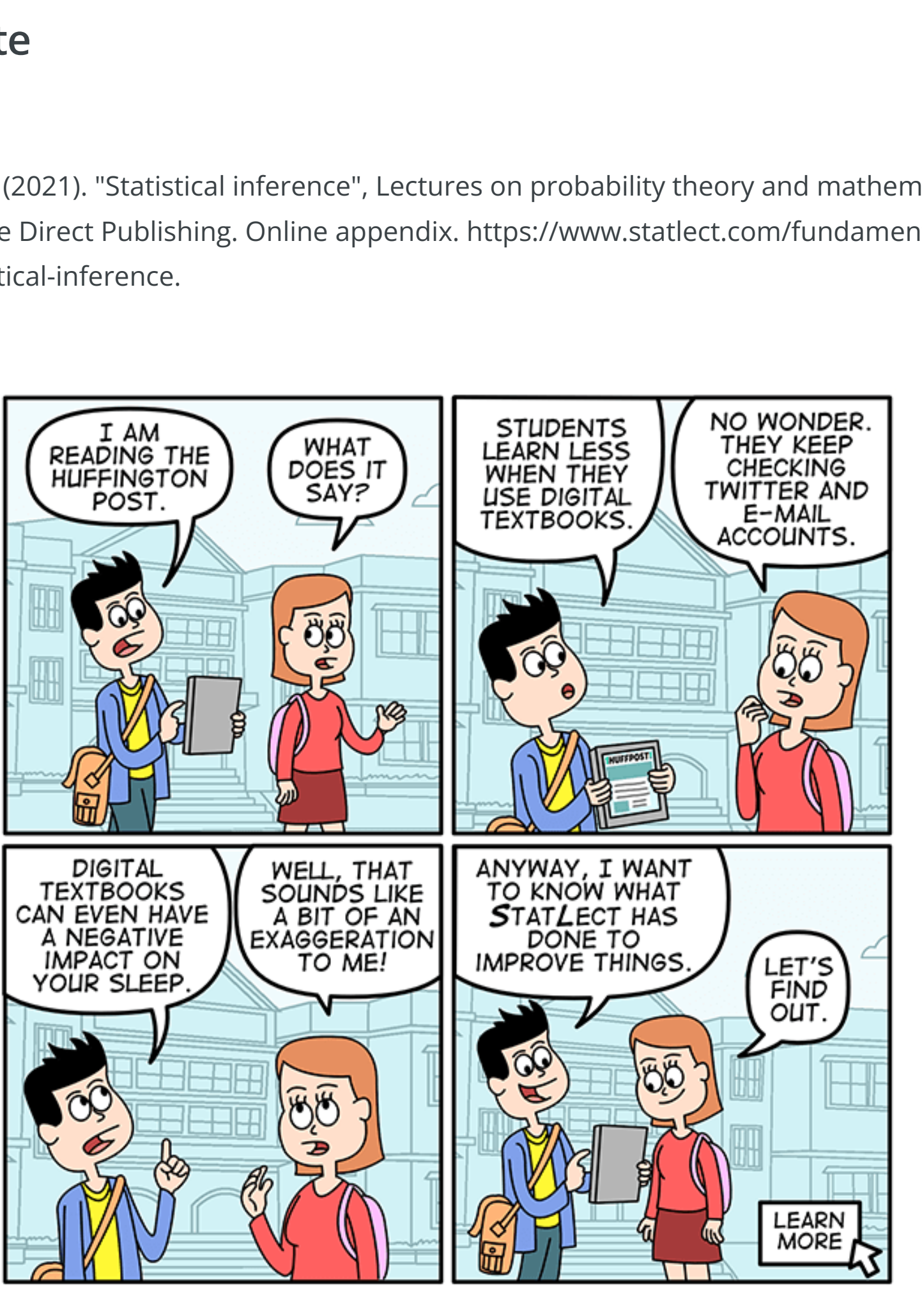
In this lecture we have touched on several important topics. You can read more about these topics in the following pages:

- statistical model;
- parameter space;
- parameter.

How to cite

Please cite as:

Taboga, Marco (2021). "Statistical inference", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/fundamentals-of-statistics/statistical-inference.



The books

Most of the learning materials found on this website are now available in a traditional textbook format.

Probability and statistics

Matrix algebra

Featured pages
Chi-square distribution
Central Limit Theorem
Uniform distribution
Mean square convergence
Convergence in distribution
Student t distribution

Main sections
Mathematical tools
Fundamentals of probability
Probability distributions
Asymptotic theory
Fundamentals of statistics
Glossary

Glossary entries
Alternative hypothesis
Mean squared error
Factorial
Continuous random variable
Power function
Estimator

Explore
Maximum likelihood
Score test
Characteristic function

About
About Statlect
Contacts
Cookies, privacy and terms of use

Share
To enhance your privacy,
we removed the social buttons,
but **don't forget to share**.