

## TOWARD THE APPLICABILITY OF STATISTICS: A REPRESENTATIONAL VIEW

MAHDI ASHOORI

*Iranian Institute of Philosophy, IRAN*

ashoori@irip.ir

S. MAHMOUD TAHERI

*University of Tehran, IRAN*

sm\_taheri@ut.ac.ir

---

**Abstract.** The problem of understanding how statistical inference is, and can be, applied in empirical sciences is important for the methodology of science. It is the objective of this paper to gain a better understanding of the role of statistical methods in scientific modeling. The important question of whether the applicability reduces to the representational properties of statistical models is discussed. It will be shown that while the answer to this question is positive, representation in statistical models is not purely structural. In spite of the fact that representation in statistical models is based on the structural similarities between the statistical model and the empirical systems under study, these relationships are shown to be appropriate for representing relations in the target system by agent function, too. A second aspect of the paper involves the claim that agent-based components of statistical modeling are: a) interpretation of random variables, b) selection of the goal of statistical research, and c) selection of estimator properties. To justify these claims, a preliminary discussion will be presented on the role of statistics in modeling, as in regression and other structural models. This role will be explored and realized using a structural viewpoint. Also the role of statistical estimation in statistical modeling is discussed to explain the representational role of models and the inferential role of the agent in modeling.

**Keywords:** Applicability • Statistics • Scientific representation • Structural model • Statistical methods.

---

RECEIVED: 02/05/2017

REVISED: 22/01/2018; 15/01/2019

ACCEPTED: 25/01/2019

### 1. Introduction

Nowadays, statistical inference has an undeniable role in scientific activities. Statisticians and researchers, who employ statistical methods, are, intuitively, aware of the fact that statistics is applicable to scientific modeling. The question then arises, is how this applicability can be explained? (Freedman, 1995). A well-known approach to answering this question is to note that scientific models are applicable, because



they have a representational role (Brown, 1997; Batterman, 2010). We would like to explore this role of statistics in scientific modeling, because “representation” is an important feature of a “model”. Representation is a concept that has been considered in various philosophical areas over the last few decades. The prospect of philosophers who use the concept of “scientific representation” is that they can explain the function of scientific theories and models. A theory of representation constitutes the necessary and sufficient conditions for scientific research (for example, see Suppes, 2002, pp. 51-80).<sup>1</sup>

Recent works on the nature of scientific representation have unraveled a conflict between the following two points of view: 1) informational views which focus on the objective relations between representations (theories, models, simulations, diagrams, etc.) and the target systems, as necessary and sufficient conditions, and 2) functional views which draw upon cognitive activities and the roles, such as interpretation<sup>2</sup> and inference,<sup>3</sup> played by the agent<sup>4</sup> in connection with these target systems (e.g. Chakravartty, 2010). In the structural viewpoint, the objective relations could be isomorphism,<sup>5</sup> partial isomorphism,<sup>6</sup> and homomorphism.<sup>7</sup>

Models have a pivotal role in many scientific contexts. Statistics has been frequently used in scientific modeling. The main function served by each model has, however, been one of representation. Such models are, typically, some simple representation or description of the entity or system being studied. This paper deals with the following two main areas:

1. Structural relations between statistical models and their target systems; and
2. The role of the agent in the representation by statistical models.

Initially, in Section 2, we consider structural accounts as the philosophical context and the conceptual toolbox to reconstruct statistical models. Implicit, to a purely structural account, is the assumption that the representational role of a statistical model is primitive and exhaustive, and that its other roles are reduced only to its representational properties. In Section 3, we explore the representational role of models to show that a purely structural account fails to explain the role of statistics in modeling. Rather, it is shown that statistics has an extrinsic role in the representation of scientific models. In Section 4, we discuss the inferential role of the agent in estimating model parameters. Section 5 introduces a hybrid account of the “representation of statistical models” from 1) necessary (Section 2) however insufficient (Section 3) structural relations between statistical models and target systems and 2) agent-based components of representation (Section 4). Finally, Section 6 provides some conclusions and remarks.

## 2. Structural reconstruction of the applicability of statistics

In this section, we explore the question of whether the applicability of statistics will reduce to the representational properties of statistical models. In representational theories, two aspects must be considered: 1) the structural relations between the model and the target system, and 2) the agent-based components. For this purpose, we will reconstruct statistical models from a structural perspective (Section 2.2) in an attempt to explicate the elements of the models and the target systems and their components. To do this, first, we will review the structural literature in the philosophy of mathematics (Section 2.1).

### 2.1. Structuralism in the philosophy of mathematics

The literature on structuralism has centered on a variety of questions and problems. The main question in mathematical structuralism is ontological about mathematical objects; however in this paper, we focus on the structural answer to the question of applicability.<sup>8</sup> In this regard, two types of applicability need to be distinguished: 1) the applicability of sciences in general, and 2) the applicability of a specific discipline to other sciences. Mathematics and statistics exemplify the second type, but they play different roles in sciences and are applicable in varied ways. Many questions may be raised about the application of statistics that have been neglected by philosophers. However, studies have been reported on the applicability of mathematics, which we can exploit as useful to our present concerns.

In the philosophy of mathematics, different ideas have been put forth about mathematical entities and their semantics, and a number of structural accounts have been presented to explain the applicability and utility of mathematics. The ‘structuralist view’ of mathematical objects, in reference to mathematical objects, is always in the context of some background structure, and that the objects have no more to them than can be expressed in terms of the “basic relations of the structure” (Parsons, 1990, p. 303); i.e., from the metaphysical point of view, the mathematical objects do not have an intrinsic nature, or the structure requires a whole of mathematical objects together. Structuralists believe that “mathematics is a rich source of structures”, and some mathematical structures are either designed to, or otherwise found to, accurately capture the important structural relations of an empirical set-up (Bueno and Colyvan, 2011, p. 346).

The structural view of applicability maintains that there exist certain structure-preserving mappings between the world and the mathematical structure in question (Pincock, 2007), and that mathematical structures represent the physical set-up. Structural accounts explain the utility of mathematics in some applied situations by demonstrating the existence of the right kind of “map” from a mathematical structure

to some appropriate physical set-up. In other words, the applicability of mathematics is accounted for by appreciating the relevant structural similarities, e.g. isomorphism or homomorphism, between the empirical system under study, and the mathematics used in the investigation of that system.

Based on this view, the usefulness of mathematics is much like that of a city street “map”; so that, such view is called the “mapping” account. This account is based on the implicit assumption that mathematics is no more than a convenient representational system.<sup>9</sup> The mapping account and the purely structural view are instantiations of informational accounts of representation. Informational views focus on the necessary and sufficient objective relations between mathematics and the physical set-up that is mapping function, in a purely structural view. But in functional view, representation and applicability are constituted by some cognitive functions.

## 2.2. Structural reconstructing on statistical models

In this section, we try to reconstruct the statistical models in a structural view, so we can examine the claim that statistical models have a structural similarity to a physical set-up. A structure is usually taken to be a set of objects and a set of relations on these. Let  $\mathbf{T}$  be the empirical situation to be represented (or a “target system”), and  $\mathbf{M}$  be the structure to be used in a model of  $\mathbf{T}$ . According to the structural view,  $\mathbf{M}$  and  $\mathbf{T}$  are set-theoretic structures;<sup>10</sup>  $\mathbf{M} = \langle A, R \rangle$  and  $\mathbf{T} = \langle A', L \rangle$ , where  $A$  is a set of model elements, and  $R$  is the relation between these elements.  $A'$  is a set of target system elements, and  $L$  is the relation between the target system elements. Such an account, then, states that a model  $\mathbf{M}$  represents a target system  $\mathbf{T}$  if there is a function  $f : A \rightarrow A'$  that establishes the following relations between  $R$  and  $L$ :<sup>11</sup>

$$a_i \in A, a_j \in A, a'_i \in A', a'_j \in A', \text{ and } f(a_i) = a'_i \text{ and } f(a_j) = a'_j, \\ \text{then } L(f(a_i), f(a_j)) \leftrightarrow R(a_i, a_j)$$

To explain how we can use the structural view to reconstruct statistical models, we focus on a main class of statistical models called *structural models* (SMs). SMs consist of two sets of variables:  $U$  and  $V$ , and a set of functions,  $F$ , that determine, or simulate, how values are assigned to each variable  $v_i \in V$ . Thus, for example, the equation  $v_i = f_i(v, u)$  describes a physical process by which the variable  $v_i$  is assigned as the response to the current values,  $v$  and  $u$ , of all the variables in  $V$  and  $U$ . Formally, the triplet  $\langle U, V, F \rangle$  defines an SM (for more details, Pearl (2014) may be consulted). Several statistical models/methods, such as regression models, factor analyses, path analyses, variance analyses, and structural equations models (SEM), are instances of  $F$  in the SM triplet. In this language,  $A = \{u_i, v_i\}$  and  $R = \{f_i\}$ .

Consider for instance,  $\mathbf{M} = \langle A, R \rangle$  as a simple linear regression model, in which  $A = \{Y, X\}$  and  $R : Y = \beta_0 + \beta_1 X + \epsilon$  is determined by the structural coefficients  $\{\beta_0, \beta_1, \epsilon\}$ .<sup>12</sup> In the real world, there is a structure  $W = \langle T, L \rangle$ , in which  $T = \{t_1, t_2\}$  with the actual relation  $L$  between  $t_1$  and  $t_2$  which we expect to be represented by the model  $\mathbf{M}$ . According to the structural account, there exists a structural relation  $f$  between  $R$  and  $L$  expressed by  $R(Y, X) \leftrightarrow L(f(t_1), f(t_2))$ . In such a structural view, the mapping is defined to hold between the statistical equations and the structural relations under study. The question arises at this point is, whether there exist any mapping between such statistical equations and structural relations under study. Accepting that the statistical equations are determined by the estimation of the parameters, we then need to answer the question about what the interpretation of the parameters would be. We will revisit this question in Section 4.

### 3. Structural mapping and underdetermination<sup>13</sup> of statistical models

Let us assume that any statistical model, as a system of equations, represents a target system. But what is the role of the statistical model in the representation itself? We may entertain two positive answers. The first is a strong answer, which asserts that the statistical model lies in the content of the representation; which may be called an ‘intrinsic role.’<sup>14</sup> In other words, the statistical inference enhances the content of the representation so that the statistical model, as an equation, represents a real relationship in the subject under study, and that the parameters of the equation have a real interpretation. The second one is a weak answer, which maintains that the statistical elements have the role of evaluating the representations of the assumptions and the conceptual model. In this section, we will explore these two answers.

As already explained in the previous section, in the statistical model  $\mathbf{M} = \langle A, R \rangle$ ,  $R$  is a system of equations. For example, in  $R : Y = \beta_0 + \beta_1 X + \epsilon$ , the relation  $R$  is determined by the structural parameters  $\{\beta_0, \beta_1\}$ . These parameters are usually estimated by estimators based on sample data. A key question here is whether this model can represent the target system? The weak answer runs into no controversies. However, the strong answer faces two problems: 1) underdetermination of the estimation of parameters and 2) the context-dependency of the interpretation of the parameters. In the following section, we examine these two problems.

#### 3.1. Underdetermination of the estimation of parameters

One task of statistical inference is to estimate model parameters. The statistical equations are then determined by the parameters thus estimated. Understanding the “role

of structural relations between a statistical model and the related target system” relies on the examination of the estimation process. We begin with the properties of a good estimator in classical statistics.<sup>15</sup>

The goal of point estimation is to make a reasonable guess at the unknown value of a designated population quantity, such as model parameters. This comprises an estimation procedure, or an ‘estimator.’ Typically, a point estimator is any function  $U = f(X_1, X_2, \dots, X_n)$  of a sample (Cassella and Berger, 2002: 311). The result of applying the estimator to the data is an estimate.

There are several techniques for finding point estimators. But how can we tell if a method of point estimation is best? For the classical approach to estimation, which assumes that there is some fixed, but unknown, true value for the parameter being estimated, we would ideally like to identify the superior alternative from among the different estimators available. Unfortunately, since the true value is unknown, we can never know this exactly. It is, therefore, better to study the estimation procedures and to identify the statistical properties that these random variables possess. In the present case, four properties of statistical models are worth noting:

1. *Unbiasedness.* An estimator  $\hat{\theta}$  is said to be unbiased for  $\theta$  if  $E(\hat{\theta}) = \theta$ . Thus, on average, the procedure for guessing the real value of  $\theta$  produces the correct value. The property of unbiasedness is intuitively appealing and is, in most cases, quite useful.
2. *Precision.* Usually the concept of precision is referred to as the *variance*. Thus, the smaller the value of  $\text{Var}(\hat{\theta})$ , the more precise  $\hat{\theta}$  will be.
3. *Efficiency.* Often, estimators are not both unbiased and precise. Efficiency is related to the concept of mean squared error which is expressed by  $\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$ . MSE incorporates two components, one measuring the variability of the estimator (precision) and the other measuring its bias (accuracy). The estimator  $\hat{\theta}_1$  is said to be more efficient than  $\hat{\theta}_2$  if  $\text{MSE}(\hat{\theta}_1) \leq \text{MSE}(\hat{\theta}_2)$  for all  $\theta$ , (and inequality strictly holds for at least one  $\theta$ ).
4. *Consistency.* An estimator  $\hat{\theta}$  is said to be consistent for  $\theta$  if the estimator  $\hat{\theta}$  converges in probability to  $\theta$  with increasing sample size. The property of consistency is an essential one because it ensures that the criterion converges in probability to the true value.

The aforementioned criteria can be used in the assessment of any estimator and in the selection of an appropriate estimator from among different alternatives.

From a structural viewpoint, if the parameters and coefficients are not fully determined (underdetermined), Then, the equation  $R$  and the statistical model  $\mathbf{M}$  will not be fully determined. But based on different alternative estimators, choosing the

appropriate estimator depends on the agent. In this case, no determined mapping function,  $f$ , is established between  $R$  and the relation  $L$  in the target system so that the representation of the target system by the model will depend on the agent function. This is the first problem faced within the strong answer to the representational role of statistical models.

### 3.2. Context-dependency in interpretation of parameters

There are other problems faced by the independent interpretation of coefficients. To investigate these problems in statistical models, we will reconsider the simple linear regression,  $Y = X + \epsilon$ . Let us suppose that a connection exists between  $X$  and  $Y$  that is represented by  $\beta$ . If we interpret the coefficient  $\beta$  as the change in  $E(Y)$  per unit change of  $X$ , then, after rewriting the equation as  $X = (Y - \epsilon)/\beta$ , we ought to interpret  $1/\beta$  as the change in  $E(X)$  per unit change of  $Y$ . But this is in conflict with both our intuitions and model predictions. The change in  $E(X)$  per unit change of  $Y$  ought to be zero if  $Y$  does not appear as an independent variable in the original equation.

Teachers of statistical methods generally adopt one of two escape routes to avoid this dilemma. One route involves denying that  $\beta$  represents a real connection in the target system and that  $\beta$  is settled for a purely statistical interpretation in which  $\beta$  measures the reduction in the variance of  $Y$  explained by  $X$  (Multhen, 1987). The other route permits real connections for only those coefficients that meet the “isolation” constraint (Bollen, 1989; James et al., 1982); the explanatory variable must be uncorrelated with the error in the equation. Because  $\epsilon$  cannot be uncorrelated with both  $X$  and  $Y$  (or so the argument goes),  $\beta$  and  $1/\beta$  cannot both have a real meaning or interpretation in the target system, thus, the paradox is resolved.

The first route is self-consistent, but it denies any real interpretation of the parameters. The second is vulnerable to logical criticism on the grounds that both  $\alpha$  and  $\beta$  represent a real connection in the subject under study. It is well-known that every pair of bivariate normal variables,  $X$  and  $Y$ , can be expressed in two equivalent ways,

$$Y = \beta X + \epsilon_1 \quad \text{or} \quad X = \beta Y + \epsilon_2,$$

where  $\text{Cov}(X, \epsilon_1) = \text{Cov}(Y, \epsilon_2) = 0$  and  $\alpha = r_{XY} = (\epsilon_X^2 / \epsilon_Y^2)\beta$ . Thus, if the condition  $\text{Cov}(X, \epsilon_1) = 0$  endows  $\beta$  with a real meaning and interpretation, then  $\text{Cov}(Y, \epsilon_2) = 0$  ought to endow  $\alpha$  with a real meaning as well. But this, too, is in conflict with both intuition and the intentions behind regression and other structural models. The change in  $E(X)$  per unit change of  $Y$  ought to be zero, rather than  $r_{XY}$ , if there is no real path from  $Y$  to  $X$  (Pearl, 2009: 159–160).

The two problems raised above — that are, the underdetermination of estimation and the context-dependency of the interpretation of model coefficients — lead us to say that we cannot claim any determined mapping function exists between the statistical model and the relations in the target system as the sufficient condition for representation. We cannot, therefore, claim either that the structural coefficients of scientific models contribute to the content of representation. We cannot have a real interpretation of estimated parameters, nor can we exactly know about our estimations. If so, then what is the role of statistical estimation in the representation by statistical models? Clearly, estimation of parameters is not part of the representational content; thus, the strong answer is ruled out. Indeed, we can say that the statistical elements and relations have *extrinsic* roles to play in scientific representation.

## 4. Agent-based components of statistical model representation

In the previous section, it was shown that the statistical elements of a scientific model have no more of a mapping role in the representation content than does the conceptual model; hence, the purely structural view is not adequate for explaining the representation by and the applicability of statistics. The question that awaits an answer is whether there are cognitive components in statistical modeling that have a constitutive role in representation and applicability? In what follows, we will first introduce the agent-based components in statistical modeling, and then proceed to identify the component(s) which is/are constitutive in representation.

### 4.1. Agent-based components in statistical modeling

Drawing upon different principles, different paradigms of statistical inference have been developed. The following three are more common in current use: 1) Classical or frequentist statistics, 2) Bayesian statistics, and 3) Likelihood-based statistics. The first step in the agent-based statistical modeling is the selection of a statistical paradigm. These paradigms differ in a number of ways as outlined below.

The first difference concerns an interpretational one. While the classical and likelihood-based statistics are based on the physical interpretation of random variables, the Bayesian one relies on an epistemic interpretation of random variables (Freedman, 1995; Berger, 1995; Bandyopadhyay and Forster, 2011). In the physical interpretation, random events and chances are regarded as frequencies, or proportions, in a class of similar events or items. According to the epistemic interpretation, probabilities may be taken to represent the strength or degree of belief (Gillies, 2000). In this context, the key point is that interpretations can all be connected to founda-



tional paradigms of statistical inference. Thus, certain cognitive activities, such as interpretation, are performed in drawing statistical inferences.

The second difference is one of underlying intentions. For illustration, let us consider two hypotheses:  $H$ , stating that a patient suffers from tuberculosis, and  $\sim H$ , its denial. Let us further assume that an X-ray, which is administered as a routine test, turns out to be positive for the patient. Based on this simple scenario and following Richard Royall (1997), one could raise three questions that underlie the epistemological issue at stake within the three competing statistical paradigms:

- (a) Given the datum, what should we *do*?
- (b) Given the datum, what should we *believe* and to what *degree*?
- (c) What does the datum say regarding the *evidence* supporting  $H$  over its alternative?

The first question is called the *decision* question, the second is the *belief* question, and the third is the *evidence* question. Royall maintains that classical statistics addresses the decision question, that Bayesians address the belief question, and that it is only the likelihood paradigm that addresses the evidence question (Royall, 1997). Whether or not any relationship can be established between classical statistics and the decision question, there always remain three goals and intentions in any statistical investigation and certain cognitive activities, such as intention, that are performed in statistical inference.

Based on the paradigmatic differences, there are two agent-based components involved: 1) interpretation of random variables, and 2) the intentions underlying a statistical inference.

The third agent-based component is “choosing an appropriate estimator.” The goal of point estimation is to make a reasonable guess at the values of the parameters. A difficulty that arises is cases where there are different estimators from among which the statistician needs to select an appropriate one (Figure 1). We recall from Section 3.1 that the properties of a good estimator in classical statistics include biasedness, consistency, precision, and efficiency which are used by the statistician as the criteria in selecting an appropriate estimator. This selection process is an agent-based component.

It is clear thus far that, in statistical modeling, there exist the three agent-based components of 1) interpretation of random variables, 2) intentions underlying a statistical inference, and 3) choosing an appropriate estimator.

## 4.2. Constitutive components

Based on the mapping account presented above, an agent could use a representation for understanding a target system. There can be agent-based components in model-

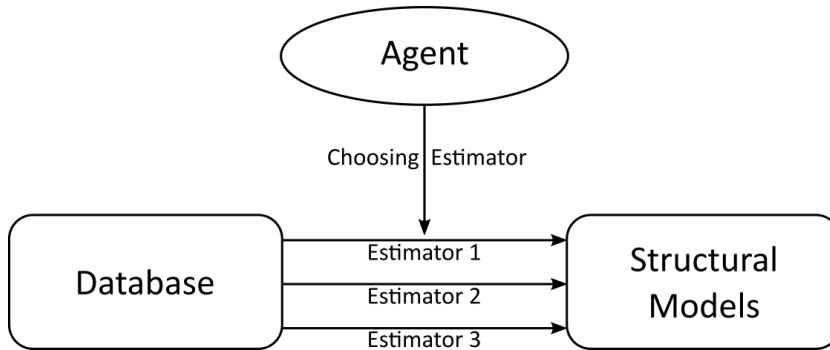


Figure 1: Agent-based components in estimation.

ing, but this representation is not constituted by the agent. That is, the ‘vehicle’ of representation represents the ‘target’ of the representation in virtue of the fact that the vehicle produces a state with the appropriate content (Bueno and French, 2011, p. 871-873). Now, we may ask whether these functional activities are constitutive for scientific representation. In other words, there are alternative relations between source and target, but a specific relation is not determined, unless, an agent determines a special relation, so this special relation as a representation is constituted by the agent.

From a functional viewpoint, a scientific representation is a tripartite relationship,  $R_S(M, T, A)$ , where  $M$  is the model,  $T$  is the target system, and  $A$  is the set of agent-based components. From an informational viewpoint, the scientific representation is a bipartite relationship,  $R_S(M, T)$ . Within the framework of the informational approach, one can ask whether or not the model,  $M$ , represents the target system,  $T$ ? In the functional approach, however, the question of whether the operating model,  $M$ , represents the target system,  $T$  should be asked from the agent,  $A$ ?

The agent is constitutive for the representation. If there are several distinct and rival mappings,  $F : \{f_1, f_2, \dots, f_n\}$ , between the model,  $M$ , and the target system,  $T$ , then each mapping,  $f_i$ , leads to different structural similarities, and it is the agent who selects one from among them. The change in the mapping leads to changes in the representation. Thus, the representation not only depends on the model,  $M$ , and the target system,  $T$ , but also on the “selection”, “intention”, or “interpretation” rendered by the agent.

The next question that may be raised is whether the representation by a statistical model depends on the paradigms and the estimator selected by the modeler? To answer this question, two scenarios may be considered (Figure 2):

- (a) The agent-based components take place prior to the modeling while there are also different models to choose from; or

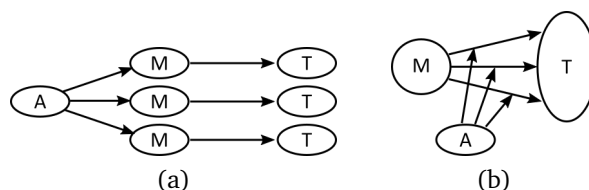


Figure 2: Two scenarios for the relationship between the model and the target system. (a) The agent-based components are prior to modeling. (b) The agent-based components are posterior to modeling.

(b) The agent-based components are posterior to modeling.

Under scenario (a), the models depend on the agent, but there is a function mapping between the model and the target system. Under scenario (b), the representation depends on the agent-based components. It seems that the selection of the “goal of inference” and the “estimator” takes place prior to the modeling, while the interpretation of the statistical model is second. If the agent judges that a paradigm or an estimator is appropriate, there will be no structural relationship between the model and the target system. In scenario (a),  $M$  and  $T$  depend on  $A$ . Thus, the representation depends on the agent as well.

The three steps of mathematical application of Bueno and Colyvan (2011) are helpful in describing the inferential concept of statistical models:

1. Immersion: the first step consists of establishing a mapping from the empirical set-up to a convenient mathematical structure.

It is doubtful whether this step can be accomplished in physical models since these models are not constructed based on data. In statistical modeling, however, immersion can be implemented by data collection.

2. Derivation: the second step consists in drawing the consequences from mathematical formalism using the mathematical structure obtained in the immersion step. This is the key point in the application process, whereby consequences from mathematical formalism are generated.

Statistical inferences, as in estimation and descriptive statistics, are derived from the database structures and turned into equations.

3. Interpretation: we interpret the mathematical consequences obtained in the derivation step in terms of the initial empirical set-up. To establish an interpretation, a mapping from the mathematical structure to the initial empirical

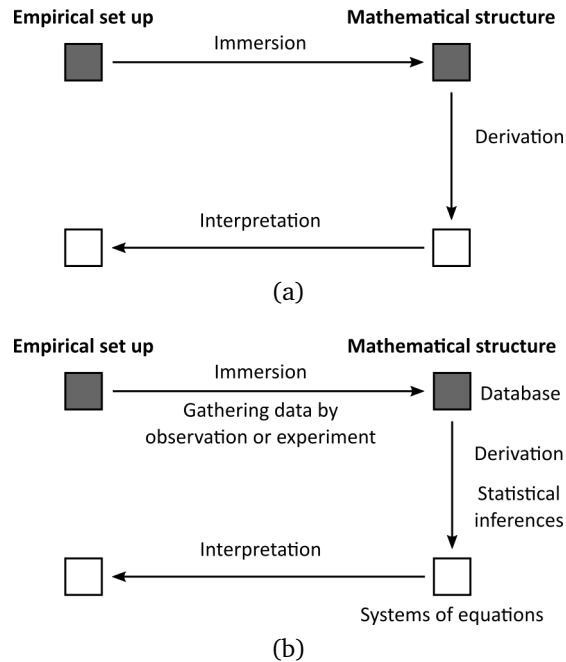


Figure 3: The three-step scheme of applicability. (a) Bueno and Colyvan's scheme. (b) Three steps of applicability of statistics.

set-up is needed. This mapping need not be simply the inverse of the mapping used in the immersion step.

We can expect that statistical models represent aspects of the world, and that it is the interpretation step that carries out the task of representation. Figure 3 develops these steps by Bueno and Colyvan (2011) in statistics.

According to Bueno and Colyvan, the inferential conception involves the following two steps: 1) immersion, and 2) interpretation (2011, p. 356), however, we show in this section, the derivation step in statistics is also agent-based. In addition, the agent is in a position to affect the mapping between the empirical set-up and the statistical structure.

## 5. Representation of statistical models

A theory of representation expresses the necessary and sufficient conditions for scientific representations along the following lines:

The model or the scientific theory,  $M$  (source), represents the target system,  $T$ , if and only if (1) ..., (2) ..., and ...

These theories state how the source of scientific representation represents the target system. As already mentioned, two accounts may be suggested as for the nature of scientific representation: 1) the informational account, which emphasizes the objective structural model and their target systems; and 2) the functional account, which emphasizes the cognitive activities and the agent's role in connection with these targets (Chakravartty, 2010; Giere, 2010).

In the informational account, the statistical models are representative if there are relevant structural similarities between the statistical model and the relations in the empirical system under study. In its broad sense, there could be a homomorphism between the statistical model and the relations in the empirical system.

In Section 2, we showed the structural virtues in statistical models that can represent the physical set-up as a target system.

In Section 3, we showed that no exact structural similarity could exist between statistical parameters determining the structural model and the real relations in the target system under study. In other words, the true values of the parameters and the statistical equations determined by these parameters are unknown. The estimated values might be true, but we cannot use them to interpret the structural relations. Then, in the representations by the statistical models, the informational view does not meet the condition of sufficiency.

In Section 4, we showed that there are agent-based components that might be constitutive in the representations by statistical models.

We can, thus, claim that a statistical model represents the subject matter itself, if:

1. There are some relationships, as structural virtues, in the statistical model that resemble the target system; and
2. The agent identifies these relationships as appropriate for representing the relations in the subject under study based on: 2a) the interpretation of random variables, 2b) the goal of the statistical research, and 2c) the properties of the estimators.

This theory of "Representation of Statistical Models" is a hybrid account from informational and functional that is based on structural similarities between the statistical model and the empirical system under study, and assessment of the agent in using these structural similarities to represent the target.

## 6. Conclusion

In this paper, we evaluated the claim that “statistical models are applicable if they represent the empirical system under study”. In representation theories, we studied: 1) structural similarities between the model and the target system, and 2) the agent-based components of statistical modeling to propose a hybrid account of representation by statistical models. Based on this account, we may claim that statistical models are applicable if they represent the empirical system under study. Statistical inference does not, however, play a role in the content of representation because the estimation of parameters and the statistical equations are unknown.

Moreover, the following claims were investigated:

1. We can explain the applicability of statistics by its representational properties; but
2. The representation of statistical models is not purely structural;
3. The representation of statistical models is based on structural similarities between the statistical model and the empirical system under study; but
4. It is the agent who identifies these relationships as appropriate for representing the relations in the target system; and
5. The agent-based components of statistical modeling are: a) interpretation of random variables, b) selection of the goals of the statistical research, and c) selection of the estimator properties.

In future studies, the relationships between explanation and representation in statistical models may be explored.

## References

- Bandyopadhyay, P.; Forster, M. (eds.) 2011. *Philosophy of Statistics*. Handbook for the Philosophy of Science Series. London: Elsevier.
- Bartels, A. 2006. Defending the structural concept of representation. *Theoria*, **55**: 7–19.
- Batterman, R. 2010. On the Explanatory Role of Mathematics in Empirical Science. *British Journal for the Philosophy of Science*, **61**: 1–25.
- Berger, J. 1995. Discussion of David Freedman’s “Some issues in the foundation of statistics”. *Foundations of Science*, **1**: 41–44.
- Bollen, K. A. 1989. *Structural Equations with Latent Variables*. Hoboken: J. Wiley.
- Brown, J. R. 1997. What is Applied Mathematics?. *Foundations of Science*, **2**, pp. 21–37.
- Bueno, O.; Colyvan, M. 2011. An inferential conception of the application of mathematics. *Nous*, **45**(2): 345–374.

- Bueno, O.; French, S. 2011. How theories represent. *British Journal of the Philosophy of Science*, **62**: 857–894.
- Cassella, G.; Berger, R. 2002. *Statistical Inference*. 2nd Ed. Duxbury: Cengage Learning.
- Chakravartty, A. 2010. Informational versus functional theories of scientific representation. *Synthese*, **172**: 197–213.
- Contessa, G. 2007. Scientific representation, interpretation, and surrogative reasoning. *Philosophy of Science*, **74**(1): 48–68.
- Da Costa, N.; French, S. 2003. *Science and Partial Truth*. Oxford: Oxford University Press.
- Freedman, D. 1995. Some issues in the foundation of statistics. *Foundations of Science*, **1**(1), pp. 19–39.
- French, S. 2003. A model-theoretic account of representation. *Philosophy of Science*, **70**: 1472–1483.
- Frigg, R. 2006. Scientific representation and the semantic view of theories. *Theoria*, **50**: 49–65.
- Giere, R. N. 2010. An agent-based conception of models and scientific Representation. *Synthese*, **172**: 269–281.
- Gillies, D. 2000. *Philosophical Theories of Probability*. London: Routledge.
- Hughes, R. I. G. 1997. Models and representation. *Philosophy of Science*, **64**: 325–336.
- James, L. R.; Mulaik, S. A.; Brett, J. M. 1982. *Causal Analysis: Assumptions, Models, and Data*. Beverly Hills: Sage Publications.
- Multhen, B. 1987. Response to Freedman’s critique of path analysis: Improve credibility by better methodological training. *Journal of Educational Statistics*, **12**: 178–184.
- Nodelman, U.; Zalta, E. N. 2014. Foundations for Mathematical Structuralism. *Mind*, **123**(1): 39–78.
- Parsons, C. 1990. The Structuralist View of Mathematical Objects. *Synthese* **84**: 303–46.
- Pearl, J. 2009. *Causality: Models, reasoning, and inference*. 2nd Ed. Cambridge, UK: Cambridge University Press.
- Pearl, J. 2014. The Deductive Approach to Causal Inference. *J Causal Inference* **2**(2): 115–129.
- Pincock, C. 2004. A Revealing Flaw in Colyvan’s Indispensability Argument. *Philosophy of Science*, **71**(1): 61–79.
- . 2007. A Role for Mathematics in the Physical Sciences. *Nous* **41**: 253–275.
- . 2012. *Mathematics and Scientific Representation*. Oxford: Oxford University Press.
- Robert, C. P. 2007. *The Bayesian Choice*. 2nd Ed. New York: Springer.
- Royall, R. 1997. *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall.
- Shapiro, S. 1997. *Philosophy of mathematics: structure and ontology*. New York: Oxford University Press.
- . 2004. Foundations of mathematics: metaphysics, epistemology, structure. *Philosophical Quarterly* **54**: 16–37.
- Suárez, M. 2004. An Inferential Conception of Scientific Representation. *Philosophy of Science*, **71**(5): 767–779.
- Suppes, P. 2002. *Representation and Invariance of Scientific Structures*. Stanford: CSLI Publications.
- van Fraassen, B. C. 2008. *Scientific representation*. Oxford: Oxford University Press.

## Notes

<sup>1</sup>Regarding fictionalism, scientific representation is not an essential property (Frigg, 2006, p 50).

<sup>2</sup>For interpretation, see Hughes (1997, p. 332) and Contessa (2007, pp. 57-61).

<sup>3</sup>For inference, see Suarez (2004, p. 770).

<sup>4</sup>For an agent-based account, see Giere (2010, p. 274).

<sup>5</sup>For isomorphism, see van Fraassen (2008, p. 18 and p. 208), and French (2003, p. 1482).

<sup>6</sup>For partial isomorphism, see da Costa and French (2003, Chap. 3).

<sup>7</sup>Bartels (2006, p. 17) says that similarity is a symmetrical relation, and homomorphism is not. But Chakravartty (2010, p. 199) thinks that if there is a homomorphism from  $A$  onto  $B$ , then  $A$  and  $B$  have some structure in common, and are thus similar (indeed, exactly similar) in that respect.

<sup>8</sup>For a review of mathematical structuralism see Nodelman and Zalta (2014), Shapiro (1997, 2004) and Parsons (1990). But Pincock (2007), Bueno and Colyvan (2011), and Batterman (2010) focus on the applicability problem.

<sup>9</sup>According to a purely structural account, as labeled by Pincock (2004), “the existence of an appropriate mapping from a mathematical structure to a physical structure is *sufficient* to fully explain the particular application of the mathematical structure in question” (Bueno and Colyvan, 2011: 345–346, our emphasis). Batterman points out that “Bueno and Colyvan are a bit uncharitable in their sufficiency claim of mapping according to Pincock” (Batterman, 2010: 9). Whether this interpretation is correct or not, in a *purely* structural view, “the applicability of mathematics is *fully* accounted for by structural relations” (Bueno and Colyvan, 2011). In other words, the representational role is primitive and other roles are reduced to representational properties. In the general structural view, the mapping account of the applicability of mathematics is correct as far as it goes, but it does not go far enough — so some modifications are required in the purely structural view of applicability.

<sup>10</sup>According to structuralism, mathematical objects relate to each other and are relational things, while set-theory provides a language to present them.

<sup>11</sup>For more details about the structural and set-theoretic views about models, see da Costa and French (2003: Chapters 2 and 3).

<sup>12</sup>The structural coefficients of equations consist of two groups: parameters and random variables.  $\beta_i$ 's are parameters that are numbers that summarize data for an entire population, but  $\epsilon$  is a random variable, because this coefficient represents factors that are ignored.

<sup>13</sup>In philosophy of science, the predicate of underdetermination is used for both entities and scientific theories. The most famous use of this term is related to the question of the rationality of scientific theories and model selection. The thesis of underdetermination claims that empirical evidence alone cannot set a theory against competing theories, that is, we cannot select a theory only by empirical evidence. The opponents of rationalism, using the thesis of underdetermination, are struggling to rationalize scientific theories. However, the use of underdetermination in this paper, focuses on determining the statistical model by the data set. Our claim is semantic, not epistemological. We want to show that statistical models are determined, but this determination is not realized only by data, they are determined by data and the functions of agent.

<sup>14</sup> The distinction between intrinsic and extrinsic roles in scientific representation are due



to Pincock (2012, Chapter 1). He distinguishes between the mathematical content of the representation itself and the mathematical techniques that were used to arrive at that representation. The mathematics of a representation is called *intrinsic* when it figures in the content of the representation. But there are usually additional *extrinsic* mathematical elements that contribute to the representation beyond those that figure in the content.

<sup>15</sup>The Bayesian approach to statistical inference considers some different criteria to produce and to evaluate an estimator. The difference between classical and Bayesian approaches to statistical inference is out of the scope of this article. According to the Bayesian approach, the parameters are random variables and they are radically unknown. The interested reader is referred to Royal (1997) and Robert (2007).

## Acknowledgments

The authors would like to thank Professor O. Bueno and Professor P. Bandyopadhyay and two anonymous referees for valuable discussions and useful comments on an earlier version of paper. They are grateful to Dr. R. Ramezan for her useful comments/corrections on the pre-final manuscript.