Elements of Structural Modelling:

What is an identification problem?

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

Laboratorio Interdisciplinario de Estadística Social

The Problems of Statistics

- Fisher (1922):
  - Scope of statistical methods: reduction of data without loosing relevant information.
  - *No human mind is capable of grasping in its entirety the meaning of any considerable quantity of numerical data*.
  - First reduction: to consider the observation under analysis as a random sample of an infinite population.
  - Infinite population: a metaphor that allows us to "identify" a population of interest with the probability distribution generating it.

Statistical model

- All the relevant information comes from the observations.
- Such a relevant information is captured by the probability distribution (sampling distribution).
- The characteristics of such a distributions are (should be) functionals of the sampling distribution.
- Such functionals are called parameters, which *are sufficient to describe [the population] exhaustively in respect of all qualities under discussion* (Fisher).

$$
\begin{tikzcd}
\theta \arrow[r] \arrow[dr] & P^{\theta} \arrow[d, shift right] \\
 & \text{Observations} \arrow[u, shift right]
\end{tikzcd}
$$

The problems of statistics

- Problems of Specification arise in the choice of the mathematical form of the probability distribution which characterizes the observed population. Its choice *is not arbitrary, but requires an understanding of the way in which the data are supposed to, or did in fact, originate*.
- Problems of Estimation.
- Problems of Distribution.

A relevant question, I

- We have observations from a population. We attribute concepts to such observations in order to understand or describe the population:

  Are such characteristics represented as functional of the sampling distribution?

  Are such characteristics (function of the) parameters of the sampling distribution?

- When the answer is negative, we face an identification problem.

A relevant question, II

Laboratorio Interdisciplinario de Estadística Social

- Haavelmo (1007):

  The one type of problems, seemingly paradoxical, grew out of a rather intricate consequence of successful economic theory. Strangely enough, the fact is that if an economic theory, an economic relation, is a good theory, true to reality, it may not be possible to quantify it by using data from the economic of which that relation is a part. This is the so-called "problem of identification".

A relevant question, II

- Haavelmo (1007):

The one type of problems, seemingly paradoxical, grew out of a rather intricate consequence of successful economic theory. Strangely enough, the fact is that if an economic theory, an economic relation, is a good theory, true to reality, it may not be possible to quantify it by using data from the economic of which that relation is a part. This is the so-called "problem of identification".

$$
\begin{tikzcd}
\theta \arrow[r] \arrow[dr] & P^{\theta} \arrow[d, shift right] \\
& \text{Observations} \arrow[u, shift right]
\end{tikzcd}
$$

A relevant question, III

- Koopmans and Reiersol (1950):

In many fields the objective of the investigator's inquisitiveness is not just a "population" in the sense of a distribution of observable variables, but a physical structure projected behind this distribution, by which the latter is thought to be generated. The word "physical" is used merely to convey that the structure concept is based on the investigator's ideas as to the "explanation" or "formation" of the phenomena studied, briefly, on his theory of these phenomena, whether they are classified as physical in the literal sense, biological, psychological, sociological, economic or otherwise.

Statistical model, I

- The specification problem leads to define a statistical model
    - Sample space $(S, \mathcal{S})$, where S is the set of all possible outcomes, and $\mathcal{S}$ is the σ-field of subsets of S: the elements of $\mathcal{S}$ are considered as the events of interest.
    - A probability distribution P^θ defined on the sample space $(S, \mathcal{S})$, where θ is a parameter. P^θ is accordingly called sampling distribution or sampling probability.
    - $\theta \in \Theta$, where the set Θ is called parameter space.
- A statistical model can compactly be written as

$$\mathcal{E} = \{(S, \mathcal{S}), P^\theta : \theta \in \Theta\}.$$

Statistical model, I

- If the statistical model is dominated (that is, when for all $\theta \in \Theta$, P^θ is absolutely continuous with respect to a σ-finite measure λ defined on $(S, \mathcal{S})$), the statistical model is equivalently described through the likelihood function p^θ, where $p^\theta = \frac{dP^\theta}{d\lambda}$:

$$\mathcal{E} = \{(S, \mathcal{S}) : p^\theta : \theta \in \Theta\}.$$

Statistical model, II

- This concept covers what it is commonly called parametric, semi-parametric and non-parametric models: the distinction depends on the structure of Θ.
  - The parameter space Θ of a parametric model is contained in a finite-dimensional space.
  - For non-parametric models, the parameter space Θ is an infinite-dimensional space.
  - Semi-parametric models can be described by saying that Θ is a Cartesian product of the form $\boldsymbol{A} \times \boldsymbol{\Psi}$, where $\boldsymbol{A}$ is a finite-dimensional space and $\boldsymbol{\Psi}$ is an infinite-dimensional space.

Example I

- Let us consider an ANOVA 1 Fixed Factor (balanced case):

$$Y_{ij} \overset{\text{indep.}}{\sim} \mathcal{N}(\alpha + \beta_j, \sigma^2), \quad i = 1, \dots, n;\ j = 1, \dots, J. \tag{1}$$

  Let

$$\mathbf{Y}_j = (Y_{1j}, \dots, Y_{nj})' \quad j = 1, \dots, J; \quad \mathbf{Y} = (\mathbf{Y}_1', \dots, \mathbf{Y}_J')'.$$

- The mutual independence implies that

$$\mathbf{Y} \sim \mathcal{N}_{nJ}\left(\alpha \mathbb{1}_{nJ} + \begin{pmatrix} \beta_1 \mathbb{1}_n \\ \vdots \\ \beta_J \mathbb{1}_n \end{pmatrix},\ \sigma^2 I_{nJ}\right), \tag{2}$$

  where $\mathbb{1}_n$ is a n-dimensional vector of 1's and I_{nJ} is the $nJ \times nJ$ identity matrix. This, (2) characterizes the fixed- effect model: the data are mutually independent and the means depend of j.

Example I

- A random-effect model is specified through a marginal-conditional decomposition: $i = 1, \ldots, n$; $j = 1, \ldots, J$

$$\text{(i) } (Y_{ij} \mid \beta_j) \overset{\text{c.ind}}{\sim} \mathcal{N}(\alpha + \beta_j, \sigma^2), \quad \text{(ii) } \beta_j \overset{\text{ind}}{\sim} \mathcal{N}(0, \tau^2). \tag{3}$$

- The specification (3) does not correspond to the statistical model because a statistical model describes the *observations*.

- (3.i) is a conditional distribution, whereas (1) is a marginal distribution: therefore, both distributions are *not* related between them and consequently one specification can not be viewed as a kind of "re-interpretation" of the other.

- c.ind means that the for each j, $\{Y_{ij} : i = 1, \ldots, n\}$ are mutually independent conditionally on β_j.

Example I

- The induced statistical model is obtained after integrating out the β_j's:

$$\mathbf{Y}_j \sim \mathcal{N}_n\left(\alpha \mathbb{1}_n, \begin{pmatrix} \sigma^2+\tau^2 & \tau^2 & \cdots & \tau^2 \\ \tau^2 & \sigma^2+\tau^2 & \cdots & \tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau^2 & \tau^2 & \cdots & \sigma^2+\tau^2 \end{pmatrix}\right). \tag{4}$$

- This means that two observations Y_{ij} and $Y_{i'j}$ are correlated between them.

- A random effect model is used when the scientist have reason to believe that the observations nested in the j-group are correlated in the sense that the j-group induces such a correlation. This is not the case in a fixed-effect model. The choice depends on substantive considerations related to the fact that the j-group induces correlation or not.

Identification problems, I

- The definition of statistical model implicitly involves a relationship between the parameter space and the set of sampling distributions defined on $(S, \mathcal{S})$.
- If we denote this set as $\mathcal{P}(S, \mathcal{S})$, such a relationship corresponds to the mapping

$$\mathbf{\Phi} \ : \Theta \longrightarrow \mathcal{P}(S, \mathcal{S})$$

such that $\mathbf{\Phi}(\theta) = P^{\theta}$ for $\theta \in \Theta$.

- The parameter θ is identified if the mapping Φ is injective:

$$\theta_1 \neq \theta_2 \Longrightarrow P^{\theta_1} \neq P^{\theta_2}.$$

Identification problems, II

- Statistical inference unsupported by a substantive theory applies to whatever statistical regularities and, consequently, stable relationships can be discerned in the data. Such purely empirical relationships are likely to be due to the presence and persistence of underlying structural relationships.

- However, the direction of this deduction cannot be reversed (from the empirical to the structural relationships) expect possibly with the help of a substantive theory which specifies the form of the structural relationships. The possibility of this deduction necessarily implies to solve an *identification problem* at the model construction level.

Identification problems, III

- Conclusion: the properties of a population under study are represented by a parameter. Consequently, if a property supposed to be a characteristic of a population can not be represented by a parameter of the sampling probability, such a property is not supported by the observations: this corresponds to an identification problem. Therefore, the statistical meaning of parameters becomes explicit by an identification analysis.

Example II

- Let us illustrate these considerations with the ANOVA 1 Fixed factor.
- Two standard identification restrictions:

$$\beta_1 = 0 \quad \text{or} \quad \sum_{j=1}^{J} \beta_j = 0.$$

- Under the first restriction:

$$\alpha = \mu_1 = E(Y_{i1}), \qquad \beta_j = \mu_j - \mu_1 = E(Y_{ij}) - E(Y_{i1}).$$

Example II

- Under the second restriction:

$$\alpha = \frac{1}{J}\sum_{j=1}^{J}\mu_j = \frac{1}{J}\sum_{j=1}^{J}E(Y_{ij}),$$

$$\beta_j = \mu_j - \frac{1}{J}\sum_{j=1}^{J}\mu_j = E(Y_{ij}) - \frac{1}{J}\sum_{j=1}^{J}E(Y_{ij}).$$

Example III

- Rasch model: let

$$Y_{ih} = \begin{cases} 1, & \text{if person } i \text{ correctly answers item } j; \\ 0, & \text{in any other case;} \end{cases} \quad ,$$

where $Y_{ij} \sim \text{Bern}\,(p_{ij})$ mutually independent.

- Rasch model focuses on *relative comparisons*:

$$p_{ij} = \frac{\lambda_{ij}}{1+\lambda_{ij}}, \quad \lambda_{ij} \geqslant 0,$$

which is equivalent to

$$\lambda_{ij} = \frac{P(Y_{ij}=1)}{P(Y_{ij}=0)}.$$

Example III

- The model specification is completed by assuming that

$$\lambda_{ij} = \frac{\epsilon_i}{\eta_j} \quad i = 1, \ldots, I, \; j = 1, \ldots, J,$$

  where ϵ_i is a characteristic of person i and η_j is a characteristic of item j.

- The question is whether these parameters can be written as a function of λ_{ij}'s. The answer is *not*.

Example III

- Let $\{\epsilon_i, \eta_j\}$ be a set of parameters satisfying the previous equalities; and let $\{\epsilon'_i, \eta'_j\}$ be other different set. It follows that

$$\frac{\epsilon_i}{\eta_j} = \frac{\epsilon'_i}{\eta'_j} \qquad \forall\,(i,j) \in \{1,\ldots,I\} \times \{1,\ldots,J\},$$

  which is equivalent to

$$\frac{\epsilon_i}{\epsilon'_i} = \frac{\eta_j}{\eta'_j} \qquad \forall\,(i,j) \in \{1,\ldots,I\} \times \{1,\ldots,J\},$$

  which necessarily implies that

$$\frac{\epsilon_i}{\epsilon'_i} = \frac{\eta_j}{\eta'_j} = \alpha \qquad \forall\,(i,j) \in \{1,\ldots,I\} \times \{1,\ldots,J\},$$

  where α is a constant.

- $\{\epsilon_i, \eta_j\}$ are not identified and, consequently, can not be statistically interpreted.

Example III

- In order to remove the inherent indeterminacy, we choose an item, for instance $j = 1$, as a *standard item* in the sense that we fix a "unit". Thus, $\eta_1 = 1$ and therefore
$$\epsilon_i = \lambda_{i1} = \frac{P(Y_{i1} = 1}{P(Y_{i1} = 0)}.$$

- The parameter that characterizes person i corresponds to the betting odd of answering correctly the standard item 1. In particular,
$$\epsilon_i > \epsilon_{i'} \Longleftrightarrow P(Y_{i1} = 1) > P(Y_{i'1} = 1).$$

  This comparison explains why psychometricians call ϵ_i *ability of person i*, although any psychological phenomenon underlies ϵ_i.

Example III

- Similarly,

$$\eta_j = \frac{P(Y_{i1} = 1)}{P(Y_{i1} = 0)} \frac{P(Y_{ij} = 0)}{P(Y_{ij} = 1)}.$$

- Then
  - $\eta_j > 1 = \eta_1 \iff P(Y_{i1} = 1) > P(Y_{ij} = 1)$ for all person i.
  - $\eta_j > \eta_k \iff P(Y_{ik} = 1) > P(Y_{ij} = 1)$ for all person i.

- These inequalities explains why psychometricians call η_j *difficulty of item j*. Thus, for instance, item j is more difficult that item k if and only if the probability to correctly answer item j is smaller than the probability to correctly answer item k.

Example III

- The Rasch model allows us to compare the individual characteristic ϵ_i and the item characteristic η_j:

$$\epsilon_i > \eta_j \Longleftrightarrow P(Y_{ij} = 1) > P(Y_{ij} = 0).$$

- The ability of person i is greater than the difficulty of item j if and only if the probability that person i answers correctly item j is greater than the probability of answering incorrectly. This property is know as the simultaneous representation p"person $\times$ items"

Concluding remarks

Laboratorio Interdisciplinario de Estadística Social

- Identification analysis is a relevant component of model construction: the empirical interpretation of the parameters of interest depend on this type of analysis.
- Furthermore, all property that a scientist assumes to be a characteristic of a population under study *is* actually a characteristic of it *only if is a function of the corresponding sampling distribution*.
- This aspect becomes relevant for empirical research because such a research *always* involves an identification problem. The way in which the identification problem is solved will constitute the so-called *maintained assumption*.

Concluding remarks

- When this type of assumptions is combined with empirical evidence, we reach scientific conclusions and thus we conform a corpus of scientific knowledge.
- Based on this knowledge, science becomes useful for informing policy decisions. Nevertheless, we will emphasize that the maintained assumptions are not unique and, therefore, we will make explicit the subjective side of scientific knowledge. This is one of the aspects we will discuss in what follows.