Discrete Mathematics for Statistical and Probability Problems

Christos P. Kitsos and Thomas L. Toulias

Abstract This paper offers a compact presentation of the solid involvement of Discrete Mathematics in various fields of Statistics and Probability Theory. As far as the discrete methodologies in Statistics are concerned, our interest is focused on the foundations and applications of the Experimental Design Theory. The set-theoretic approach of the foundations of Probability Theory is also presented, while the notions of concepts and fuzzy logic are formulated and discussed.

Introduction

The aim of this paper is to provide a Discrete Mathematics point of view of some statistical applications. Two are our main lines of thought: Design Theory and statistical distances. The Design Theory attracts interest from the group theory and projective geometry. Design Theory is discussed, while some emphasis is given to the Latin squares. We also recall the theory of ideals and provide some aspects from the probability theory that, we believe, deserves more attention.

The notion of distance is fundamental in Statistics. In mathematical analysis, especially in metric spaces, the distance serves as a criterion to check the convergence of a sequence, while a sequence in Statistics (with typical example being the Robbins-Monro iterative scheme) is asked to converge *in distribution*, which is usually the normal distribution; see [13] for details.

The reason is that such sequences can provide maximum likelihood estimators (MLE), being within the classical statistical framework, while other methods might not.

The notion of "concept" associated with the "objects" and "attributes" is introduced, from which the idea of a set-theoretic distance, in a Discrete Mathematics sense, is emerged.

C. P. Kitsos · T. L. Toulias (✉)
Technological Educational Institute of Athens, Athens, Greece
e-mail: xkitsos@teiath.gr

N. J. Daras, Th. M. Rassias (eds.), *Modern Discrete Mathematics and Analysis*,
Springer Optimization and Its Applications 131,
https://doi.org/10.1007/978-3-319-74325-7_12

Geometric methods, and therefore distance metrics methods, are adopted in various problems in Statistics. In optimal Experimental Design Theory for the continuous case, geometric methodologies are considered on the induced design space and the relative geometrical aspects have been discussed by Kitsos et al. [20]. For the discrete case, the geometrical approach is tackled in a compact form in this paper.

In principle, the usual (geometrical) distance metric in Statistics is considered to be the Euclidean distance, based on the ℓ_2 norm, but this is not the case for Discrete Mathematics; see section "Finite Geometry and Design Theory." In section "Discrete Mathematics and Statistics," the relation between Discrete Mathematics and Statistics is developed. The Experimental Design Theory is also discussed, especially the Latin squares. Moreover, a finite geometry approach is also developed in a compact form. In section "Discrete Mathematics and Probability," the applications of Discrete Mathematics to Probability is presented, while in section "Discrete Distance Measures," certain distance measures are discussed.

Discrete Mathematics and Statistics

Introduction

Discrete Mathematics offers a strong background to statistical problems, especially to the Design Theory. We shall trace some of these applications, bringing practice with theory. Consider the practical problem where a manufacturer is developing a new product. Let us assume that he/she wishes to evaluate v varieties (of the product) and asks a number of consumers to test them. However, it seems impossible in practice to ask each consumer to test all the varieties. Therefore, two lines of thought might be adopted:

1. Each consumer tests the same number of varieties, say $k < v$.
2. Each variety should be tested by r consumers.

The above problem gives rise to the following generalization: Let X be any set of size v, i.e., $v := |X|$. We say that a set $\mathscr{B}$ of k-subsets of X is a *design*, denoted by $D(v, k, r)$ with parameters $v, k, r \in \mathbb{N}^* := \mathbb{N} \setminus \{0\}$, when each member $x \in X$ belongs to exactly $r \leq v$ sets of $\mathscr{B}$. In Design Theory, a subset $B \in \mathscr{B}$ is called a *block* of the design under investigation.

Suppose now that C denotes any set of k-subsets of X with $v := |X|$. In general, we say that the marks (i.e., the readings of an experiment) are members of the set

$$\mathfrak{C} := \{(x, C)\}_{x \in C}. \tag{1}$$

What in Statistics is known as a *replication* of a value x is the row total $r(x) := \#\big(\{C : x \text{ occurs in } C\}\big)$. The column total is k in all the cases by the definition

Table 1 The table of Example 1 with $k = 3$, $|C| = 4$

x	C_1	C_2	C_3	C_4	$r(x)$
1	✓	✓		✓	3
2	✓		✓		3
3		✓			1
4	✓		✓	✓	3
5			✓		1
6		✓		✓	2
k	3	3	3	3	12

of C. Therefore, it is easy to see that

$$\sum_{x \in X} r(x) = k|C|. \tag{2}$$

Example 1 The above discussion can be visualized with Table 1 where $X := \{1, 2, \ldots, 6\}$. Notice that $\sum r(x) = 12$.

In principle, when we are working on a statistical design $D(v, k, r)$ then $r(x) := r$ and as we are working with blocks B, $b := |B|$, relation (2) is then reduced to $vr = bk$, and hence

$$b = \frac{vr}{k} \leq \binom{v}{k}. \tag{3}$$

In general, it can be proved that there is a design $D(v, k, r)$ if and only if (iff) $k \mid vr$,

$$\frac{vr}{k} \leq \binom{v}{k}.$$

The condition that each object (i.e., element of X) belongs to the same number of blocks can be strengthened. In particular, it can be required that a pair of objects or, even more, that t objects are taken at a time: this is known as *t-design*, $t \in \mathbb{Z}^+$.

Let X be a set with $v := |X|$. Then a set B of k-subsets of X is said to be a *t-design*, denoted with $D(v, k, r_t)$, iff for each t-subset T of X, the number of blocks which contain T is constant r_t. It is interesting to notice that

i. If B is a t-design, it is also an s-design, $1 \leq s \leq t - 1$, $s \in \mathbb{Z}^+$.
ii. If B is a $D(v, k, r_t)$ t-design, it is then also a $D(v, k, r_s)$ s-design, with

$$r_s = r_t \frac{(v - s)(v - s - 1) \cdots (v - t + 1)}{(k - s)(k - s - 1) \cdots (k - t + 1)}. \tag{4}$$

iii. For $0 \leq s \leq t - 1$, it is required that

$$(k - s)(k - s - 1) \cdots (k - t + 1) \mid r_t(v - s)(v - s - 1) \cdots (v - t + 1).$$

iv. A recursive formula holds:

Table 2 A 4×4 Latin square L_1

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

$$r_{t-1} = r_t \frac{v - t + 1}{k - t + 1}, \quad t \in \mathbb{N}.$$

v. If $b = |B|$, then

$$b = r_0 = r_1 \frac{v}{k}.$$

Usually a 2-design with $k = 3$ and $r_2 = 1$ is called as *Steiner Triple System* (STS). Such a system can be seen by considering two words, say u_1 and u_2, both of length n in the alphabet $\{0, 1\}$. Let $u_1 + u_2$ denote the word obtained adding the corresponding digits of u_1 and u_2. Then, if we consider X to be the set of all such words with the exception of $00\ldots0$, the set of all three subsets of X, formed by $\{u_1, u_2, u_1 + u_2\}$, is a 2-design such as $D(2^{n-1}, 3, 1)$ which is an STS design with $2^n - 1$ varieties.

Latin Squares

In Statistics, and especially in Experimental Design Theory, the Latin squares (LS), as proposed by Fisher in [5], play an important role; see [2, 23, 24] and [3] among others. The traditional example is the following: Suppose we have to plan an agriculture experiment in which four new kinds of fertilizers, say A, B, C, and D, are to be tested in a square field. The "scheme" has to be as in Table 2 in order to describe an LS.

In principle, an LS of order n is an $n \times n$ array in which each of the n elements occurs once in each row and once in each column. The statistical term LS comes from the fact that R.A. Fisher used "Latin letters" to cover the "square" field in his agricultural experiments.

We shall denote with $L(i, j)$ the (i, j) entry of an LS L, while the labels for the rows and columns of L as well as for all its (i, j) elements shall be considered as $\mathbb{Z}_m$-valued, with $\mathbb{Z}_m$ being the set of nonnegative integers modulo m.

Given two Latin squares of the same order n, say L_1 and L_2, the relation for the orthogonality between L_1 and L_2 can be defined. Indeed, $L_1 \perp L_2$ iff for every $k, l \in \mathbb{Z}_m$ there are just two $i, j \in \mathbb{Z}_m$ for which

$$L_1(i, j) = k, \quad \text{and} \quad L_2(i, j) = l.$$

Following this terminology, Euler in 1786 was the first who suggests constructing an orthogonal pair of arrays of a six-order LS. He was unable to solve this problem. It is now known that no such pair exists. Actually, the problem that Euler suggested was

> Given 36 officers of 6 different ranks from 6 different regiments, can they be arranged in a square in such a way that each row and each column contains one officer of each rank and one officer from each regiment?

What to admire? The physical problem itself or the underlined mathematical insight which tries to tackle the problem? We shall come to this problem later.

Despite Euler's problem that has no solution, there is a way of constructing orthogonal LS. Theorem 2 below offers a method of constructing orthogonal LS (OLS) based on properties of $\mathbb{Z}_p$ with p being a prime.

Theorem 1 *For each $v \geq 2$, the $v \times v$ array defined by $L(i, j) = i + j$, $i, j \in \mathbb{Z}_v$ is an LS.*

See Appendix for the proof.

Theorem 2 *Let p be a prime and $0 \neq a \in \mathbb{Z}_p$ given. Then, the rule*

$$L_a(i, j) = ai + j, \quad i, j \in \mathbb{Z}_p, \tag{5}$$

defines an LS. Furthermore, for given $b \neq a$, $b \in \mathbb{Z}_p$, it holds that

$$L_b \perp L_a. \tag{6}$$

See Appendix for the proof.

Example 2 The LS, say L_2, of Table 3 below is orthogonal to LS L_1 of Table 2, as the 16 pairs $(A, A), (A, B) \ldots, (D, D)$ occur in one of the 16 positions, i.e., $L_2 \perp L_1$ by the LS orthogonality definition.

Based on Theorem 2, we say that we can obtain a set of $p-1$ *mutually orthogonal LS* (MOLS) of order p (MOLS$_p$) for each prime p.

Example 3 For $p = 3$ we get two MOLS$_3$, i.e.,

$$L_1: \begin{matrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{matrix} \quad \text{and} \quad L_2: \begin{matrix} 0 & 1 & 2 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \end{matrix}$$

Table 3 A 4×4 Latin square L_2

A	B	C	D
C	D	A	B
D	C	B	A
B	A	D	C

So far, we discussed that for a prime p, using the properties of the field $\mathbb{Z}_p$, it is possible to construct a set of $p-1$ MOLS_p. The same result holds if we replace p with a prime power $q := p^r, r \in \mathbb{Z}^+$. Indeed:

Theorem 3 *For q being a prime r-power of p, it is possible to construct $q-1$ mutually orthogonal Latin squares of order q (MOLS_q).*

Proof Apply Theorem 2 where $a \in \mathbb{Z}_p$ is now replaced by $a \in \mathbb{F}_q$, with $q-1$ nonzero and $\mathbb{F}_q$ being a finite field in place of $\mathbb{Z}_p$.

In practice, given any field with n elements, we would like to construct $n-1$ MOLS. Due to Theorem 3 the arisen question is:

Question Is it possible to construct $n-1$ mutually orthogonal Latin squares of order n (MOLS_n) when n is not a prime power?

Answer In Discrete Mathematics and Statistics, this is one of the most well-known unsolved problems. For the case of $n = 6$, it is known already that there is not a set of 5 MOLS_6 (recall Euler's problem mentioned earlier which is unsolved).

Another approach to Design Theory, under the context of Discrete Mathematics, is through geometry and particular through finite geometry on projective planes. Recall that there are two main approaches of finite plane geometry: affine and projective. In an affine plane, the normal sense of parallel lines is valid. In a projective plane, by contrast, any two lines intersect at a unique point, so parallel lines do not exist. The finite affine plane geometry and finite projective plane geometry can be described by simple axioms. An affine plane geometry is a nonempty set X (whose elements are called "points"), along with a nonempty collection L of subsets of X (whose elements are called "lines"), such that:

(a) For every two distinct points, there is exactly one line that contains both points.
(b) Given a line ℓ and a point p outside ℓ, there exists exactly one line ℓ' containing p such that $\ell \cap \ell' = \varnothing$ (Playfair's axiom).
(c) There exists a set of four points, no three of which belong to the same line.

The last axiom ensures that the geometry is not trivial (either empty or too simple to be of interest, such as a single line with an arbitrary number of points on it), while the first two specify the nature of the geometry. Note that the simplest affine plane contains only four points; it is called the *affine plane of order* 2, where the order of an affine plane is the number of points on any line. Since no three are collinear, any pair of points determines a unique line, and so this plane (of four points) contains six lines. It corresponds to a tetrahedron where nonintersecting edges are considered "parallel" or a square where not only opposite sides but also diagonals are considered "parallel"; see Fig. 1.

In optimal Experimental Design Theory, a geometry is constructed via the induced design space; see [12, 27] and [18] among others. In this paper, the geometric approach is realized in the following through the finite field $\mathbb{F}_q$.

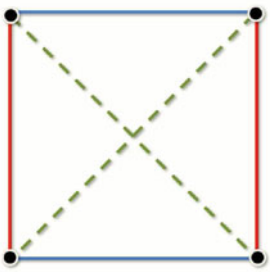

Fig. 1 A finite affine plane of order 2. The two diagonal lines do not intersect

Finite Geometry and Design Theory

Let $x, y \in \mathbb{F}_q$, where $\mathbb{F}_q$ being a finite field. Then, the "coordinate" or analytic plane geometry for $(x, y) \in \mathbb{R}^2$ is still valid for the elements of $\mathbb{F}_q$. As all the algebraic "manipulations" hold in $\mathbb{R}^2$, the sense of "line" and "plane" for a finite number of "points" $x, y \in \mathbb{F}_q$ is also parent in $\mathbb{F}_q$. In particular, the lines in a 2-design are the blocks on the set points; see Theorem 4 below. Thus, a line satisfies the analytic expression $ax + by + c = 0$ where $x, y, a, b, c \in \mathbb{F}_q$ with $a^2 + b^2 \neq 0$.

Theorem 4 *Consider a finite field $\mathbb{F}_q$ equipped with lines and planes as above. Then, the lines of $\mathbb{F}_q$ are the blocks of a 2-design $D(v, k, r_2)$ of the set of points of $\mathbb{F}_q$. In particular, the design is $D\left(q^2, q, 1\right)$.*

See Appendix for the proof.

The $D\left(q^2, q, 1\right)$ design, described in Theorem 4, is usually known as the *affine plane over* $\mathbb{F}_q$ (see also the proof in Appendix). Recall the point (iv) in sub-Section 2.1. For the 2-design above, i.e., for $t := 2$, it is

$$r_1 = r_2 \frac{v - 2 + 1}{k - 2 + 1} = r_2 \frac{v - 1}{k - 1} = 1 \times \frac{q^2 - 1}{q - 1} = q + 1. \tag{7}$$

According to property (v), as in Section 3.2, it is

$$b = r_0 = r_1 \frac{v}{k} = (q + 1) \frac{q^2}{q} = q(q + 1). \tag{8}$$

Example 4 Let $q = 3$, i.e., $\mathbb{F}_3 \equiv \mathbb{Z}_3$ is under consideration. There are $v = q^2 = 9$ points and $k = q = 3$. Those nine points, say $P_i, i = 1, 2, \ldots, 9$, are

$$\begin{array}{lll} P_1 = (0, 0), & P_2 = (0, 1), & P_3 = (0, 2), \\ P_4 = (1, 0), & P_5 = (1, 1), & P_6 = (1, 2), \\ P_7 = (2, 0), & P_8 = (2, 1), & P_9 = (2, 2). \end{array}$$

The $b = r_0 = q(q + 1) = 12$ lines, say $\ell_i, i = 1, 2, \ldots, 12$, are presented in Table 4:

Table 4 Lines of Example 4

Line	Equation	Points
ℓ_1	$x = 0$	P_1, P_2, P_3
ℓ_2	$x = 1$	P_4, P_5, P_6
ℓ_3	$x = 2$	P_7, P_8, P_9
ℓ_4	$y = 0$	P_1, P_4, P_7
ℓ_5	$y = 1$	P_2, P_5, P_8
ℓ_6	$y = 2$	P_3, P_6, P_9
ℓ_7	$x + y = 0$	P_1, P_6, P_8
ℓ_8	$x + y = 1$	P_2, P_4, P_9
ℓ_9	$x + y = 2$	P_3, P_5, P_7
ℓ_{10}	$2x + y = 0$	P_1, P_5, P_9
ℓ_{11}	$2x + y = 1$	P_2, P_6, P_7
ℓ_{12}	$2x + y = 2$	P_3, P_4, P_8

Notice that there are four classes with three lines as

$$\big\{(\ell_1, \ell_2, \ell_3), (\ell_4, \ell_5, \ell_6), (\ell_7, \ell_8, \ell_9), (\ell_{10}, \ell_{11}, \ell_{12})\big\}.$$

Each class of parallel lines has no intersection (common point) between them, while when there is an intersection, one common point exists (as in the Euclidean case of $\mathbb{R}^2$). However, if we adopt the projective geometry's approach, i.e., assume that every two lines have always one common point, we are in a finite version of projective geometry [4], and its relation with the Design Theory. Considering a prime power, Theorem 5 holds where the projective plane property over $\mathbb{F}_q$ is demonstrated in comparison with the affine plane over $\mathbb{F}_q$, according to Theorem 4.

Theorem 5 *For any prime power q, there is a 2-design $D(v, k, r_2) = D\big(q^2 + q + 1, q + 1, 1\big)$. This particular design has the additional property that any two blocks have just one member in common.*

See Appendix for the proof.

Calculating r_1 and $r_0 = b$, due to the relations (iv) and (v) in Section 3.2, it holds

$$r_1 = \frac{v-1}{k-1} r_2 = \frac{\big(q^2+q+1\big)-1}{(q+1)-1} - 1 = q + 1, \tag{9a}$$

$$b = r_0 = \binom{v}{k} r_1 = q^2 + q + 1. \tag{9b}$$

See the similarity between (7)–(9b) and (9a)–(9b). Moreover, we can notice that:

- There are $q^2 + q + 1$ points and $q^2 + q + 1$ lines.
- Each line contains $q + 1$ points and each point belongs to $q + 1$ lines.
- any two points belong to one common line and any two lines have one common point.

Example 5 Consider the affine plane over $\mathbb{F}_3$, as in Example 4. We need the parallel lines of $\mathbb{F}_3$ to meet, so we add to each line a new (arbitrary) point, which corresponds to the projective geometry's point at infinity or *infinity point*. In particular,

$$\begin{aligned}
\ell_i' &:= \ell_i \cup \{X_1\}, \quad i = 1, 2, 3, \\
\ell_i' &:= \ell_i \cup \{X_2\}, \quad i = 4, 5, 6, \\
\ell_i' &:= \ell_i \cup \{X_3\}, \quad i = 7, 8, 9, \\
\ell_i' &:= \ell_i \cup \{X_4\}, \quad i = 10, 11, 12.
\end{aligned}$$

A new line $\ell_\infty \supseteq \{X_1, X_2, X_3, X_4\}$, containing the newly introduced points X_i, $i = 1, 2, 3, 4$, is then introduced and called as the *line at infinity* or *infinity-line*. Therefore, the projective plane over $\mathbb{F}_3$ has in total 13 lines, i.e., ℓ_i', $i = 1, 2, \ldots, 12$, and ℓ_∞, and 13 points, i.e., the given P_i, $i = 1, 2, \ldots, 9$, and X_j, $j = 1, 2, 3, 4$. Each line contains four points, i.e., each block contains four elements, and each pair of points belongs exactly to one line. Hence, the 2-design $D(13, 4, 1)$ is obtained.

Applications of Experimental Design

In practice, a *complete randomized block design* (CRBD) is analyzed as a two-way ANalysis Of VAriance (ANOVA); see the pioneering work of [26]. The *incomplete balanced* needs a special ANOVA to analyze the collected data while an *incomplete general block design* is analyzed through regression analysis; see [8] among others. A (complete) Latin square is analyzed through a three-way ANOVA in industry. That is, ANOVA and regression analysis are adopted to analyze real data with the assistance of an appropriate software; see [15] among others.

The theoretical insight of experimental design provides food for thought for different branches of mathematics. We tried to present some of them in a compact form. The experimenter faces often the need of a strong mathematical background when analyzing a 2^n *factorial experiment*, defined as in [30], and especially for a portion of it. When we are talking about a *confounded experiment*, one may consider the number-theoretic Kempthorne method [11]; see [9] among others, for a number-theoretic development. A "traditional" example is the following:

Example 6 Construct two blocks in a 2^3 factorial experiment, so that the term ABC is confounded. We then have the linear combination $L = X_1 + X_2 + X_3$ with the value of L is evaluated as follows:

$$\begin{aligned}
(1) &= \mathrm{A}^0\mathrm{B}^0\mathrm{C}^0,\ L = 0 + 0 + 0 = 0, \\
a &= \mathrm{A}^1\mathrm{B}^0\mathrm{C}^0,\ L = 1 + 0 + 0 = 1, \\
b &= \mathrm{A}^1\mathrm{B}^1\mathrm{C}^0,\ L = 1, \\
ab &= \mathrm{A}^1\mathrm{B}^1\mathrm{C}^0,\ L = 2 = 0 \ (\mathrm{mod}2),
\end{aligned}$$

$$
\begin{aligned}
c &= \mathrm{A}^0\mathrm{B}^0\mathrm{C}^1,\ L = 1,\\
ac &= \mathrm{A}^1\mathrm{B}^0\mathrm{C}^1,\ L = 2 = 0\ (\mathrm{mod}2),\\
bc &= \mathrm{A}^0\mathrm{B}^1\mathrm{C}^1,\ L = 2 = 0\ (\mathrm{mod}2),\\
abc &= \mathrm{A}^1\mathrm{B}^1\mathrm{C}^1,\ L = 3 = 1\ (\mathrm{mod}2).
\end{aligned}
$$

So, for

$$
\begin{aligned}
L = 0 &\ \text{ the block is } \ (1), ab, ac, bc,\\
L = 1 &\ \text{ the block is } \ (1), a, b, c, abc.
\end{aligned}
$$

Therefore, if we decide to apply a $\frac{1}{2}2^3 = 2^{3-1}$ experiment, i.e., a half 2^3 factorial experiment, we have to choose one of the two blocks as above.

Different rules have been proposed to overpass confounding. For Fisher's multiple confounding rule, see [23]. The violation of the structure of a 2^n factorial design, by adding center points, dominates EVolutionary OPeration (EVOP), [1]. Then we moved to a new "model" by adding more points, i.e., $2^n +$ center $+$ "star", to study response surface exploration; see [25] among others.

The nonlinear (optimal) experimental design, as it was studied by Kitsos [12], has no relation with the elegant mathematical approach of Fisher. The nonlinear Design Theory suffers from parameter dependence [12], and the practical solution is to adopt sequential design; see [6]. The induced design space offers the possibility of a geometrical consideration, [18, 20]. The compromise of a quasi-sequential approach [14] was also proposed, while some technics based on polynomial root-finding, for nonlinear problems, were studied in [28].

This section offers a quick attempt to complete the development of the experimental design topic, the backbone of Statistics.

Discrete Mathematics and Probability

As we already discussed in section "Discrete Mathematics and Statistics," discrete mathematical ideas appear to have an aesthetic appeal in Statistics. Especially during the Fisher's and Kolmogorov's era, there was an attempt to develop a theoretical discrete framework to shed a new light in statistical problems. In this section, we show influence of Discrete Mathematics in other research areas related to Statistics. The measure-theoretic approach of Kolmogorov had a competitor from algebra, i.e., probability algebra, which was never successfully formulated. The measure theory context of probability was accepted completely. But, still, the algebraic definition of probability space deserves some further attention, especially when the foundation needs some discrete treatment.

Probability algebra (PA), as introduced in [10], is defined to be the pair (α, P), with $\alpha \neq \varnothing$ where its set elements $A, B, C, \ldots \in \alpha$ are called *events*. P is a real

function on α and is called *probability*. The binary operations $A \vee B$ and $A \wedge B$ and the unitary operation A^{c} (not in A) equip α with the algebraic structure of a Boolean algebra. For the probability P we assume that:

i. P is positive definite, i.e.,

$$\mathrm{P}(A) \geq 0 \text{ for every } A \in \alpha, \text{ and } \mathrm{P}(A) = 0 \Longleftrightarrow A = \varnothing \in \alpha.$$

ii. P is normed i.e.,

$$\mathrm{P}(E) = 1 \text{ where } E \in \alpha \text{ is the unitary element.}$$

iii. P is additive, i.e.,

$$\mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B) \text{ when } A \wedge B = \varnothing.$$

Let β be a Boolean sub-algebra of α. Then the restriction of the function P to β is probability on β. If $\alpha := \{\varnothing, E\}$ with $\mathrm{P}(E) = 1$, $\mathrm{P}(\varnothing) = 0$, the probability algebra (α, P) is called *improper*.

The terms probability space and Borel probability field, introduced by Kolmogorov in his pioneering work [22], are also constructed through the algebraic approach, while the distribution function was also well defined; see [10, Theorem 5.4].

For given probability algebras (α_1, P_1) and (α_2, P_2) with α_i, $i = 1, 2$, Boolean algebras, consider the isomorphism

$$\varphi : \alpha_1 \longrightarrow \alpha_2, \text{ where } A \longmapsto \varphi(A).$$

Then, we say that the two probability algebras are *isometric* iff

$$P_1(A) = P_2\big(\varphi(A)\big), \quad A \in \alpha_1.$$

Example 7 Let $A := \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, $n \geq 2$. We define α_n to be the class of all subsets of A forming a Boolean algebra. Let P_i, $i = 1, 2, \ldots, n$, $0 < \mathrm{P}_i < 1$ with $\sum_i \mathrm{P}_i = 1$ be associated with α_i, $i = 1, 2, \ldots, n$. For every subset of A of the form $\{\alpha_{\ell_1}, \alpha_{\ell_2}, \ldots, \alpha_{\ell_k}\}$, we define the probability P as follows:

$$\mathrm{P}\Big(\{\alpha_{\ell_1}, \alpha_{\ell_2}, \ldots, \alpha_{\ell_n}\}\Big) := \mathrm{P}_{\ell_1} + \mathrm{P}_{\ell_2} + \cdots + \mathrm{P}_{\ell_k}.$$

Then (α_n, P) is a probability algebra with 2^n elements, provided that $\mathrm{P}(\varnothing) = 0$.

With the above, we tried to provide some elements of the algebraic foundations of probability. Problems such us convergence in stochastic spaces, expectations of random variables, moments, etc. can be defined appropriately this algebraic approach to probability, [10]. The multivariate problem, the sequential line of thought, and other statistical fields have not been tackled yet.

Algebraic Approach to Concept

We introduce now the term *concept* through lattice theory. Recall that a *lattice* is an abstract order structure. It consists of a partially ordered set in which every two elements have a unique supremum (also called a *least upper bound* or *join*) and a unique infimum (also called a *greatest lower bound* or *meet*). An example is given by the natural numbers, partially ordered by divisibility, for which the unique supremum is the least common multiple and the unique infimum is the greatest common divisor.

The main question, from a statistical point of view, and not only, might be: why lattice? When the study of hierarchies is one of the target of the research, the hierarchy of concept can be proved dominant, using *subconcepts* and *superconcepts*. A typical example of a subconcept is "human with a disease" where the superconcept is "healthy human being" which is a subconcept of the superconcept "being." The concept has to be determined from all the objects belonging to the concept under consideration as well as from all attributes necessarily valid for the defined objects. Usually it is not expected the experimenter to consider all objects and attributes describing the given concept, i.e., a certain amount of sets of objects and attributes is initially fixed.

A *concept* is every set function ϕ of a set O, called the *objects*, to another set A, called the *attributes*. We shall use the notation (O, A).

The above definition of concept provides the mathematical insight of the expression: the "object" O has "attributes" A. Let now all the objects under consideration form the set $\mathfrak{O}$, which has a finite number of elements. Similarly, all the attributes form the set $\mathfrak{A}$ which also has a finite number of elements. We emphasize that $\phi(O)$ does not define a unique set A while, equivalently, $\phi^{-1}(A)$ does not define a unique set O. Based on the collected data, A, O and φ can be defined appropriately.

The data we examine (e.g., any qualitative attributes, "yes" or "no" to a given disease, to human or animals, or the level of quality of an industrial product) act as a generator ϕ of concepts:

$$\phi : \mathfrak{O} \longmapsto \mathfrak{A}, \quad \varphi(O) = A. \tag{10}$$

From the above discussion, the set $\mathfrak{O}$ represents a set of "humans" or "animals," and O can represent "humans with a disease" and $\varphi(O) = A = \{0, 1\}$ with $0 :=$ "no" and $1 :=$ "yes".

When we are interested to create a new concept, we must consider the simple laws of set theory.

The *concept union* $\curlyvee$ of two concepts is a new concept of the form $(O_1, A_1) \curlyvee (O_2, A_2) := (O_1 \cup O_2, A_1 \cap A_2), (O_1, A_1), (O_2, A_2) \in \mathfrak{C}$. The *concept intersection* $\curlywedge$ is defined, respectively, as $(O_1, A_1) \curlywedge (O_2, A_2) := (O_1 \cap O_2, A_1 \cup A_2), (O_1, A_1), (O_2, A_2) \in \mathfrak{C}$; see [17].

It is easy to verify that the concept $(\varnothing, A)$ is the neutral (zero) element for the union in the sense that $(O, A) \curlyvee (\varnothing, \mathfrak{A}) = (O, A)$.

The definition of the union between two concepts is not only mathematically valid but also practical, as you can assign to the empty set any attribute. The neutral element for the intersection $\frown$ is the element $(\mathfrak{O}, \varnothing)$. It can be proved that the set of all concepts with operation either $\smile$ or $\frown$ is a commutative Abelian semigroup with the appropriate neutral element. The set of all concepts $\mathfrak{C}$ it can be proved, as far as the union and intersection of concepts are concerned, to be commutative and associative and, therefore, a lattice.

Two concepts (O_1, A_1) and (O_2, A_2) belonging to $\mathfrak{C}$ are *equivalent* if and only if $A_1 = A_2$. We shall write $(O_1, A_1) \cong (O_2, A_2) \overset{\text{def.}}{\Longleftrightarrow} A_1 = A_2$.

Proposition 1 *The equivalence between two concepts is a genuine equivalence relation among concepts. Therefore, we can create a partition of the concepts (equivalent with each other) coming from the collected data* $(\mathfrak{C}/\cong)$.

See Appendix for the proof.

Therefore, all the concepts of the form (O_i, A) are equivalent to $(O_i \cup O_j, A)$, $O_i, O_j \in \mathfrak{C}$ under the relation "$\cong$," and the whole class of equivalence is formed by taking the "concept union" $\smile$. Consequently, from the objects $O_i \in \mathfrak{O}$, we create the new object elements $O_i \smile O_j$ of the power set $\mathscr{P}(\mathfrak{O})$. This is a way to classify concepts depending on attributes.

To classify concepts depending on concepts themselves, we define another equivalence relation of the form $(O_1, A_1) \equiv (O_2, A_2) \Longleftrightarrow O_1 = O_2$.

Relation "$\equiv$" is an equivalence relation as it can be proved similarly to Proposition 1.

We call the set O as the *extension* of a concept, while set A shall be called as the *intension* of it.

Now, from the definition of the concept union $\smile$, we realize that by taking the union of two concepts, we find common attributes (similarities) of another "greater" object. Correspondingly, thinking about the concept intersection, we find that "less extension implies greater intension." Lattice means order, as we mentioned already; so for every two elements of it, there exists another "upper" or "preceding" element and another "lower" or "following" element. It is not a hierarchy (a tree), but it is a network as in Figs. 2 and 3.

We now define the order relations "$\preceq$" of the lattice for the already existing operations $\smile$ and $\frown$. Indeed:

The concept (O_1, A_1) *follows* concept (O_2, A_2) or, equivalently, the concept (O_2, A_2) *precedes* concept (O_1, A_1), if and only if $O_1 \subseteq O_2$ and $A_1 \supseteq A_2$, i.e., $(O_1, A_1) \preceq (O_2, A_2) \overset{\text{def.}}{\Longleftrightarrow} O_1 \subseteq O_2$ and $A_1 \supseteq A_2$.

Moreover a ring structure can be proved for the set $\mathfrak{C}$ of concepts. The complement $(O, A)^{\mathrm{c}}$ of the concept $(O, A) \in \mathfrak{C}$ is the concept consisted by the usual set-theoretic complements of $O \in \mathfrak{O}$ and $A \in \mathfrak{A}$, i.e., $(O, A)^{\mathrm{c}} := \left(O^{\mathrm{c}}, A^{\mathrm{c}}\right)$; see [17].

The complement of a concept as above is well defined because:

(a) $O^{\mathrm{c}} \subseteq \mathfrak{O}$ and $A^{\mathrm{c}} \subseteq \mathfrak{A}$ imply that $\left(O^{\mathrm{c}}, A^{\mathrm{c}}\right) \in \mathfrak{C}$.
(b) There is only one complement O^{c} of O and only A^{c} of A; hence, there is only one complement of the concept $(O, A) \in \mathfrak{C}$.

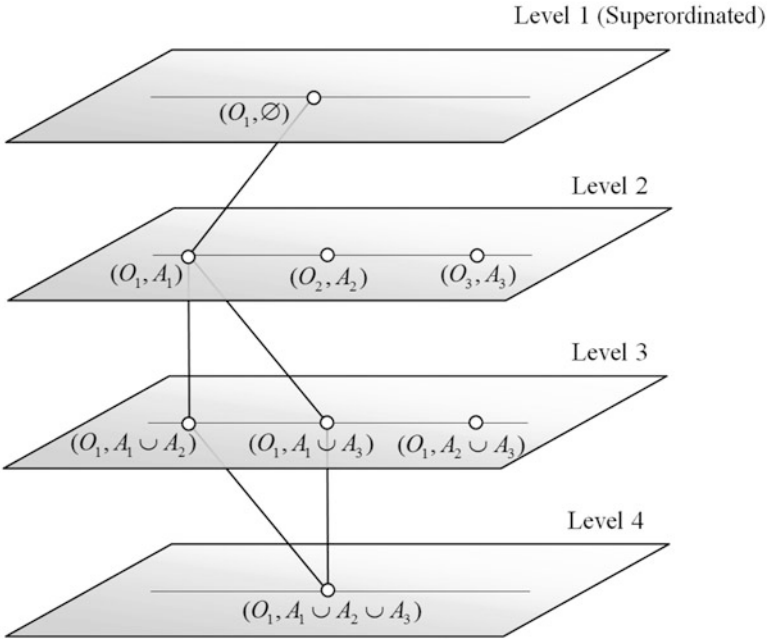

Fig. 2 From fourth level to super-ordinated

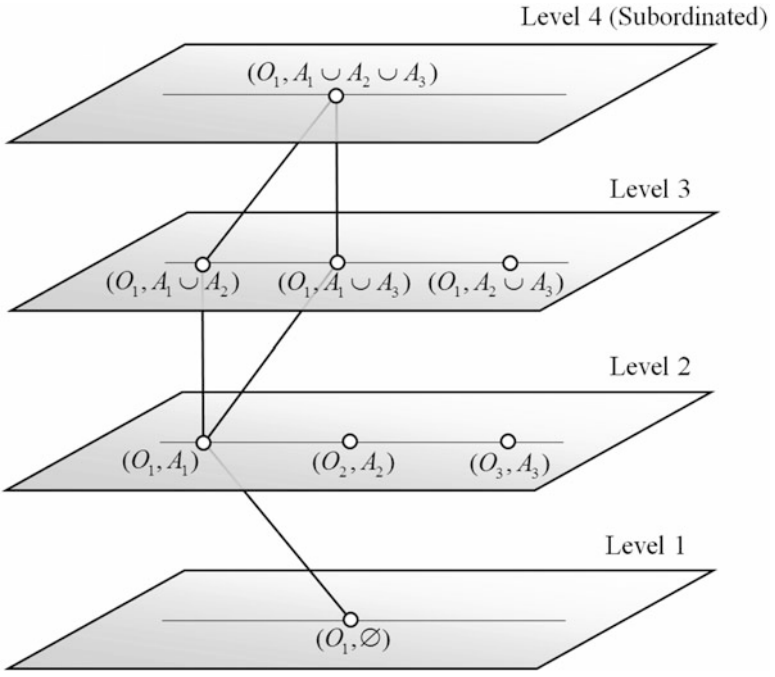

Fig. 3 The construction of concept $\left(O_1, \bigcup_{i=1}^{3} A_i\right)$ from $(O_1, \varnothing)$

The *symmetric difference* or *disjunctive union* $(O_1, A_1) \odot (O_2, A_2)$ between two concepts (O_1, A_1) and (O_2, A_2) belonging to $\mathfrak{C}$ is the concept $\left(O_1 \ominus O_2, (A_1 \ominus A_2)^c\right) \in \mathfrak{C}$ where $O_1 \ominus O_2 := (O_1 \cup O_2) \setminus (O_1 \cap O_2)$ and $A_1 \ominus A_2 := (A_1 \cup A_2) \setminus (A_1 \cap A_2)$ are the usual set-theoretic symmetric differences (or disjunctive unions) of $O_1, O_2 \in \mathfrak{O}$ and $A_1, A_2 \in \mathfrak{A}$, respectively. The following can be proved; see [17].

Theorem 6 *The set $\mathfrak{C}$ enriched with the operation $\odot$ is a group.*

Moreover, the set C enriched with the operations $\odot$ and $\circledcirc$ is a ring, where $\odot$ plays the role of "addition" and $\circledcirc$ of "multiplication." It is commutative (due to the commutative property of the operation $\circledcirc$) with unit (because of the neutral element $(\mathfrak{O}, \varnothing)$ of the operation $\circledcirc$) and distributive from both sides.

Since "$\cong$" is an equivalence relation, we can define equivalence classes of concepts, through this similarity relation in which classes are disjoint. Therefore, one can define the "orbit" in the geometrical sense, and not only, as we are moving from one class to another class; see [7] for a general affine geometric point of view of Statistics and [16] for an affine geometric approach for the logit problem. As we know, the classes are disjoint sets and their union makes the set of reference, $\mathfrak{C}$ in this paper. So, in this case, we have a partitioning of $\mathfrak{C}$ according to the defined equivalence relation.

Discrete Distance Measures

It is known that the number of coefficients in which vectors X and Y differ is a distance $d(X, Y)$, known as Hamming distance. If we let $X := (0, 0, 1, 1, 1)$ and $Y := (1, 1, 0, 0, 0)$ with $X, Y \in \mathbb{F}_2^{(5)}$, then $d(X, Y) = 5$. Such a definition is used in binary codes where the minimum distance is always desired. In particular, if we define the weight $w := w(a)$ of a word a, to be the number of ones in a, then $w(a) = d(a, 0)$ and $d(a, y) := w(a - y)$ with y being another word.

The above discussion provides us food for thought to work on the introduction of a discrete distance measure between two given concepts.

The *object distance* $d(O_1, O_2)$, i.e., the distance between two finite objects $O_1, O_2 \in \mathfrak{O}$, is defined to be the nonnegative integer expressing the number of elements of their symmetric difference $O_1 \ominus O_2$, i.e.,

$$d(O_1, O_2) := |O_1 \ominus O_2|. \tag{11}$$

Proposition 2 *The defined object distance $d(O_1, O_2)$ as above is a genuine distance metric, i.e., it satisfies the three properties of a metric: positive definiteness, symmetricity, and triangularity.*

Proof Trivial.

Recall the symmetric difference between two concepts, i.e., $C_1 \odot C_2 = \big(O_1 \ominus O_2, (A_1 \ominus A_2)^{\mathrm{c}}\big)$. The distance d between objects, as in (11), is not that informative, since it is a quantitative but not a qualitative one: two sets of objects may have many different elements, coming from the same (we assume homogenous) population, but we are not measuring the data differences qualitatively but quantitatively. Besides, we are not working with objects or attributes, but with both of them, i.e., with concepts.

The symmetric difference $O_1 \ominus O_2$ between two objects acts between attributes to create a new one of the form $(A_1 \ominus A_2)$. Thus, if we want a qualitative distance between $O_1, O_2 \in \mathfrak{O}$, we must check $(A_1 \ominus A_2)$. In such a case, we can then define

$$d(A_1, A_2) := |A_1 \ominus A_2| = |\mathfrak{A}| - \big|(A_1 \ominus A_2)^{\mathrm{c}}\big|, \quad A_1, A_2 \in \mathfrak{A}. \tag{12}$$

Note that if the distance between two attributes is increasing, i.e., there are many noncommon attributes, then (12) yields that the number of elements of $(A_1 \ominus A_2)^{\mathrm{c}}$ is decreasing.

Suppose now we have two objects, O_1 and O_2. As a measure of "comparison," we introduce the normalized distance

$$d_{\mathrm{n}}(O_1, O_2) := \frac{|O_1 \ominus O_2|}{|O_1| + |O_2|}, \quad O_1, O_2 \in \mathfrak{O}. \tag{13}$$

For a discussion and a number of calculations for different cases of objects, see the Proof of Claim in Appendix.

Claim The normalized distance between objects is a function depending on the number of different elements between them, ranging from value 0 (no differences) to 1 (everything is different).

See Appendix for the proof.

Fuzzy Logic Approach

The fuzzy logic extends the classical binary response: a sentence is either true or false (not true), and hence it belongs to the $\{0, 1\}$ binary set. This binary set can be extended to the $[0,\ 1]$ interval. In strictly mathematical terms, the *characteristic function* of a given set $Q \in \Omega$, i.e.,

$$I_Q :\ Q \longrightarrow \{0, 1\}, \quad \text{with}\ \ Q \ni x \longmapsto I_Q(x) \in \{0, 1\},$$

is now considered as the *membership function* of a fuzzy set $Q \subseteq \Omega$, i.e.,

$$M_Q :\ Q \longrightarrow [0,\ 1], \quad \text{with}\ \ Q \ni x \longmapsto M_Q(x) \in [0,\ 1].$$

The value of $M_Q(x)$ declares the degree of participation of the element $x \in Q$ which belongs/participates to the fuzzy set $Q \in \Omega$. In particular,

$$M_Q(x) := \begin{cases} 1, & \text{declares that } x \text{ belongs to } Q, \\ 0, & \text{declares that } x \text{ does not belongs to } Q, \\ q \in (0,\ 1), & \text{declares that } x \text{ belongs "partially" (in some degree) to } Q. \end{cases}$$

The above introductory elements are useful to realize the extensions succeeded by fuzzy logic, i.e., adopting an interval of values rather than a single binary value to describe a phenomenon.

The interval mathematics, as described in [29], offers another approach to develop intervals, different than the fuzzy logic one; see [21] and [31] for the corresponding applications.

Recall that, under the fuzzy logic approach, the subset-hood between two sets A and B, subsets of the "universe" set Ω, is

$$S(A, B) = \frac{k(A \cap B)}{k(A)},$$

where $k(A) := \text{card}(A) = |A|$ is the generalized cardinal number and represents the degree to which B is a subset of A. If $A \subseteq B$ then $S(A, B) = 1$. Based on this, the *fuzzy entropy* of a set A, denoted with $E_{\text{F}}(A)$, can be defined as

$$E_{\text{F}}(A) := \frac{k(A \cap A^{\text{c}})}{k(A \cup A^{\text{c}})}.$$

The fuzzy entropy of A measures "how much" underlap $A \cup A^{\text{c}}$ and overlap $A \cap A^{\text{c}}$ violate the existent laws $A \cap A^{\text{c}} = \varnothing$, $A \cup A^{\text{c}} = \Omega$. That is, the fuzzy entropy measures eventually "how much" of $A \cup A^{\text{c}}$ is included to $A \cap A^{\text{c}}$.

The following theorem rules the fuzzy entropy theory:

Theorem 7 (of Fuzzy Entropy) *It holds that*

$$E_{\text{F}}(A) = S\big(A \cup A^{\text{c}}, A \cap A^{\text{c}}\big).$$

Proof From the definition (14), it holds that

$$S\big(A \cup A^{\text{c}}, A \cap A^{\text{c}}\big) = \frac{k\Big(\big(A \cup A^{\text{c}}\big) \cap \big(A \cap A^{\text{c}}\big)\Big)}{k(A \cup A^{\text{c}})} = \frac{k(A \cap A^{\text{c}})}{k(A \cup A^{\text{c}})} = E_{\text{F}}(A).$$

Example 8 It holds

$$E_{\text{F}}(\Omega) = S\big(\Omega \cup \Omega^{\text{c}}, \Omega \cap \Omega^{\text{c}}\big) = S\big(\Omega \cup \varnothing, \Omega \cap \varnothing\big) = S(\Omega, \varnothing) = 0.$$

Therefore, the universe set Ω has fuzzy entropy 0, while for the middle point M, it holds

$$E_{\mathrm{F}}(M) = S\big(M \cup M^{\mathrm{c}}, M \cap M^{\mathrm{c}}\big) = 1.$$

Based on the above discussion, we can define the *fuzzy entropy deviance*, i.e.,

$$\delta(O_1, O_2) := E_{\mathrm{F}}(O_1) - E_{\mathrm{F}}(O_2) = \frac{k\big(O_1 \cap O_1^{\mathrm{c}}\big)}{k\big(O_1 \cup O_1^{\mathrm{c}}\big)} - \frac{k\big(O_2 \cap O_2^{\mathrm{c}}\big)}{k\big(O_2 \cup O_2^{\mathrm{c}}\big)}.$$

It is always a problem to define a distance measure when information divergences are under consideration; see [19] for the continuous case of the Kullback–Leibler divergence.

We would like to emphasize that fuzziness and randomness are different ideas. They seem similar, but they are not identical. The randomness concerns problems where the event is well defined but it is uncertain if it will take place or not. The fuzziness concerns situations which are not well defined and can only be described in a sufficient way when it is known how we shall move between different classes. That is, we are moving under a fuzzy event to the "probability of a fuzzy event," which is still (even to such a probability oriented procedure) closer to a measure-theoretic approach than to an algebraic approach (via probability algebras). Moreover, in fuzzy logic, the additivity due to a "measure" is not existing, so it is not a defined probability measure, while the term "possibility" replaces the term "probability."

Discussion

This paper studied the general influence of the Discrete Mathematics line of thought to Statistics and probability. In particular, the Experimental Design Theory employs many discrete statistical concepts, which offer a solid framework, although the real data analysis is mainly performed by the ANOVA approach. Furthermore, geometry plays an important role in all the mathematical scenarios—so does in the Experimental Design Theory in its finite formulation.

The foundations of probability theory are mainly based on measure-theoretic concepts. But still, there is also a set-theoretic approach. Lattice theory is applied in concepts and Boolean algebra supports the fuzzy logic extension. Distance measures offer criteria to decide "how close" two estimates or two sets or two concepts are. A brief discussion to the subject was also offered in this paper.

Appendix

Proof (of Proposition 1) We shall prove that the introduced relation $\cong$ is reflective, symmetric, and transitive. Indeed:

i. Reflexivity of relation "$\cong$". Indeed, $(O, A) \cong (O, A) \Leftrightarrow A = A$, which is true due to the reflexivity of the equality relation "=" for sets.
ii. Symmetricity of relation "$\cong$". Indeed, $(O_1, A1) \cong (O_2, A_2) \Leftrightarrow A_1 = A_2 \Leftrightarrow A_2 = A_1$ (symmetricity of the equality relation "=" for sets) $\Leftrightarrow (O_2, A_2) = (O_1, A_1)$.
iii. Transitivity of relation "$\cong$". Indeed, it holds that

$$(O_1, A_1) \cong (O_2, A_2) \Leftrightarrow A_1 = A_2 \text{ and} \tag{14a}$$

$$(O_2, A_2) \cong (O_3, A_3) \Leftrightarrow A_2 = A_3, \tag{14b}$$

which are both equivalent to $A_1 = A_3$, due to the transitivity of the equality relation "=" for sets, and hence $(O_1, A_1) \cong (O_3, A_3)$.

Proof (of Theorem 1) Let us consider (i, j) and $(i.j')$ be the same symbols for the position. Then,

$$L(i, j) = L(i, j') \Longrightarrow i + j = i + j'.$$

As $i, j, j' \in \mathbb{Z}_V$, then $-i$ exists and thus

$$(-i) + i + j = (-i) + i + j' \Longrightarrow j = j'.$$

This means each symbol occurs once in row i. Since there are v symbols and v positions, each symbol occurs exactly once. The same line of thought is followed for the columns. Therefore, $L(i, j)$ is an LS.

Proof (of Theorem 2) Following the same line of thought of Theorem 1, we prove that $L_a = L_a(i, j)$ is an LS. Indeed, for $L_a(i, j) = L_a(i, j')$, it holds $ai + j = ai + j'$. Since $a, i, j \in \mathbb{Z}_p$, then $a^{-1}, -j \in \mathbb{Z}_p$, and hence $j = j'$. Similarly, $L_a(i, j) = L_a(i, j')$ yields $j' = j$. Consider now the position, say (i_1, j_1) of L_a, and a different position, say (i_2, j_2) of L_b. Moreover, let k_1, k_2 be the symbols for both positions. Then, for

$$L_a(i_1, j_1) = k_1, \quad L_b(i_1, j_1) = k_2, \text{ it is}$$
$$ai_1 + j_1 = k_1, \quad bi_1 + j_1 = k_2,$$

and for

$$L_a(i_2, j_2) = k_1, \quad L_b(i_2, j_2) = k_2, \text{ it is}$$
$$ai_2 + j_2 = k_1, \quad bi_2 + j_2 = k_2.$$

Thus,

$$a(i_1 - i_2) = j_2 - j_1 \text{ and } b(i_1 - i_2) = j_2 - j_1.$$

Assuming that $i_1 \neq i_2$ then $(i_1 - i_2)^{-1} \in \mathbb{Z}_p$ and

$$a = b = (i_1 - i_2)^{-1}(j_2 - j_1).$$

However, we assumed that $a \neq b$; thus, k_1 and k_2 are equal in only one position. Therefore, $L_a \perp L_b$.

Proof (of Theorem 4) Since x, y are elements of $\mathbb{F}_q$, they can take q different values. So, there are $v = q^2$ points. As far as the block is concerned, we must prove that every line has exactly q points and that any two points of $\mathbb{F}_q$ belong to exactly one line. Indeed:

Consider the line $ax + by + c = 0$, $b \neq 0$. Then,

$$y = -b^{-1}(c + ax),$$

such that (x, y) is on the line, and hence the line has q points. If $b = 0$, $a \neq 0$, it holds

$$x = -a^{-1}c.$$

In such a case, there are q possible values of y in $\mathbb{F}_q$ and q points of the form $\left(-a^{-1}c, y\right)$ lie on the line.

Now, suppose that (x_1, y_1) and (x_2, y_2) are two given distinct points, and hence $x_2 - x_1$ and $y_1 - y_2$ are not both zero. The equation of the line "passing" (actually "containing") the two points is

$$\ell : (y_1 - y_2)x + (x_2 - x_1) = x_2y_1 - x_1y_2,$$

is the equation of a line. Moreover, it contains the two given points. If another line is containing the two given points and described by the analytic form

$$ax + by + c = 0,$$

then it holds

$$ax_1 + by_1 + c = 0 \text{ and } ax_2 + by_2 + c = 0, \text{ i.e.,}$$
$$a(x_2 - x_1) = b(y_1 - y_2).$$

The value $(x_2 - x_1)^{-1}$ exists in $\mathbb{F}_q$, provided $x_1 \neq x_2$, and hence

$$a = b(x_2 - x_1)^{-1}(y_1 - y_2) = \lambda(y_1 - y_2).$$

So we have

$$b = \lambda(x_2 - x_1) \text{ and}$$
$$c = -ax_1 - by_1 = -ax_1 - \lambda(x_2 - x_1)y_1 = -\lambda(y_1 - y_2)x_1 - \lambda(x_2 - x_1)y_1$$
$$= \lambda(x_1 y_2 - x_2 y_1).$$

Thus, the line is the same with the above defined since lines ℓ and $\lambda\ell$ coincide, in finite geometry. Therefore, only one line exists and "contains" the two points.

Proof (of Theorem 5) In the affine plane over $\mathbb{F}_q$, the lines

$$ax + by + c = 0 \text{ and } a'x + b'y + c' = 0,$$

are said to be *parallel* if $ab' = a'b$ in $\mathbb{F}_q$. There are $q + 1$ equivalence classes of parallel lines of the form

$$x + \lambda y = 0, \quad \lambda \in \mathbb{F}_q,$$

and the $y = 0$ line. Any point of the affine plane belongs to just one line in each class.

We introduce now $q + 1$ points X_λ, $\lambda \in \mathbb{F}_q$, and X_∞, all belonging to a new line ℓ_∞; see also Example 5 for the discussion on ℓ_∞ line. The X_λ points lie to each line parallel to line $x + \lambda y = 0$, while X_∞ to each line parallel to $y = 0$. We have to prove that the lines are blocks of a design with parameters stated above. There are $q^2 + q + 1$ points and each line contains $q + 1$ points. Thus, we proved that any two distinct points, say H and I, belong to just one line. Then, the following cases can be true:

1. The points H and I are both parts of the initial affine plane; let us called them "old" points. So, H and I belong to a unique line of this plane, which corresponds uniquely to a line on the extended plane (with infinity point).
2. If H is an "old point" and I is a "newly added" point, i.e., $I := X_\lambda$, then H belongs already to precisely one "old" line in the parallel class represented by X_∞, and, therefore, the corresponding new line in the unique line ℓ' contains points H and I. The same is certainly true if $X := X_\infty$.
3. If H and I are both "new" points, then they belong to ℓ_∞ by definition. Therefore, any two points belong to just one line (i.e., to one block). Now, from the above construction, any two lines have just one common point.

Proof (of Claim) We distinguish the following cases:

- *Case* $O_1 \cap O_2 = \varnothing$ (disjointed objects). In general, it holds

$$O_1 \ominus O_2 = (O_1 \setminus O_2) \cup (O_2 \setminus O_1) = O_1 \cap O_2, \quad O_1, O_2 \in \mathfrak{O}, \tag{15}$$

and, hence, in this case. we obtain

$$|O_1 \ominus O_2| = |O_1 \cup O_2| = |O_1| + |O_2| - |O_1 \cap O_2| = |O_1| + |O_2|, \tag{16}$$

which is—in principle—the maximum possible number of elements of the symmetric difference between two objects. Thus, the normalized distance d_{n} between $O_1, O_2 \in \mathfrak{O}$ equals 1. Indeed, via (15),

$$\begin{aligned} |O_1 \ominus O_2| &= \big|(O_1 \setminus O_2) \cup (O_2 \setminus O_1)\big| \\ &= \|(O_1 \setminus O_2| + |O_2 \setminus O_1| - \big|(O_1 \setminus O_2) \cap (O_2 \setminus O_1)\big|) \\ &= |O_1 \setminus O_2| + |O_2 \setminus O_1| - |O_1 \cap O_2| \\ &\leq |O_1 \setminus O_2| + |O_2 \setminus O_1| \leq |O_1| + |O_2|, \end{aligned} \tag{17}$$

where the equality holds iff $O_1 \setminus O_2 = O_1$ and $O_2 \setminus O_1 = O_2$, which is equivalent to $O_1 \cap O_2 = \varnothing$. Furthermore, the normalized distance between O_1 and O_2 is confirmed to be 1 since

$$d_{\mathrm{n}}(O_1, O_2) := \frac{|O_1 \ominus O_2}{|O_1| + |O_2|} = \frac{|O_1| + |O_2|}{|O_1| + |O_2|} = 1. \tag{18}$$

- *Case* $O_1 \subseteq O_2$ (included objects). If O_1 is a subset of O_2, then

$$\begin{aligned} |O_1 \ominus O_2| &= |O_1 \setminus O_2| + |O_2 \setminus O_1| - \big|(O_1 \setminus O_2) \cap (O_2 \setminus O_1)\big| \\ &= |\varnothing| + |O_2 \setminus O_1| - \big|\varnothing \cap (O_2 \setminus O_1)\big| \\ &= |O_2 \setminus O_1| - |\varnothing| \leq |O_2|, \end{aligned} \tag{19}$$

where the equality holds iff $O_1 = \varnothing$. In principle, the normalized distance d_{n} between objects is less than or equal to 1. Indeed, for $O_1, O_2 \in \mathfrak{O}$,

$$\begin{aligned} d_{\mathrm{n}}(O_1, O_2) &= \frac{|O_1 \setminus O_2| + |O_2 \setminus O_1| - \big|(O_1 \setminus O_2) \cap (O_2 \setminus O_1)\big|}{|O_1| + |O_2|} \\ &\leq \frac{|O_1 \setminus O_2| + |O_2 \setminus O_1|}{|O_1| + |O_2|} \leq \frac{|O_1| + |O_2|}{|O_1| + |O_2|} = 1. \end{aligned} \tag{20}$$

The above is also confirmed, via (19), for the specific case of $O_1 \subseteq O_2$, as

$$d_{\mathrm{n}}(O_1, O_2) := \frac{|O_1 \ominus O_2}{|O_1| + |O_2|} = \frac{|O_2 \setminus O_1|}{|O_1| + |O_2|} \leq \frac{|O_2|}{|O_1| + |O_2|} \leq 1, \tag{21}$$

where, again, the equality holds iff $O_1 = \varnothing$.

- *Case* $O_2 \subseteq O_1$ (included objects). If O_2 is a subset of O_1, then we obtain dually that

$$d_{\mathrm{n}}(O_1, O_2) := \frac{|O_1 \ominus O_2|}{|O_1| + |O_2|} = \frac{|O_1 \setminus O_2|}{|O_1| + |O_2|} \leq \frac{|O_1|}{|O_1| + |O_2|} \leq 1, \tag{22}$$

  where the equality holds iff $O_2 = \varnothing$.
- *Case* $O_1 = O_2$ (equated objects). If object O_1 coincides with object O_2, then

$$\begin{aligned} d_{\mathrm{n}}(O_1, O_2) &= \frac{|O_1 \setminus O_2 + |O_2 \setminus O_1| - \big|(O_1 \setminus O_2) \cap (O_2 \setminus O_1)\big|}{|O_1| + |O_2|} \\ &= \frac{|\varnothing| + |\varnothing| - |\varnothing \cap \varnothing|}{|O_1| + |O_2|} = 0. \end{aligned} \tag{23}$$

References

1. Box, G.E.P.: Evolutionary operation: a method of increasing industrial productivity. J. R. Soc. Ser. B **26**, 211–252 (1957)
2. Box, I.F.: R.A. Fisher: The Life of a Scientist. Wiley, New York (1978)
3. Cochran, W.C., Box, G.M.: Experimental Design. Wiley, New York (1957)
4. Coxeter, H.S.M.: Projective Geometry. Springer, New York (1987)
5. Fisher, R.A.: The Design of Experiments. Oliver and Boy, London (1947)
6. Ford, I., Kitsos, C.P., Titterington, D.M.: Recent advances in nonlinear experimental design. Technometrics **31**, 49–60 (1989)
7. Fraser, D.A.S.: Structural Inference. Wiley, New York (1968)
8. Hicks, C.R.: Fundumental Concepts in the Design of Experiments. Holt, Rinehart and Winston, New York (1973)
9. Hunter, J.: Numder Theory. Oliver and Boyd, Clasgow (1964)
10. Kappos, D.: Probability Algebra and Stochastic Spaces. Academic, New York, London (1969)
11. Kempthorne, O.: The Design and Analysis of Experiments. Wiley, New York (1952)
12. Kitsos, C.P.: Design and Inference for Nonlinear Problems. PhD Thesis, University of Glasgow (1986)
13. Kitsos, C.P.: Fully sequential procedures in nonlinear design problems. Comput. Stat. Data Anal. **8**, 13–19 (1989)
14. Kitsos, C.P.: Quasi-sequential procedures for the calibration problem. In: Dodge, Y., Whittaker, J. (eds.) COMPSTAT 92 Proceedings, vol. 2, pp. 227–231 (1992)
15. Kitsos, C.P.: Statistical Analysis of Experimental Design (in Greek). New Technologies, Athens (1994)
16. Kitsos, C.P.: The SPRT for the Poisson process. In: e-Proceedings of the XII-th Applied Stochastic Models and Data Analysis (ASMDA 2007), Chania, Crete, 29 May–1 June 2007
17. Kitsos, C.P., Sotiropoulos, M.: Distance methods for bioassays. Biometrie und Medizinische Informatik Greifswalder Seminarberichte **15**, 55–74 (2009)
18. Kitsos, C.P., Toulias, T.L.: On the information matrix of the truncated cosinor model. Br. J. Math. Comput. Sci. **4**(6), 759–773 (2014)
19. Kitsos, C.P., Toulias, T.L.: Hellinger distance between generalized normal distributions. Br. J. Math. Comput. Sci. **21**(2), 1–16 (2017)
20. Kitsos, C.P., Titterington, D.M., Torsney, B.: An optimal design problem in rhythmometry. Biometrics **44**, 657–671 (1988)

21. Klir, G., Yan, B.: Fuzzy Logic. Prentice Hall, New Jersey (1995)
22. Kolmogorov, A.N.: Grundbegriffe der Wahrscheinlichkeitsrechung. Springer, Berlin (1933)
23. Mead, R.: The Design of Experiments. Cambridge University Press, Cambridge (1988)
24. Montgomery, D.G.: Design and Analysis of Experiments. Wiley, New York (1991)
25. Myers, R.H.: Response Surface Methodology. Allyn and Bacon Inc., Boston (1971)
26. Scheffe, H.: The Analysis of Variance. Wiley, New York (1959)
27. Silvey, S.D.: Optimal Design - An Introduction to the Theory for Parameter Estimation. Springer Netherlands, Dordrecht (1980)
28. Toulias, T.L., Kitsos, C.P.: Fitting the Michaelis-Menten model. J. Comput. Appl. Math. **296**, 303–319 (2016)
29. Wolfe, M.A.: Interval mathematics, algebraic equations and optimization. J. Comput. Appl. Math. **124**, 263–280 (2000)
30. Yates, F.: Design and Analysis of Factorial Experiment. Imperial Bureau of Soil Sciences, London (1937)
31. Zandeh, L.A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst. **90**, 111–127 (1997)