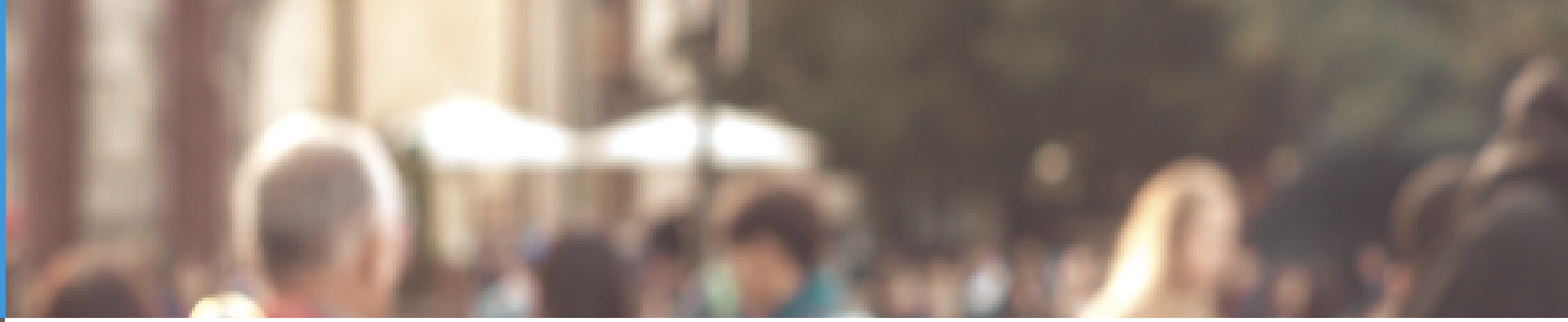


Elements of Structural Modelling

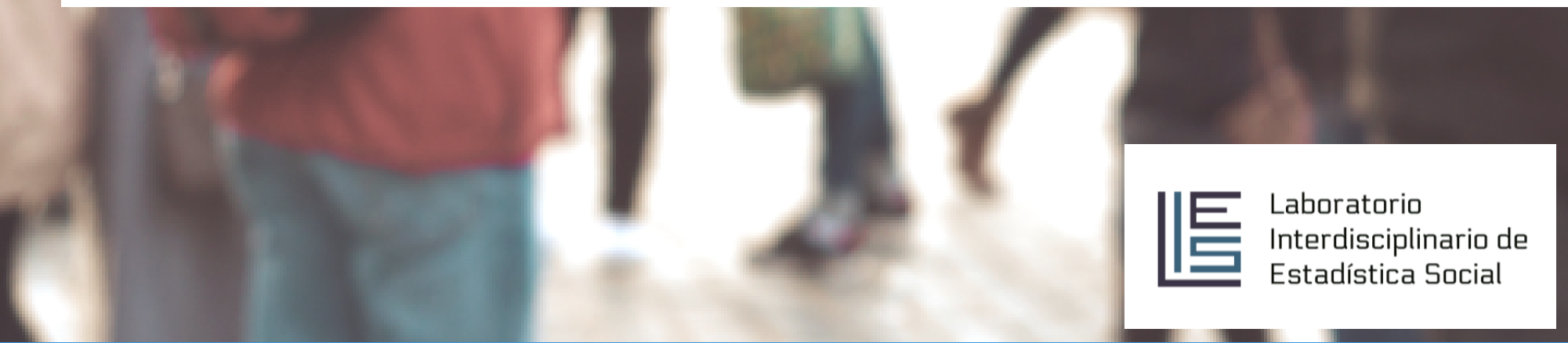
Overview

Cursillo de Verano 2020, Facultad de Matemáticas, UC

Ernesto San Martín

Laboratorio Interdisciplinario de Estadística Social LIES, UC
The Economics School of Louvain, Université Catholique de Louvain, Belgium

January 2021



Statistical Modelling

- Let's start with some apparent naive questions:

Statistical Modelling

- Let's start with some apparent naive questions:
  - What does statistical modelling mean?

Statistical Modelling

- Let's start with some apparent naive questions:
  - What does statistical modelling mean?
  - What inputs are required for statistical modelling?

Statistical Modelling: A procedure?

Laboratorio Interdisciplinario de Estadística Social

- What is required to make the meaning of statistical modeling explicit?

Statistical Modelling: A procedure?

- What is required to make the meaning of statistical modeling explicit?
- Is it sufficient to describe the steps of such a procedure? (assuming that statistical modelling is a (kind of) procedure!

Statistical Modelling: A procedure?

- What is required to make the meaning of statistical modeling explicit?
- Is it sufficient to describe the steps of such a procedure? (assuming that statistical modelling is a (kind of) procedure!
- Are "data" a fundamental ingredient of statistical modelling? Or, is statistical modelling a procedure that is "used on those data"?

Statistical Modelling: A procedure?

- What is required to make the meaning of statistical modeling explicit?
- Is it sufficient to describe the steps of such a procedure? (assuming that statistical modelling is a (kind of) procedure!
- Are "data" a fundamental ingredient of statistical modelling? Or, is statistical modelling a procedure that is "used on those data"?
- How is data defined or characterized? Is a taxonomy of "types of data" enough: quantitative data, categorical data, nominal data, . . . ? In fact, the teaching of Statistical Methods is usually organized on the basis of such taxonomy.

Statistical Modelling, Statistical Models?

- When we teach and learn linear regression, are we teaching and learning how to model statistically?

Statistical Modelling, Statistical Models?

- When we teach and learn linear regression, are we teaching and learning how to model statistically?
- Or do we learn what statistical modeling is when we note that a linear relationship between two or more variables is not enough to "model" that relationship? We learn new "statistical models" such as non-linear regression, do we then learn to model?

Breiman's (2001) Perspective

- Statistical Modelling: Two Cultures (*Statistical Science 16*, 199-231):

  Statistics starts with data. Think of the data as being generated by a black box in which a vector of input variables **x** *(independent variables) go in one side, and on the other side the response variables* **y** *come out. Inside the black box, nature functions to associate the predictor variables with the response variables* [. . .]

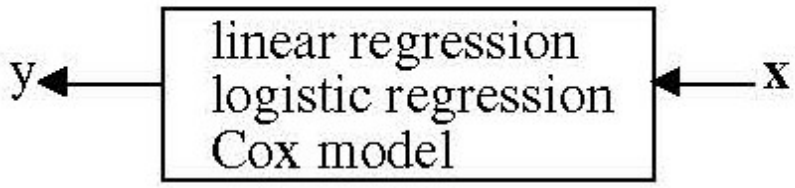
  There are two goals in analyzing data:
  Prediction. *To be able to predict what responses are going to be to future input variables.*
  Information. *To extract some information about how nature is associating the response variables to the input variables.*

Breiman's (2001) Perspective

- Breiman distinguishes two different approaches:
  - Data modelling culture: the analysis in this culture starts with assuming a stochastic data model for the inside of the black box. The model validation is performed through goodness of fit.
  - Algorithmic modelling culture: The analysis in this culture considers the inside of the box complex and unknown. Their approach is to find a function $f(\mathbf{x})$ -an algorithm that operates on $\mathbf{x}$ to predict the responses $\mathbf{y}$.

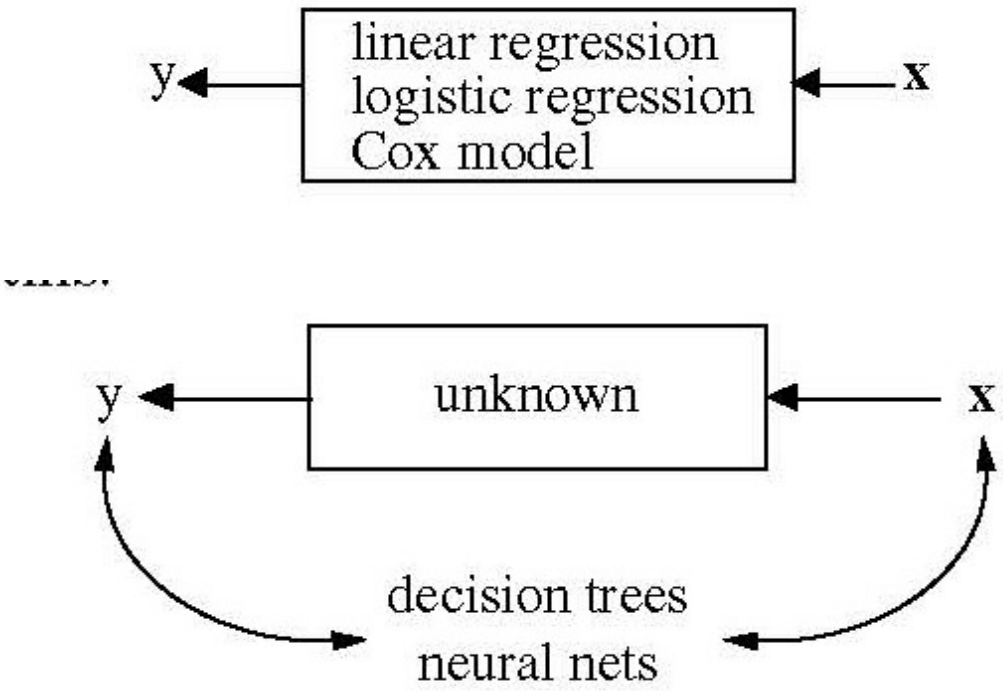
Breiman's (2001) Perspective

- Breiman summarizes both approaches as follows:



Breiman's (2001) Perspective

- Breiman summarizes both approaches as follows:



Some questions

- Is there anything missing from Breiman's perspective?

Some questions

- Is there anything missing from Breiman's perspective?
- Are all the terms he uses in his description basic undefined terms?

Re-reading . . .

- Let us re-read Breiman's (2001) claim:

  Statistics starts with data. Think of the data as being generated by a black box in which a vector of input variables **x** *(independent variables) go in one side, and on the other side the response variables* **y** *come out. Inside the black box, nature functions to associate the predictor variables with the response variables* [. . .]

- What do the terms highlighted in red or blue mean?

- Can they be defined or, as in an axiomatic theory, are they the undefined elements of such a theory?

Box's (1976) Perspective

- Many statisticians often quote the following statement attributed to Box:

  All models are wrong, but some are useful.

Box's (1976) Perspective

- Many statisticians often quote the following statement attributed to Box:

  All models are wrong, but some are useful.

- Is it a cliche (i.e., an expression whose content has been lost) or a principle of statistical modelling?

Box's (1976) Perspective

- Many statisticians often quote the following statement attributed to Box:

  All models are wrong, but some are useful.

- Is it a cliche (i.e., an expression whose content has been lost) or a principle of statistical modelling?

- The above question makes sense because how do you define wrong, and what do you mean by useful?

Box's (1976) Perspective

- Many statisticians often quote the following statement attributed to Box:

  All models are wrong, but some are useful.

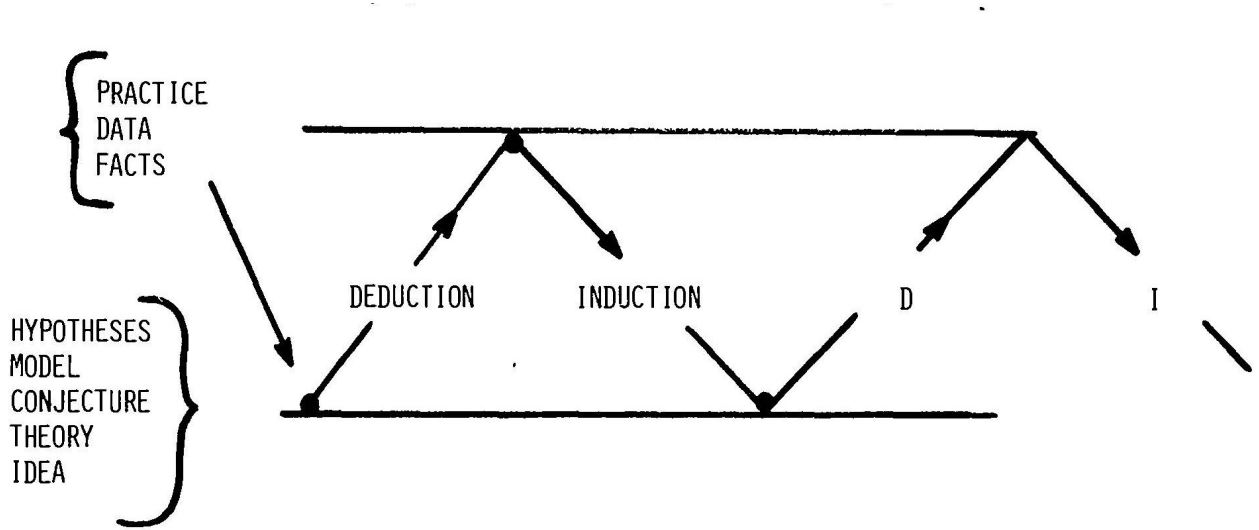
- Is it a cliche (i.e., an expression whose content has been lost) or a principle of statistical modelling?

- The above question makes sense because how do you define wrong, and what do you mean by useful?

- How is model defined or characterized?

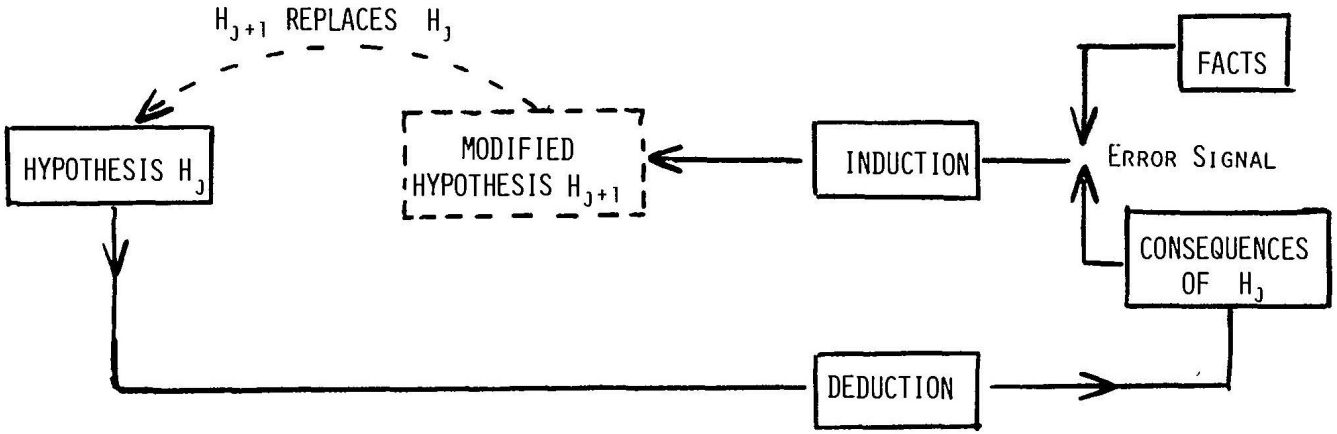
Box's (1976) Perspective

- Box (1976) describes the aspects of scientific method:
  - Iteration between theory and practice:



Box's (1976) Perspective

- Flexibility:



Box's (1976) Perspective

- Parsimony:

Box's (1976) Perspective

- Parsimony:

  Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity.

Box's (1976) Perspective

- Parsimony:

  Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity.

- Box's statement may thus be seen as a critique of the (current) sophistication of statistical models.

Box's (1976) Perspective

- Worrying Selectively:

Box's (1976) Perspective

- Worrying Selectively:

  Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.

Box's (1976) Perspective

- Worrying Selectively:

  Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.

- How to define criteria to identify what is more and less important? Will the field of application have any role to play here?

Box's (1976) Perspective

- Worrying Selectively:

  Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.

- How to define criteria to identify what is more and less important? Will the field of application have any role to play here?

- It should be remembered that the fields of application induce a dichotomy in the teaching of statistical methods: biostatistics, psychometrics, econometrics, sociometrics, epidemiology, ...

Box's (1976) Perspective

- Role of Mathematics in Science:

Box's (1976) Perspective

- Role of Mathematics in Science:

  Equally, the statistician knows, for example, that in nature there never was a normal distribution, there never was a straight line, yet with normal and linear assumptions, known to be false, he can often derive results which match, to a useful approximation, those found in the real world.

  The researcher hoping to break new ground in the theory of experimental design should involve himself in the design of actual experiments. The investigator who hopes to revolutionize decision theory should observe and take part in the making of important decisions. An appropriately chosen environment can suggest to such an investigator new theories or models worthy to be entertained. Mathematics artfully employed can then enable him to derive the logical consequences of his tentative hypotheses and his strategically selected environment will allow him to compare these consequences with practical reality. In this way he can begin an iteration that can eventually achieve his goal.

Statistics and Statistical Modelling

- Is all statistical development consistent with any statistical modelling perspective?
- Are all statistical methodologies useful for any of these perspectives?

Structural Modelling

- We will focus our attention on Structural Modelling.

Structural Modelling

- We will focus our attention on Structural Modelling.
- It is a movement developed since the 40's: Frisch, Haavelmo, Hurwicz, Koopmans ...

Structural Modelling

- We will focus our attention on Structural Modelling.
- It is a movement developed since the 40's: Frisch, Haavelmo, Hurwicz, Koopmans ...
- Rasch, Reiersol, Herman Rubin and H. Simon also participated to this movement.

Structural Modelling

- We will focus our attention on Structural Modelling.
- It is a movement developed since the 40's: Frisch, Haavelmo, Hurwicz, Koopmans ...
- Rasch, Reiersol, Herman Rubin and H. Simon also participated to this movement.
- The original discussion papers can be found at https://cowles.yale.edu/.

Structural Modelling: Key features

- The concept of structural model was developed under the motto no measurement without theory (Koopmans, 1947).
- Structural models connect theories and facts.

Questions

- What does *combining facts and theories* mean?
- Is it enough to say that the context of the application providing the data is taken into account?
- Is it enough to choose the variables that come into play according to a specific substantive theory?
- Do these considerations impact the way statistics is developed?

Structural Modelling: Key features

- Model corresponds to a Statistical model.
- To define a statistical model, we use a basic tool: probability:

Structural Modelling: Key features

- Model corresponds to a Statistical model.
- To define a statistical model, we use a basic tool: probability:

  When we speak of the probability *of a certain object fulfilling a certain condition, we imagine all such objects to be divided into two classes, according as they do or do not fulfil the condition. This is the only characteristic in them of which we take cognisance. For this reason probability is the most elementary of statistical concepts. It is a parameter which specifies a simple dichotomy in an infinite hypothetical population, and it represents neither more nor less than the frequency ratio which we imagine such a population to exhibit (Fisher, 1922, p. 312).*

Structural Modelling: Key features

- Model corresponds to a Statistical model.
- To define a statistical model, we use a basic tool: probability:

  When we speak of the probability *of a certain object fulfilling a certain condition, we imagine all such objects to be divided into two classes, according as they do or do not fulfil the condition. This is the only characteristic in them of which we take cognisance. For this reason probability is the most elementary of statistical concepts. It is a parameter which specifies a simple dichotomy in an infinite hypothetical population, and it represents neither more nor less than the frequency ratio which we imagine such a population to exhibit (Fisher, 1922, p. 312).*

- Do we agree that statistics builds its claims and methodologies using probabilities?

Structural Modelling: Key features

- Haavelmo (1944):

  This study is intended as a contribution to econometrics. It represents an attempt to supply a theoretical foundation for the analysis of interrelations between economic variables. It is based upon the modern theory of probability and statistical inference (p. iii).

Structural Modelling: Key features

- Haavelmo (1944):

  This study is intended as a contribution to econometrics. It represents an attempt to supply a theoretical foundation for the analysis of interrelations between economic variables. It is based upon the modern theory of probability and statistical inference (p. iii).

- Neyman (1937), when discussing the type of objects to which it is possible to apply "probability,"says.

  The problem of the definition of measure in relation to the theory of probability has been recently discussed by Łomnicki and Ulam (1934), who quote extensive literature. A systematic outline of the theory of probability based on that of measure is given by Kolmogoroff (1933). See also Borel (1925-26); Lévy (1925); Fréchet (1937).

Kolmogorov 1933/1950

- A set S the elements of which are called elementary events.
- A family $\mathcal{S}$ of subsets of S which contains all the permissible combinations of subsets of S:
  - $S \in \mathcal{S}$.
  - $A \in \mathcal{S} \Longrightarrow A^c \in \mathcal{S}$.
  - $A, B \in \mathcal{S} \Longrightarrow A \cup B \in \mathcal{S}$.
  - $\{A_n : n \in \mathbb{N}\} \in \mathcal{S} \Longrightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{S}$.

  The elements of $\mathcal{S}$ are called random events.

Kolmogorov 1933/1950

- A set S the elements of which are called elementary events.
- A family $\mathcal{S}$ of subsets of S which contains all the permissible combinations of subsets of S:
  - $S \in \mathcal{S}$.
  - $A \in \mathcal{S} \Longrightarrow A^c \in \mathcal{S}$.
  - $A, B \in \mathcal{S} \Longrightarrow A \cup B \in \mathcal{S}$.
  - $\{A_n : n \in \mathbb{N}\} \in \mathcal{S} \Longrightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{S}$.

  The elements of $\mathcal{S}$ are called random events.
- $P : \mathcal{S} \longrightarrow [0, 1]$ such that $A \longmapsto P(A)$, which is characterized by the following properties:
  - $P(S) = 1$.
  - $A, B \in \mathcal{S}$ with $A \cap B = \varnothing$ then $P(A \cup B) = P(A) + P(B)$.
  - For a decreasing sequence of events

    $$A_1 \supset A_2 \supset \cdots \subset A_n \cdots$$

    such that $\bigcap_{n\in\mathbb{N}} A_n = \varnothing$ then $\lim_{n\to\infty} P(A_n) = 0$.

  The number $P(A)$ is called the probability of event A.

Some remarks

- Random event is just a designation of the elements of $\mathcal{S}$. We prefer to call them events of interest and therefore $\mathcal{S}$ contains all the information that is of interest.
- Probability of an event A is just a number.

Some remarks

- Random event is just a designation of the elements of $\mathcal{S}$. We prefer to call them events of interest and therefore $\mathcal{S}$ contains all the information that is of interest.
- Probability of an event A is just a number.
- Random and Probability are undefined terms. It is important that statisticians always remember this!

Construction of a Probability Space

- The simplest fields of probability are constructed as follows. We take an arbitrary finite set $S = \{\xi_1, \xi_2, \ldots, \xi_k\}$ and an arbitrary set $\{p_1, p_2, \ldots, p_k\}$ of non-negative numbers with the sum

$$p_1 + p_2 + \cdots + p_n = 1.$$

  $\mathcal{S}$ is taken as the set of all subsets of E, and we put

$$P\{\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_\lambda}\} = p_{i_1} + p_{i_2} + \cdots + p_{i_\lambda}.$$

- In such cases, $p_1, p_2, \ldots, p_n$ are called probabilities of the elementary events $\xi_1, \xi_2, \ldots, \xi_n$ or simply elementary probabilities.

- In this way we derived all possible *finite* fields of probability in which $\mathcal{S}$ consists of the set of all subsets of S (Kolmogorov, 1933/1950, pp. 2-3).

Specification Problems

- Fisher (1922) defined the task of the statistician:

  The object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data.

Specification Problems

- Fisher (1922) defined the task of the statistician:

  The object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data.

- This object is accomplished by constructing a hypothetical infinite population, of which the actual data are regarded as constituting a random sample.

Specification Problems

- Fisher (1922) defined the task of the statistician:

  The object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data.

- This object is accomplished by constructing a hypothetical infinite population, of which the actual data are regarded as constituting a random sample.

- The law of distribution of this hypothetical population is specified by relatively few parameters, which are sufficient to describe it exhaustively in respect of all quantities under discussion. Any information given by the sample, which is of use in estimating the values of these parameters, is relevant information.

Specification Problems

- Since the number of independent facts supplied in the data is usually far greater than the number of facts sought, much of the information supplied by any actual sample is irrelevant. Is the object of the statistical processes employed in the reduction of data to exclude this irrelevant information, and to isolate the whole of the relevant information contained in the data.
- It will be seen [...] that the process of the reduction of data is, even in the simplest cases, performed by interpreting the available observations as a sample from a hypothetical infinite population.

Specification Problems

- The problems which arise in reduction of data may be conveniently divided into three types:
  - Problems of Specification. These arise in the choice of the mathematical form of the population.
  - Problems of Estimation. These involve the choice of methods of calculating from a sample statistical derivates, or as we shall call them statistics, which are designed to estimate the values of the parameters of the hypothetical population.
  - Problems of Distribution. These include discussions of the distribution of statistics derived from samples, or in general any functions of quantities whose distribution is known.

Digression

- The specification problems deal with the specification of the probability distribution that represents the infinite hypothetical population of which the actual data are considered as a random sample. This probability distribution depends on parameters which in turn describe the population of interest.

Digression

- The specification problems deal with the specification of the probability distribution that represents the infinite hypothetical population of which the actual data are considered as a random sample. This probability distribution depends on parameters which in turn describe the population of interest.
- More specifically:
  - Probability distributions $\{P^a : a \in A\} \Longleftrightarrow$ Infinite hypothetical population.
  - Parameters $a \in A$ are supposed to be interpreted with respect to the population under analysis.
  - Probability distributions describe an observed population.

Statistical model

- A statistical model $\mathcal{E}$ is defined as

$$\mathcal{E} = \{(S, \mathcal{S}),\ P^a : a \in A\},$$

where

  - $(S, \mathcal{S})$ is the sample space, that is, the space of all plausible observations.
  - P^a is a sampling probability defined on the sample space; it is indexed by a parameter a.
  - A is the parameter space.

Statistical model

- The statistical modelling process requires, therefore,
  - To make explicit the observations.
  - To construct the sampling probabilities $\{P^a : a \in A\}$, making explicit the parameters of interest, namely a function of $a \in A$.

Example

- Empirical research is typically based on the analysis of survey data.

Example

- Empirical research is typically based on the analysis of survey data.
- Most (if not all) the collected data are in the form of variables taking values in finite sets.
- Examples include educational attainment, language proficiency, workers' union status, employment status, health conditions, and health/functional status.

Example

- Income level: typically intervals of income.
- Let suppose there are 5 levels of income levels:

$$w = 1, \quad w = 2, \quad w = 3, \quad w = 4, \quad w = 5.$$

Example

- Income level: typically intervals of income.
- Let suppose there are 5 levels of income levels:

$$w = 1, \quad w = 2, \quad w = 3, \quad w = 4, \quad w = 5.$$

- Is

$$\{P(w = j) : j = 1, \ldots, 5\}$$

the statistical model?

Example

- If it is, then $P(w = j)$ with $j = 1, \ldots, 5$, constitute the characteristics of the population under study.

Example

- If it is, then $P(w = j)$ with $j = 1, \ldots, 5$, constitute the characteristics of the population under study.
- The implicit assumption is that what is observed is what effectively characterizes the respondent's income level.

Example

- If it is, then $P(w = j)$ with $j = 1, \ldots, 5$, constitute the characteristics of the population under study.
- The implicit assumption is that what is observed is what effectively characterizes the respondent's income level.
- However, if we take into account theories of human behavior in relation to transmitting information that seems sensitive, then these probabilities do not represent the parameters of interest.

Example

- If it is, then $P(w = j)$ with $j = 1, \ldots, 5$, constitute the characteristics of the population under study.
- The implicit assumption is that what is observed is what effectively characterizes the respondent's income level.
- However, if we take into account theories of human behavior in relation to transmitting information that seems sensitive, then these probabilities do not represent the parameters of interest.
- How the parameters of interest can be specified?

Example

- Let x be a random variable that represents the true income level. Therefore, the parameters of interest are

$$\{P(x = j) : j = 1, \ldots, 5\}.$$

Example

- Let x be a random variable that represents the true income level. Therefore, the parameters of interest are

$$\{P(x = j) : j = 1, \ldots, 5\}.$$

- How the parameters of interest are related to the parameters of the statistical model?

Example

- Let x be a random variable that represents the true income level. Therefore, the parameters of interest are

$$\{P(x = j) : j = 1, \ldots, 5\}.$$

- How the parameters of interest are related to the parameters of the statistical model?

- If we assume that each respondent knows his or her income level and answers the survey taking into account it, then

$$\begin{pmatrix} P(w=1) \\ \vdots \\ P(w=5) \end{pmatrix} = \begin{pmatrix} P(w=1 \mid x=1) & \cdots & P(w=1 \mid x=5) \\ \vdots & \cdots & \vdots \\ P(w=5 \mid x=1) & \cdots & P(w=5 \mid x=5) \end{pmatrix} \begin{pmatrix} P(x=1) \\ \vdots \\ P(x=5) \end{pmatrix}.$$

Example

- Relevant modelling question: To what extent do the parameters $\{P(w = j)\}$ provide some information about the parameters $\{P(x = j)\}$?
- This question deals with an identification problem.

Example

- Relevant modelling question: To what extent do the parameters $\{P(w = j)\}$ provide some information about the parameters $\{P(x = j)\}$?
- This question deals with an identification problem.
- Note that the problem of identification arose because of what we assume about the behavior of the respondents, not from the observations themselves (the survey responses).

Specification problems

- The specification problem is related to the specification of the probability distribution of the observations.
- Typically, we think in a random variable X and its probability distribution P^a.

Specification problems

- The specification problem is related to the specification of the probability distribution of the observations.
- Typically, we think in a random variable X and its probability distribution P^a.
- However, the population of interest can be characterized by more than one random variable.
- Suppose each member of the population is characterized by two random variables X and Y. Therefore, it is necessary to specify the joint distribution of (X, Y), namely

$$P^a(X \leqslant x, Y \leqslant y), \quad (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

Specification problems

- The structural modelling perspective emphasizes that the joint process/probability $P^a(X, Y)$ can be decomposed into simpler pieces.
- Taking into account that our basic tool is the probability, such a decomposition can be only done by making marginal-conditional decompositions.

Example

- Let us consider three random variables:
  - Tabaquismo T.
  - Cáncer al sistema respiratorio C.
  - Exposición al abesto A.
  - Estatus socio-económico S.
- Nos interesa explicar el fenómeno generado por $P(T, C, A, S)$.
- ¿Qué sabemos? En ciertos lugares de Europa, se observa que:
- Aquellos de bajo S tienden a fumar más y a trabajar en medio-ambientes insalubres y pro tanto están más expuestos a A.
- Luego S es una variable causal clave que hay que incluir en un modelo de prácticas.

Ejemplo

- Es decir,

$$A \perp\!\!\!\perp T \mid S.$$

- Además, estudios clínicos muestran que A y T producen C para poblaciones diferentes. Luego

$$C \perp\!\!\!\perp SES \mid A, T.$$

- Luego, la descomposición recursiva se educa a

$$\begin{aligned} p(S, A, T, C) &= P(C \mid A, T, S)\, P(A, T \mid S)\, P(S) \\ &= P(C \mid A, T)\, P(A \mid S)\, P(T \mid S)\, P(S). \end{aligned}$$

- Aquí, $P(C \mid A, T)$ representa el mecanismo biológico, y $P(A \mid S)P(T \mid S)$ representa el mecanismo social.

Example

- Let us revisit the specification of a linear regression and its extensions.
- Let (x, y) be a 2-dimensional random vector.
- According to our discussion, we specify the joint distribution of (x, y), namely

$$P(y, x)$$

Example

- Let us revisit the specification of a linear regression and its extensions.
- Let (x, y) be a 2-dimensional random vector.
- According to our discussion, we specify the joint distribution of (x, y), namely

$$P(y, x)$$

- Thereafter we perform the following decomposition:

$$P(y, x) = P(y \mid x)P(x).$$

Why?

Example

- But typically we say

$$y = \alpha + \beta x + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2).$$

- How can we embed this specification in the previous framework?

Example

- But typically we say

$$y = \alpha + \beta x + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2).$$

- How can we embed this specification in the previous framework?
- The previous specification is equivalent to the following one:

$$(y \mid x) \sim \mathcal{N}(\alpha + \beta x, \sigma^2)$$

So, one element of the previous specification is the invariance of the form of a distribution under linear transformations.

Example

- But typically we say

$$y = \alpha + \beta x + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2).$$

- How can we embed this specification in the previous framework?
- The previous specification is equivalent to the following one:

$$(y \mid x) \sim \mathcal{N}(\alpha + \beta x, \sigma^2)$$

  So, one element of the previous specification is the invariance of the form of a distribution under linear transformations.

- Moreover, we focus our attention on one characteristic of $P(y \mid x)$, namely

$$E(y \mid x).$$

Example

- But typically we say

$$y = \alpha + \beta x + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2).$$

- How can we embed this specification in the previous framework?
- The previous specification is equivalent to the following one:

$$(y \mid x) \sim \mathcal{N}(\alpha + \beta x, \sigma^2)$$

  So, one element of the previous specification is the invariance of the form of a distribution under linear transformations.

- Moreover, we focus our attention on one characteristic of $P(y \mid x)$, namely

$$E(y \mid x).$$

- What is the role of the distribution of $P(x)$?

Example

- Therefore, when we write

$$y = \alpha + \beta x + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

we can say that

$$\alpha + \beta x$$

is a conditional expectation of $P(y \mid x)$.