

# Interpreting Infinite Numbers

Carles M. Cuadras

Department of Statistics, University of Barcelona, 08028 Barcelona, Spain; ccuadras@ub.edu

**Abstract:** The mathematical concept of infinity, in the sense of Cantor, is rather far from applied mathematics and statistics. These fields can be linked. We comment on the properties of infinite numbers and relate them to some operations with random variables. The existence of statistical parametric models can be studied in terms of cardinal numbers. Some probabilistic interpretations of Gödel's theorem, Turing's halting problem, and the Banach-Tarski paradox are commented upon, as well as the axiom of choice and the continuum hypothesis. We use a basic but sufficient mathematical level.

**Keywords:** transfinite arithmetic; infinite cardinals; Cantor theorem; statistical models; bivariate distributions; Gödel theorem; Turing halting problem; Banach-Tarski paradox; continuum hypothesis

**MSC:** 03D10; 03E25; 11A99; 60E05

## 1. Introduction

The concept of infinite quantity is very old, and is often considered as a negative sentence. Though it has been around for a long time, the notion of infinite quantity is still often viewed negatively. In addition, limitless power and perfection were connected with the divine.

The distinction between finite and infinite started in ancient Greece. Zenon denied the infinity through paradoxes. Aristotle considered the infinity as something potential, never actual, an approach accepted for a long time (Gauss said that infinity was only a *façon de parler*). The infinity, as a number or object, was unaccepted, as it does not exist. Thomas d'Aquino in *Summa Theologica*, and Galileo Galilei in *Discursi e dimostrazione matematiche*, also denied the infinity. The night darkness goes against the paradox of Kepler-Olbers on the possible infinite sky. In contrast, Agustin de Hipona and Spinoza (*Ethica*) used the idea of the absolutely infinite.

Fontenelle in *Eléments de la Géométrie de l'infini* (published in 1727) refers to the infinity as an actual object, which can be operated algebraically. Bolzano in *Paradoxien des Unenlidchen* (published in 1854) studied some paradoxes on the infinity and defended the concept of actual infinity. The mathematical infinity was an academic leivmotiv; for instance, the discussion of Domenech y Estapa [1] on geometrical absurdities engendered by the interpretation of the mathematical infinity. He does not agree to define the straight line as a limit of a circumference whose diameter increases (so Nicholas of Cusa illustrated the meaning of infinity), as this definition may identify the extremes, i.e., the positive and the negative infinities.

The rigorous mathematical definition and study of infinite quantities starts with Georg Cantor. Galileo's argument against the infinity, since if accepted, then natural and even numbers have the same size, was used by Cantor to define the infinite cardinality of a set. In fact, Gregory of Rimini (about 1350) had advanced that a subpart of something infinite could be equivalent to the whole.

Cantor found, compared, and operated infinities of different sizes, and despite Kronecker's and Poincaré's opposition, he imposed his theory and founded the so-called Transfinite Arithmetic, today accepted everywhere. It was even suggested by the academic Rodriguez-Salinas [2] that *Homo Sapiens*, after Cantor, should be *Homo Trans-Sapiens*.



**Citation:** Cuadras, C.M. Interpreting Infinite Numbers. *Axioms* **2023**, *12*, 314. <https://doi.org/10.3390/axioms12030314>

Academic Editors: Hans J. Haubold and Mircea Merca

Received: 19 January 2023

Revised: 9 March 2023

Accepted: 16 March 2023

Published: 22 March 2023



**Copyright:** © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

This article deals with the following related concepts:

1. The infinite cardinality (size) of sets.
2. Operating random variables and transfinite arithmetic.
3. Gödel’s theorem and parametric statistics.
4. The cardinality of some bivariate distributions.
5. The axiom of choice and the continuum hypothesis.
6. The probabilistic analogy of the Banach-Tarski theorem.

Statistics is a very wide subject, and most statisticians do not know some topics on infinite cardinals. The aim of this article is to show, in a basic way, these concepts, and to establish a parallelism between transfinite arithmetic and some operations with random variables, as well as the dimensionality of statistical models and other concepts in the field of statistics and probability.

For a rigorous and coherent introduction to the set theory, see [3]. For a history of the concept of infinity, see chapter 2 in [4]. See [5,6] for a description of the infinities of different sizes. The biography and contributions of Cantor are well explained in [7].

## 2. Essentials on Cardinals and Transfinite Arithmetic

Any non-empty set  $\mathbb{A}$  has a cardinal number denoted  $\#\mathbb{A}$ . If the elements of  $\mathbb{A}$  are in one-to-one correspondence with the elements of a proper subset of  $\mathbb{A}$ , then the cardinal  $\#\mathbb{A}$  is infinite.

Accordingly, we can give the cardinal number  $\#\emptyset = 0$  to the empty set, as well as  $\#\{a_1, \dots, a_k\} = k$  if  $k$  is a natural number and  $a_i \neq a_j$  for  $i \neq j$ . The most interesting infinite cardinals are  $\#\mathbb{N}$  and  $\#\mathbb{R}$ , where  $\mathbb{N} = \{0, 1, \dots, n, \dots\}$  is the infinite countable set—called denumerable—of the natural numbers, and  $\mathbb{R}$  is the uncountable set of the real numbers. The notations for these infinite cardinals are:

$$\#\mathbb{N} = \aleph_0, \quad \#\mathbb{R} = \mathfrak{c}.$$

The infinite cardinal  $\mathfrak{c}$  is the power of the continuum. It is well known in set theory that

$$\aleph_0 < \mathfrak{c} = 2^{\aleph_0} = \#\mathcal{P}(\mathbb{N}),$$

where  $\mathcal{P}(\mathbb{N})$  is the family of all subsets of  $\mathbb{N}$ .

The inequality  $\#\mathbb{N} < \#\mathcal{P}(\mathbb{N})$  is a particular case of Cantor’s theorem: Any non-empty set  $\mathbb{A}$  satisfies  $\#\mathbb{A} < \#\mathcal{P}(\mathbb{A})$  [3].

If  $\leftrightarrow$  stands for the one-to-one correspondence between two sets, then the equality, inclusion, and two elementary operations with cardinals are:

Equality	$\#\mathbb{A} = \#\mathbb{B}$	if	$\mathbb{A} \leftrightarrow \mathbb{B}$
Inequality	$\#\mathbb{A} \leq \#\mathbb{B}$	if	$\mathbb{A} \subseteq \mathbb{B}$
Sum	$\#\mathbb{A} + \#\mathbb{B} = \#(\mathbb{A} \cup \mathbb{B})$	if	$\mathbb{A} \cap \mathbb{B} = \emptyset$
Product	$\#\mathbb{A} \times \#\mathbb{B} = \#(\mathbb{A} \times \mathbb{B})$	if	$\mathbb{A}, \mathbb{B}$ are general sets.

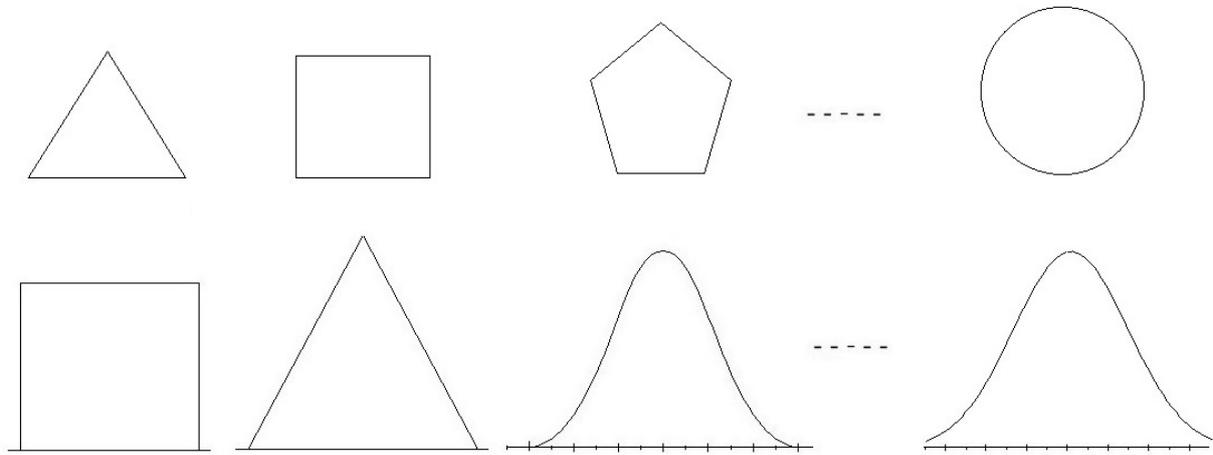
For instance,  $\mathbb{N} \leftrightarrow \mathbb{Q}$ , where  $\mathbb{Q}$  is the set of rational numbers, so  $\mathbb{Q}$  is infinite countable. In transfinite or cardinal arithmetic, the following rules hold, where  $n$  is any finite natural number:

$$\begin{aligned} n + \aleph_0 &= \aleph_0, & \aleph_0 + \aleph_0 &= \aleph_0, \\ \aleph_0 \times \aleph_0 &= \aleph_0, & \aleph_0 + \mathfrak{c} &= \mathfrak{c}, \\ \mathfrak{c} + \mathfrak{c} &= \mathfrak{c}, & \mathfrak{c} \times \mathfrak{c} &= \mathfrak{c}. \end{aligned}$$

Thus,  $\#\mathbb{R} = \#\mathbb{R}^2 = \#\mathbb{R}^3 = \mathfrak{c}$ . Hence, the real line, the Euclidean plane, and the three-dimensional space have the same uncountable cardinality  $\mathfrak{c}$ , the power of the continuum.

If we consider a regular polygon  $\mathbb{V}_n$  with  $n$  vertices, circumscribed in a circumference, the limit, as  $n$  increases, is  $\mathbb{V}_\infty$ , the circumference with infinite countable points. Although  $\aleph_0 < \mathfrak{c}$ , it is impossible to distinguish geometrically between  $\mathbb{V}_\infty$  (cardinality  $\aleph_0$ ) and

the complete continuous circumference (cardinality  $\mathfrak{c}$ ); see Appendix A. The central limit theorem also illustrates the meaning of infinity. Both examples (geometric and probabilistic) are depicted in Figure 1. Another paradigm, described by Nicholas of Cusa (about 1445), of the infinite cardinality  $\mathfrak{c}$  is the straight line, thought of as the limit of a circle whose diameter increases constantly.



**Figure 1.** Two examples of infinity. The limit of the vertices of a regular polygon tends to the countable circumference. If we add one, two, three, etc., independent uniform random variables, the limit is the normal distribution.

### 3. Operating Poisson and Gaussian Distributions

Let us indicate  $\text{Poisson}(n)$ , the Poisson random variable with parameter  $n$  (both the expectation and variance are  $n$ ), where  $n$  is fixed and  $n \in \mathbb{N}$ . If  $\text{Poisson}(m)_1$  and  $\text{Poisson}(n)_2$  are independent, then the sum of both variables is also  $\text{Poisson}(m + n)$ .

When  $n$  is ranging in  $\mathbb{N}$ , this family of discrete random variables is indicated by  $[\text{Poisson}(n)]$ .

Similarly, if  $\text{Normal}(\mu, \sigma^2)$  is the normal random variable with mean  $\mu$  and variance  $\sigma^2$ , then  $[\text{Normal}]$  is the family of all normal random variables. The sum of independent normal variables is also normal. Accordingly, the bivariate normal family can be indicated by  $[\text{Normal}, \text{Normal}]$ , and all continuous univariate and bivariate variables are indicated by  $[\text{All r.v.}]$  and  $[\text{All r.v.}, \text{All r.v.}]$ , respectively.

It is clear that  $\#[\text{Poisson}(n)] = \aleph_0$  and  $\#[\text{Normal}] = \mathfrak{c}$ . As  $[\text{All r.v.}]$  is described by the all positive curves (continuous or not continuous) with area 1, and it is well-known that the cardinality of this family is  $\#[\text{All}] = 2^{\mathfrak{c}} = \#\mathcal{P}(\mathbb{R})$ . In addition, if  $[\text{Poisson}(n)]_1$  and  $[\text{Poisson}(n)]_2$  are independent, as each variable is an application from the sample population to  $\mathbb{N}$ , and we can interpret that there are no common variables, so (ignoring the constant 0) the intersection of both sets is empty. By  $\mathbb{Q} + [\text{Normal}]$ , we indicate all the possible sums  $q + X$ , where  $q$  is rational and  $X$  is random normal.

With these notations, since  $\aleph_0 = \#\mathbb{Q}$ ,  $\aleph_0 = \#[\text{Poisson}(n)]$ ,  $\mathfrak{c} = \#[\text{Normal}]$ , etc., we have a relation between some operations with random variables and cardinal numbers; see Table 1.

**Table 1.** Relatiing some operations with cardinals and random variables.

Random Variables	Cardinals
$[\text{Poisson}(n)]_1 + [\text{Poisson}(n)]_2 = [\text{Poisson}(n)]$	$\aleph_0 + \aleph_0 = \aleph_0$
$\mathbb{Q} + [\text{Normal}] = [\text{Normal}]$	$\aleph_0 + \mathfrak{c} = \mathfrak{c}$
$[\text{Normal}]_1 + [\text{Normal}]_2 = [\text{Normal}]$	$\mathfrak{c} + \mathfrak{c} = \mathfrak{c}$
$[\text{Normal}] + [\text{All r.v.}] = [\text{All r.v.}]$	$\mathfrak{c} + 2^{\mathfrak{c}} = 2^{\mathfrak{c}}$
$[\text{Normal}, \text{Normal}] = [\text{Bivariate normal}]$	$\mathfrak{c} \times \mathfrak{c} = \mathfrak{c}$
$[\text{All r.v.}, \text{All r.v.}] = [\text{Bivariate all}]$	$2^{\mathfrak{c}} \times 2^{\mathfrak{c}} = 2^{\mathfrak{c}}$

#### 4. Gödel's Theorem

Bolzano had proven that the number of mathematical propositions is infinite. Gödel proved that there are also infinite propositions on the integer numbers, which cannot be reduced to a finite number of axioms. This is a special case of the famous theorem enunciated by Kurt Gödel in 1931. This theorem says that in a logical system based on axioms and containing the Arithmetic, we have some propositions which cannot be proven. These propositions are undecidable: they can be accepted or rejected. The acceptance should be considered as a new axiom. An example is the fifth postulate of Euclides on parallel lines, which cannot be proven as a consequence of the other four axioms. Another example is the hypothesis of the continuum (see below), posed by Cantor: there does not exist a set  $\mathbb{G}$  with cardinality  $\mathfrak{g}$  between the infinite countable (denumerable) and the uncountable; i.e., it is not possible to find  $\mathfrak{g}$  such that  $\aleph_0 < \mathfrak{g} < \mathfrak{c}$ . However, this hypothesis can be accepted or rejected, without prejudice to the Arithmetic.

Let us understand Gödel's theorem with a broad view, and interpret this theorem from a statistical perspective, namely, in terms of observable events, parametric models, inference, etc. A statistical model is a family of probability distributions parametrized by a parameter  $\theta$ . The Poisson distribution with parameter  $\lambda$  and the normal distribution with parameters  $\mu, \sigma^2$ , are two examples, with a parameter space of dimensions one and two, respectively. If  $F_\theta$  is a statistical model, with  $\theta$  belonging to a region of  $\mathbb{R}^k$ , with positive hypervolume, then clearly,  $\#F_\theta = \mathfrak{c}$ .

Question: Does there exist a universal parametrization covering all distributions? The answer is no. Let us consider all probability distributions with the same support. This set has cardinal  $2^{\mathfrak{c}}$ . If we suppose all of the models described by  $F_\theta$ , since  $\#F_\theta = \mathfrak{c} < 2^{\mathfrak{c}}$ , there are many distributions out of this family. Thus, a parametrization can not cover all distributions. See an analytic proof in Appendix A.

As well, as a logical system is incomplete, since some propositions are undecidable, a statistical model, as wide as possible, is also incomplete, since some probability distributions are not contained in the model. This justifies the so-called non-parametric statistics, an approach that makes inferences on functional expressions, considering the whole set of distributions, but avoiding the use of parameters.

Another analogy is as follows. Given a sample of size  $n$ , to perform an inference on the mean  $\mu$  in the normal model  $N(\mu, \sigma_0^2)$ , with the variance  $\sigma_0^2$  being known, the statistic  $\sum x_i$  is sufficient and complete. However,  $\sum x_i$  is not sufficient and complete to perform an inference on  $N(\mu, \sigma^2)$ , where both parameters are unknown. We need two statistics, namely  $\sum x_i$  and  $\sum x_i^2$ .

Similarly, let us consider the Poisson distribution whose support is  $\mathbb{N}$ . Theoretically, we can observe any subset of  $\mathbb{N}$ . However,  $\#\mathcal{P}(\mathbb{N}) = \mathfrak{c}$ , and the cardinality is too large. Thus, some events can be observed, e.g., "a value  $k$  is even number", but many other events cannot be enunciated using our limited language; hence, these events cannot be observed.

#### 5. Turing's Halting Problem

This problem was posed by Allan Turing in 1935. There is no general algorithm that is capable of determining whether or not a computer program will finish running. This is another example of an undecidable proposition in Gödel's sense. From a statistical perspective, Chaitin's approach [8] is quite interesting. A summary is next given.

Let us consider the Cantor space of all binary infinite sequences. A computer program is a subset of this space. Let  $P$  be the set of halting programs. If  $p$  is a halting program of  $|p|$  bits, the probability that a randomly chosen binary sequence of length  $|p|$  coincides with  $p$  is  $2^{-|p|}$ , provided that we generate this sequence of 0s and 1s independently with the same probability  $1/2$ . In general, when choosing a program at random, the probability of achieving a halting program is

$$\Omega = \sum_{p \in P} 2^{-|p|}.$$

That is,  $\Omega = \sum_n a_n/2^n$ , where  $a_n$  is the number of halting programs of  $n$  bits, taking into account that if the program with sequence  $\mathbf{b} = b_1 \cdots b_{n-1}$  of  $(n - 1)$  bits halts, then the  $a_n$  programs of  $n$  bits cannot begin with  $\mathbf{b}$ . This constrains  $a_n$ .

If  $\Omega = 0$ , no program halts. If  $\Omega = 0.16$ , then there is 16% halt. If  $\Omega = 1$ , all programs halt. However,  $\Omega$  is not fixed, but is random. We need an algorithm of  $n$  bits to determine the first  $n$  bits of  $\Omega$ . This probability behaves as if it were randomly generated, just as explained above. There is not a law reducing the computation of  $\Omega$ . None, some, or all programs finish running. Consequently, the halting problem is undecidable.

A short, informal proof of the halting problem, based on cardinal comparison, is given in Appendix A.

### 6. Cardinality of Bivariate Distributions

A bivariate cumulative distribution function  $H(x, y)$  of two random variables  $X, Y$ , with univariate marginal distributions  $F(x), G(y)$ , is usually described by a parametric model. Suppose that the range of the variables  $X, Y$  are the intervals  $[a, b], [c, d]$ . If the degree of dependence between the random variables is quantified by means of a unique parameter, the cardinality of  $H$  is 1. However, in general,  $H$  admits the canonical decomposition

$$dH(x, y) = dF(x)dG(y) \left[ 1 + \sum_{n \geq 1} \rho_n a_n(x) b_n(y) \right],$$

where  $\rho_n$  are the canonical correlations, all positive, and  $a_n(x), b_n(y)$  are the canonical functions. Then, the cardinality of  $H(x, y)$  is  $\#\{\rho_n\}$ , the power of the set of canonical correlations. This number, also called rank of  $H$ , is the dimensionality from a geometrical point of view, related to the so-called chi-squared distance between two observations  $x, x'$  of the variable  $X$ ,

$$\delta^2(x, x') = \int_c^d \left[ \frac{dH(x, y)}{dF(x)dG(y)} - \frac{dH(x', y)}{dF(x')dG(y)} \right]^2 dG(y).$$

See [9] for a continuous correspondence analysis interpretation. Some (finite and infinite) cardinalities of  $H(x, y)$  are reported in Table 2.

**Table 2.** Cardinality of some bivariate distributions, defined as the size of the set of canonical correlations. Note the power of the continuum cardinality of the last distribution.

Bivariate Distributions Where $F(x), G(y)$ Are Indicated by $F, G$	Cardinality
$FG$ (stochastic independence)	0
$FG[1 + \theta(1 - F)(1 - G)]$	1
$FG + \theta_1(F - F^2)(G - G^2) + \theta_2(2F^2 - F)(1 - F)(2G^2 - G)(1 - G)$	2
$FG/[1 - \theta(1 - F)(1 - G)]$	$\aleph_0$
$\min\{F, G\}^\theta (FG)^{1-\theta}$	$\mathbf{c}$

In all cases, we can consider  $0 \leq \theta \leq 1$ . Thus,  $H(x, y) = F(x)G(y)$ , corresponding to the stochastic independence, the cardinality being 0 because of the absence of positive canonical correlations. The cardinality of the second model is 1 because the only canonical correlation is  $\rho_1 = \theta/3$ .

Any distribution with infinite cardinality, e.g.,  $FG/[1 - \theta(1 - F)(1 - G)]$ , can be approximated by another one with finite cardinality [10].

It is worth noting the power of the continuum cardinality of the fifth model (see Table 2), defined and studied in [11,12]. In the uniform marginal case,  $F(x) = x, G(y) = y$ ,  $H(x, y) = \min\{x, y\}^\theta (xy)^{1-\theta}$ , the set of canonical correlations is the function  $\theta\rho^{1-\theta}$ , where  $0 \leq \rho \leq 1$ . This continuous function ranges between 0 and  $\theta$ . Thus, if  $\theta > 0$ , the power of the set of canonical correlations is  $\mathbf{c} = \#\{\theta\rho^{1-\theta}\}$ .

Another continuous correlation model is  $F(x)G(y)[1 - \theta \ln \max\{F(x), G(y)\}]$ . In the uniform case with  $\theta > 0$ , the cardinality of this distribution is  $\mathfrak{c} = \#\{\theta\rho\}$ , which is also the power of the continuum.

These two distributions admit an integral expansion, instead of a series expansion, and the set of canonical correlations is not countable, but continuous. This transition from countable to continuous uncountable cardinality agrees with the hypothesis explained in the next section.

## 7. The Continuum Hypothesis

This hypothesis stated by Cantor in 1878 [13] says that there is no set with cardinality between the infinite countable  $\aleph_0$  and the uncountable  $\mathfrak{c}$ . It can be expressed as

$$\aleph_0 < \mathfrak{c} = \aleph_1,$$

where  $\aleph_1$  is the next immediate infinity after  $\aleph_0$ . According to Cantor,  $\aleph_1$  exists and is  $\mathfrak{c}$ . Nonetheless, this is considered undecidable after the results by Gödel and Cohen in 1940 and 1963, respectively. Nowadays, we should write

$$\aleph_0 < \mathfrak{c} \leq \aleph_1.$$

It is known of Cantor's futile attempts in showing this hypothesis, which was included as an unsolved question in the list of 23 problems posed by Hilbert in 1900. Gödel showed that the acceptance of this hypothesis is not contradictory, whereas Cohen showed that it can be considered false and that it is not contradictory either. Thus, the continuum hypothesis is independent of the axioms of the Zermelo-Fraenkel theory of sets [3].

It is quite surprising that some authors [14–17], many years later, ignored this essential difficulty and mentioned this problem as interesting but not solved yet. The first widespread reference to Cohen and the independence of the continuum hypothesis appeared in a textbook in 1965 [18]. In the Spanish literature on this topic, [19,20] are the first books paying attention to the independence of this hypothesis. Many years later, this hypothesis still has interest [6,21].

Indeed, in probability and statistics, this hypothesis is implicitly accepted. In general, only the probabilities of subsets of  $\mathbb{N}$  are considered under discrete models (such as the Poisson distribution). In addition, only Borel sets under continuous models (such as the Gaussian distribution), are taken into account. Recall that a Borel set is obtained by joining the isolated points and intervals of the real line. To define the probability of other sets is not considered, as they are unobservable. For example, accepting the axiom of choice (see below), we can construct non-measurable sets, it, however, being impossible after an experience to decide on the presence or absence of any of these sets. That is, given a value  $x$  of the random variable, it is impossible to decide whether or not  $x$  belongs to a non-measurable set of  $\mathbb{R}$ . See Appendix A.

## 8. Axiom of Choice

This axiom (stated by Zermelo in 1904) postulates that given a family of non-empty sets, we can choose an element of each set and construct a new set with these elements. Two examples are as follows. If we consider all of the circumferences centered in the origin  $(0,0)$ , we do not need this axiom, as we can choose a point of each circumference, e.g., the point on the right cutting the horizontal axis. However, if we consider the family of all closed curves in the plane, we need the axiom of choice to choose a point of each curve.

Some properties in algebra, geometry, topology, and analysis depend on this axiom. One functional analysis application is to prove the null norm of the eigenfunctions of a kernel with respect to another kernel, with both being related to the last bivariate distribution given in Table 2 [22].

With some imagination, we can establish a comparison between this axiom (choosing an element of each set) and the Bayes theorem (the probability that an observation belongs to each set or cause).

However, accepting the axiom of choice, we can prove the existence of a mathematical object, but we are not able to actually construct that object. This may be a trouble. For instance, any vector space has a basis. However, there are vector spaces, e.g.,  $(\mathbb{Q}, \mathbb{R})$ , such that the basis is unknown. This also happens with the vector space of all random variables with support in  $\mathbb{R}$ . In addition, accepting this axiom, some subsets of  $\mathbb{R}$  lack length (Lebesgue measure) or probability (see the Appendix A). Another anomaly is next commented upon.

**9. Banach-Tarski’s Theorem**

Let us suppose that  $\mathbf{B}$  is a solid ball. This theorem asserts that we can divide  $\mathbf{B}$  into  $m + n$  non-overlapping parts, i.e.,

$$\mathbf{B} = A_1 \cup \dots \cup A_m \cup B_1 \cup \dots \cup B_n,$$

and, after isometric transformations  $A_i \rightarrow A_i^*, B_j \rightarrow B_j^*$ , we can assemble these parts to yield two balls:

$$\mathbf{B} = A_1^* \cup \dots \cup A_m^* \text{ and also } \mathbf{B} = B_1^* \cup \dots \cup B_n^*.$$

Accordingly, this may be expressed as the paradoxical equidecomposition

$$\mathbf{B} = \mathbf{B}_1 \cup \mathbf{B}_2 \text{ with } \mathbf{B}_1 \cap \mathbf{B}_2 = \emptyset,$$

being that the volume is the same:  $|\mathbf{B}| = |\mathbf{B}_1| = |\mathbf{B}_2|$ . Thus, the initial ball can be duplicated. Notice that the parts  $A_i, B_j$ , are non-measurable subsets of  $\mathbf{B}$ . In fact, it is necessary to accept the axiom of choice to prove this surprising result.

There is a probabilistic analogy. Suppose that  $X$  is a normal random variable with a mean of 0 and a variance of  $\sigma^2$ . Then, we can decompose  $X$  as a sum, i.e., to express  $X = X_1 + X_2$ , the sum of two independent normal variables with mean 0. If  $[X]$  stands for the class of variables  $\alpha X$ , where  $\alpha$  is a real parameter, and similarly  $[X_1], [X_2]$ , then

$$[X] = [X_1] + [X_2] \text{ with } [X_1] \cap [X_2] = 0,$$

being  $[X] \stackrel{sd}{=} [X_1] \stackrel{sd}{=} [X_2]$ , where “ $\stackrel{sd}{=}$ ” means “same distribution”. This is so because  $\alpha X = \alpha X_1 + \alpha X_2$  and  $[X_1], [X_2]$  are independent; hence, they do not contain common variables, except for the constant 0. In spite of this lack of coincidence, these sets contain exactly the same family of normal distributions. Thus, in some sense,  $[X]$  can be duplicated.

Of course, the sum of non-coincident sets of random variables is not the union of disjoint sets, but the analogy is clear; see Table 3.

**Table 3.** Analogy between dividing a ball into two balls of the same volume, and decomposing a class of normal variables into the sum of two independent classes with the same size.

Banach-Tarski $\mathbf{B} = \mathbf{B}_1 \cup \mathbf{B}_2$	Disjoint Balls $\mathbf{B}_1 \cap \mathbf{B}_2 = \emptyset$	Same Volume $ \mathbf{B}  =  \mathbf{B}_1  =  \mathbf{B}_2 $	$\mathbf{B}$ Is Divided into Non-Measurable Parts
Normal class $[X] = [X_1] + [X_2]$	No coincident $[X_1] \cap [X_2] = 0$	Same distributions $[X] \stackrel{sd}{=} [X_1] \stackrel{sd}{=} [X_2]$	$X$ is sum of r.v.’s with unknown distribution

Furthermore, taking into account the central limit theorem, let us admit that  $X$  normal can be interpreted as the sum of a series of independent random variables, whose distributions are unknown. Namely,  $X = \sum_{i \geq 1} X_i$ , where the convergence is in law (the standardization is omitted). Then,

$$X = \sum_{i \text{ odd}} X_i + \sum_{i \text{ even}} X_i,$$

and  $X$  is the sum of two independent normal random variables. Note that “unknown distribution” would correspond to “non-measurable subset” in the above Banach-Tarski decomposition.

Finally, removing the constant 0, we can consider  $[X]$ ,  $[X_1]$ ,  $[X_2]$  projective spaces, so that we have another analogy in terms of projective geometry.

## 10. Discussion, Conclusions, and Future Work

The old pamphlet [1] is an example of how a mathematician and also architect—for a long time, the careers of Mathematics and Architecture overlapped in some universities—perceived contradictions between descriptive geometry and pure mathematics. However, most subjects of Mathematics and Statistics can be related. For instance, if  $\mathbf{E}$  is the vector space generated by random variables, the dual space  $\mathbf{E}^*$  can be interpreted as the population (set of individuals), since if  $\omega$  is an individual and  $Y$  is a random variable, we can associate the real number  $Y(\omega)$  to the pair  $(\omega, Y)$  [23]. The differential geometry can be used to define geodesic distances between the parameters of a statistical model [24,25]. The study of bivariate exchangeable distributions can be performed using functional analysis [22]. There are more examples linking different fields.

We have proven that some concepts and properties of Probability and Statistics can be useful for understanding and interpreting the main properties of the infinite cardinals.

Several proposals for future research are:

- (1) Analytic geometry. The equation  $x^{2n} + y^{2n} = 1$  defines a closed curve tending to a square as  $n \rightarrow \infty$ . Study the implicit equation of other regular polygons (see Figure 1) in the same way.
- (2) Inference. Given a sample of size  $n$ , the statistics  $\sum x_i$ ,  $\sum x_i^2$  are sufficient to perform inference on the normal model. Then, we may explore the sufficiency of  $\sum x_i$ ,  $\sum x_i^2, \dots, \sum x_i^k$  under a perspective similar to Gödel’s theorem. That is, to study what kind of inference is “undecidable” on a specific model. Note that we can make a non-parametric inference if  $k = n$ .
- (3) Bivariate distributions. We have passed from distributions with countable cardinality to uncountable continuous cardinality. Does it give enough evidence for accepting the hypothesis of the continuum?
- (4) Banach-Tarski theorem. Removing the constant 0 and interpret  $[X] = [X_1] + [X_2]$  as a decomposition of projective spaces, which can be generalized to a higher dimension. In addition, for  $k > 2$ , study the comparison between dividing a ball into  $k$  balls and decomposing a normal random variable into the sum of  $k$  independent random variables.
- (5) Statistical models. The cardinality of the Poisson (with parameter being a natural number) and the normal models are  $\aleph_0$  and  $\mathfrak{c}$ , respectively. Are there parametric models with cardinalities larger than  $\mathfrak{c}$ ?

**Funding:** This research received no external funding

**Data Availability Statement:** Data is contained within the article.

**Conflicts of Interest:** The author declares no conflict of interest.

## Appendix A

Here, we provide some direct and elementary proofs, avoiding the use of special functions and too-formal concepts.

- (1) The limiting circumference of the regular circumscribed polygons has cardinality  $\aleph_0$ .

**Proof.** We can identify the triangle and the square with the sets  $\mathbb{V}_3 = \{0, 1/3, 2/3\}$  and  $\mathbb{V}_4 = \{0, 1/4, 2/4, 3/4\}$ . In general, any regular polygon of  $n$  vertices can be identified with

$$\mathbb{V}_n = \{0, 1/n, 2/n, \dots, (n-1)/n\}.$$

Accordingly, the limiting circumference of these circumscribed polygons is  $\mathbb{V}_\infty$ . Let us consider the set  $\mathbb{W} = \bigcup_{n \geq 1} \mathbb{V}_n$ . That is, skipping repeated numbers (e.g.,  $1/2 = 2/4$ ),

$$\mathbb{W} = \{0, 1/2, 1/3, 2/3, 1/4, 3/4, 1/5, 2/5, 3/5, 4/5, 1/6, 5/6, \dots\}.$$

Then,  $\mathbb{V}_\infty \subset \mathbb{W} \subset \mathbb{Q}$  shows that  $\mathbb{V}_\infty$  is countable. Note that the complete circumference is continuous and has cardinality  $\mathfrak{c}$ , as it can be identified with the interval  $[0, 1]$ . This proof is related to Figure 1. An alternative proof follows by using polar coordinates.  $\square$

- (2) If the random variable  $X$  has finite mean and variance, and takes values in  $\mathbb{N}$  and  $X = X_1 + X_2$ , where  $X_1$  and  $X_2$  are independent with the same distribution of  $X$  (except the mean and variance), then  $X$  is Poisson.

**Proof.**  $E(X_1) = E(X_2) = \lambda$  and  $X = X_1 + X_2$  implies  $E(X) = 2\lambda$ . Then,

$$p(0; 2\lambda) = P(X = 0) = P(X_1 = 0)P(X_2 = 0) = p(0; \lambda)^2$$

implies  $p(0; \lambda) = a^{-\lambda}$  with  $a > 1$ . Now, suppose  $p(k; \lambda) = a^{-\lambda} \lambda^k / k!$ , which is true for  $k = 0$ . Then,

$$\begin{aligned} p(k+1; \lambda) &= \sum_{i=0}^{k+1} p(i; \lambda)p(k+1-i; \lambda) \\ &= 2p(0; \lambda)p(k+1; \lambda) + \sum_{i=1}^k p(i; \lambda)p(k+1-i; \lambda) \\ &= 2a^{-\lambda}p(k+1; \lambda) + a^{-2\lambda} \sum_{i=1}^k \frac{\lambda^i \lambda^{k+1-i}}{i!(k+1-i)!} \text{ (by induction).} \end{aligned}$$

But,

$$2^{k+1} = \sum_{i=0}^{k+1} \frac{(k+1)!}{i!(k+1-i)!} \text{ implies } \sum_{i=1}^k \frac{1}{i!(k+1-i)!} = (2^{k+1} - 2)/(k+1)!$$

Therefore,

$$p(k+1; 2\lambda) = 2a^{-\lambda}p(k+1; \lambda) + a^{-2\lambda}[\lambda^{k+1}/(k+1)!(2^{k+1} - 2)].$$

This equation is satisfied for  $p(k+1; 2\lambda) = a^{-2\lambda}(2\lambda)^{k+1}/(k+1)!$ . Since  $p(k; 2\lambda)$  is a probability density, we must take  $a = e$ .  $\square$

- (3) The Cramer-Levy theorem says: if a normal random variable  $X$  can be decomposed as the sum  $X = X_1 + X_2$  of two independent random variables, then  $X_1$  and  $X_2$  are also normal. We prove a more general result.

Suppose that  $X$  has a finite mean and variance  $\mu, \sigma^2$ , and take values in  $\mathbb{R}$ . Assume that  $X = aX_1 + bX_2$ , where  $X_1, X_2$  are independent and have the same distribution as  $X$ . Then,  $X$  is normal (Gaussian).

**Proof.** If  $X = aX_1 + bX_2$ , then  $\mu = (a + b)\mu$ ,  $\sigma^2 = (a^2 + b^2)\sigma^2$ , so  $\mu = 0$  and  $a^2 + b^2 = 1$ . For the sake of simplicity, let us take  $a = b = 1/\sqrt{2}$ . Then,  $X = a(X_1 + X_2)$ , so  $X \stackrel{sd}{=} a(X_1 + X_2)$ , where “ $\stackrel{sd}{=}$ ” means “same distribution”. Consider  $X_1, X_2, \dots, X_n$ , independent random variables distributed as  $X$ . Then,

$$X \stackrel{sd}{=} a(X_1 + X_2), \quad X \stackrel{sd}{=} a[a(X_1 + X_2) + X_3].$$

In general, for  $n > 1$ ,

$$X \stackrel{sd}{=} S_n = a^{n-1}(X_1 + X_2) + \dots + a^{n-1}X_3 + \dots + aX_n.$$

From the central limit theorem,  $S_n$  converges in law to the normal distribution. Since  $X \stackrel{sd}{=} S_n$ , for all  $n > 1$ , the distribution of  $X$  must be normal.  $\square$

- (4) We prove that a parametrization does not exist for the whole set of univariate distributions. Consider the cdfs (cumulative distribution functions) of random variables, taking values on the interval  $[0, 1]$ . Let us restrict the parametrization to the set of cdfs  $F$ , such that  $F(x) > x$ . Suppose that a parametrization is possible and we write a member of this class as  $F_\theta$ , where  $0 \leq \theta \leq 1$ . We say that  $F_\theta$  is a regular parametrization if the derivative  $F'_x(x)$  exists and satisfies  $m = \inf F'_x(x) > -\infty$ .

Define  $G(x) = \alpha F_x(x) + (1 - \alpha)x$ . Clearly,  $G(0) = 0$ ,  $G(1) = 1$ ,  $G(x) > x$ , for any  $\alpha$ . Suppose  $m \neq 1$  and take  $\alpha = 1/(m - 1)$ . Then,  $G'(x)$  exists and

$$\inf G'(x) = m/(m - 1) + 1 - 1/(m - 1) = 0.$$

Thus,  $G(x)$  is a cdf belonging to the class defined above. Hence,  $G = F_{\theta_0}$  for some  $\theta_0$ , and we have  $G(\theta_0) = F_{\theta_0}(\theta_0)$ . However, the equation

$$F_{\theta_0}(\theta_0) = \alpha F_{\theta_0}(\theta_0) + (1 - \alpha)\theta_0 = \theta_0.$$

implies  $F_{\theta_0}(\theta_0) = \theta_0$ . This is contradictory, so  $G$  is not a member of this parametric class. If  $m = 1$ , we may take  $\alpha = 1/2$  and run into the same contradiction.

If we consider a multiparameter  $(\theta_1, \dots, \theta_n)$ , the proof is similar, taking a common value  $\theta_0$  for the  $n$  parameters.

- (5) Short informal and indirect proof of Turing’s halting problem. Let  $p$  be a binary sequence corresponding to a halting program. The size  $|p|$  (number of bits) is finite, otherwise  $p$  will not halt. Suppose that  $t$  is an algorithm that can determine whether a computer program halts or not. Let  $P_t$  be the set of halting programs controlled by  $t$ . It is readily proven that  $P_t$  is infinite countable. Consider the family  $\mathbf{P}_t = \mathcal{P}(P_t)$  of all subsets of  $P_t$ . If  $p, p' \in P_t$ , then  $\{p, p'\} \in \mathbf{P}_t$ . Indicating by  $s$  a suitable concatenating sentence (e.g., if  $p$  ends then  $p'$  starts), we may consider the program with the binary sequence  $psp'$ . We can similarly generate many other halting programs. Consider the subfamily  $\mathbf{P}_t^f$  of all finite subsets of  $P_t$ . We have  $\#P_t \leq \#\mathbf{P}_t^f$ , and from Cantor’s theorem,  $\#P_t < \#\mathbf{P}_t$ . Thus,  $\#P_t \leq \#\mathbf{P}_t^f \leq \#\mathbf{P}_t$ . However, to determine an intermediate cardinal between  $\#P_t$  and  $\#\mathbf{P}_t$  is undecidable (continuum hypothesis), so that controlling all halting programs is impossible.

- (6) If we accept the axiom of choice, there are non-measurable sets.

**Proof.** Let  $\mathbb{D}$  be the set of all random variables with normal  $N(\mu, 1)$  distribution, and let  $\mathbb{I}$  be the subset of  $\mathbb{D}$  with mean  $0 \leq \mu \leq 1$ . We uniformly choose a variable  $N(\mu, 1)$  from  $\mathbb{I}$ , so that the probability or Lebesgue measure of  $\mathbb{I}$  is 1. We define in  $\mathbb{D}$  the relation of equivalence  $XRY$  if  $X - Y = q \in \mathbb{Q}$ . Then,  $\mathbb{D}$  splits into non-overlapping classes of equivalence. Now, we choose an element, i.e., a normal variable, from each class and build the set  $\mathbb{A}$ . Since  $\mathbb{Q}$  is countable, we can consider the sets  $\mathbb{A}_n = \mathbb{A} + q_n$  where  $q_n \in \mathbb{Q}$ . As  $\mathbb{B}_n = \mathbb{A}_n \pmod{1} \subset \mathbb{I}$ , the measure (if it exists) of  $\mathbb{B}_n$  is  $m(\mathbb{B}_n) < 1$ . All sets  $\mathbb{B}_n$  have the same measure. However,  $\cup_n \mathbb{B}_n = \mathbb{I}$  and the sigma-additivity of the measure implies  $\sum_n m(\mathbb{B}_n) = 1$ , so  $0 = 1$  or  $\infty = 1$ . This is impossible; hence,  $\mathbb{A} \cap \mathbb{I}$  is non-measurable or has no probability. If  $\mathbb{A} \cap \mathbb{I}$  represents a statistical model, we cannot choose a distribution from this model. Non-measurable is synonymous with non-observable.  $\square$

## References

- Domenech y Estapa, J. Absurdos geométricos que engendran ciertas interpretaciones del infinito matemático. *Mem. Real Acad. Cienc. Artes de Barcelona* **1894**, *1*, 315–329.
- Rodríguez-Salinas, B. Verdades no demostrable: teorema de Gödel y sus generalizaciones. In *Real Academia de Ciencias. Horizontes Culturales*; Espasa: Madrid, Spain, 2002; pp. 57–62.
- Schimmerling, E. *A Course on Set Theory*; Cambridge University Press: Cambridge, UK, 2011.
- Graham, L.; Kantor, J.-M. *Naming Infinity*; Harvard University Press: Cambridge, UK, 2009.
- Dauben, J.W. *Georg Cantor. Investigación y Ciencia*; Temas: Grandes Matemáticos, Spain, 1995; pp. 94–105.
- Delahaye, J.-P. Demostrar la hipótesis del continuo. *Investigación y Ciencia* **2020**, *524*, 71–79.
- Dauben, J.W. *Georg Cantor: His Mathematic and Philosophy of the Infinite*; Harvard University Press: Cambridge, MA, USA, 1979.
- Chaitin, G.J. A theory of program size formally identical to information theory. *J. ACM* **1975**, *22*, 329–340. [[CrossRef](#)]
- Cuadras, C.M.; Greenacre, M. A short history of statistical association: From correlation to correspondence analysis to copulas. *J. Multivariate Anal.* **2022**, *188*, 104901. [[CrossRef](#)]
- Cuadras, C.M.; Diaz, W.; Salvo-Garrido, S. Two generalized bivariate FGM distributions and rank reduction. *Commun. Stat.-Theory Methods.* **2020**, *49*, 5639–5665. [[CrossRef](#)]
- Cuadras, C.M.; Augé, J. A continuous general multivariate distribution and its properties. *Commun. Stat.-Theory Methods* **1981**, *10*, 339–353. [[CrossRef](#)]
- Ruiz-Rivas, C.; Cuadras, C.M. Inference properties of a one-parameter curved exponential family of distributions with given marginals. *J. Multivar. Anal.* **1988**, *27*, 447–456. [[CrossRef](#)]
- Sierpinski, W. *Hypothèse du Continu*; Chelsea Pub. Co.: New York, NY, USA, 1956.
- Bacmann, H. *Transfinite Zahlen*; Springer: Berlin/Heidelberg, Germany, 1955.
- Munroe, M.E. *Introduction to Measure and Integration*; Addison Wesley Pub. Co.: Reading, MA, USA, 1952; 1959.
- Obregón, I. *Teoría de la Probabilidad*; Limusa: Mexico City, Mexico, 1975.
- Orts, J.M. El principio de elección. *Mem. Real. Acad. Cienc. Artes de Barcelona* **1957**, *32*, 297–320.
- Lipschuz, S. *General Topology*; Schaum Pub. Co.: New York, NY, USA, 1965.
- Mosterín, J. *Teoría Axiomática de Conjuntos*; Ariel: Barcelona, Spain, 1971.
- Navarro, J. *La Nueva Matemática*; Salvat Editores: Barcelona, Spain, 1973.
- Rittberg, C.J. How Woodin changed his mind: new thoughts on the continuum hypothesis. *Arch. Hist. Exact Sci.* **2015**, *69*, 125–151. [[CrossRef](#)]
- Cuadras, C.M. Contributions to the diagonal expansion of a bivariate copula with continuous extensions. *J. Multivariate Anal.* **2015**, *139*, 28–44. [[CrossRef](#)]
- Dempster, A.P. *Elements of Continuous Multivariate Analysis*; Addison Wesley: Reading, MA, USA, 1969.
- Amari, S. *Differential-Geometrical Methods in Statistics*; Springer: New York, NY, USA, 2012; Volume 28.
- Oller, J.M.; Cuadras, C.M. Rao's distance for negative multinomial distributions. *Sankhya* **1985**, *47 A*, 75–83.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.