

deltamath

$h = 6.63 \cdot 10^{-34} \text{ J}\cdot\text{s}$ $\pi = 3,141592\dots$

MATHEMATICS — PHYSICS — ASTRONOMY — INFORMATICS

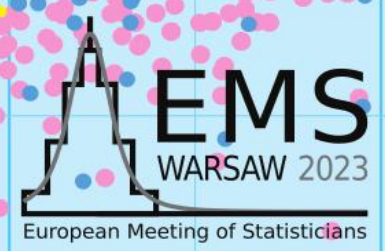
NR 7 (590) 2023

PL ISSN 0137-3005 | NR IND 35 550 X
MONTHLY
www.deltami.edu.pl

Special
Statistical
Issue

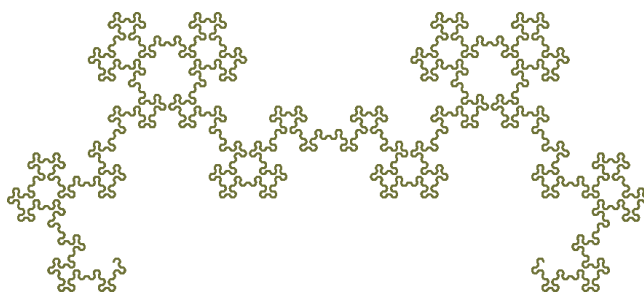


ISSN 0137-3005
9 17701371300502 07 >





In the next issue:
Fractal world of paper ribbon.




CONTENTS

ISSUE 7 (590)

Statistics with imprecise data
Przemysław Grzegorzewski

Magic fives
Oskar Skibski

 Virus means poison
Magdalena Fikus

All models are wrong,
so which ones are useful?
Przemysław Biecek

A boy or a girl?
Łukasz Rajkowski

News of Physics
New answers, new questions

Searching for order
Andrzej Dąbrowski


 Problems

Club 44

Are these distributed uniformly?
Radosław Poleski

Straight from Heaven: Wooden clock

The Night Sky in July

 Markov Chains – part 1
Bartłomiej Bzdęga

str. 1

str. 6

str. 8

str. 9

str. 12

str. 14

str. 15

str. 19

str. 20

str. 21

str. 23

str. 23

str. 25

Delta Monthly Magazine – Mathematics, Physics, Astronomy, Informatics is published by the University of Warsaw in collaboration with scientific societies: the Polish Mathematical Society, Polish Physical Society, Polish Astronomical Society, and Polish Information Processing Society.

Editorial Committee: dr Waldemar Berej, doc. dr Piotr Chrzęstowski-Wachtel, dr Krzysztof Ciesielski, prof. UJ – President, prof. dr hab. Bożena Czerny, dr Andrzej Dąbrowski, dr Tomasz Greczyło, prof. UW, dr Adam Gregosiewicz, dr Andrzej Grzesik, prof. dr hab. Agnieszka Janiuk, dr hab. Artur Jeż, prof. UW, dr hab. Bartosz Klin, prof. dr hab. Andrzej Majhofer – vice-President, dr Adam Michalec, prof. dr hab. Damian Niwiński, prof. dr hab. Krzysztof Oleszkiewicz, dr hab. Krzysztof Pawłowski, prof. PAN, dr Milena Ratajczak, dr hab. Radosław Smolec, prof. PAN, prof. dr hab. Paweł Strzelecki, prof. dr hab. Andrzej Wysmolek.

Editorial Board: Wiktor Bartol, Michał Bejger, Szymon Charzyński – Editor-in-Chief Agnieszka Chudek, Anna Durkalec, [Marta Gródek], Katarzyna Małek, Michał Miśkiewicz, Wojciech Przybyszewski, Łukasz Rajkowski – Deputy Editor-in-Chief Anna Rudnik, Krzysztof Rudnik, Oskar Skibski, Marzanna Wawro – secretary of the Editorial Board.

Correspondence address:

Delta's editorial office, ul. Banacha 2, room 4020, 02-097 Warszawa

e-mail: delta@mimuw.edu.pl tel. 22-55-44-402.

Covers and graphics:

Anna Ludwicka Graphic Design & Serigrafia.

Typeset using L^AT_EX by the Editors.

Printed at Poligrafia NOT poligrafianot.pl

Subscription:

Garmond Press: www.garmondpress.pl

Kolporter: www.kolporter.com.pl (institutions only)

RUCH S.A.: www.prenumerata.ruch.com.pl

Back Issues (from 1987 onwards) can be purchased in person at the Editorial Office or ordered via email.

Price per copy: for the last 12 months, 6 PLN; for earlier issues, 3 PLN.



Website (including archived articles, links, etc.):
deltami.edu.pl

You can also find us on
facebook.com/Delta.czasopismo

Publisher: University of Warsaw

Dear Reader,

Data is one of the resources on which modern civilization is built. Like most resources, it requires processing before it becomes fully useful. The intellectual refinery that humanity has developed to deal with the abundant deposits of information is broadly termed statistics. Statistics also serves as a common denominator for the articles published in this edition of *Delta*, the Polish popular science monthly that you are holding in your hands. In almost 50 years of its history it has been striving to bring the subject areas that it covers – mathematics, informatics, physics and astronomy – closer to its readers. In a similar way, statistics brings the knowledge hidden in data closer to researchers and, in turn, the society.

The creation of this issue is correlated with the 34th conference “European Meeting of Statisticians” held in Warsaw from July 3rd to 7th, 2023. The local organization of the conference is by the Polish Mathematical Society, Warsaw University of Technology, and the University of Warsaw; the latter is also a publisher of *Delta*. The participants have been given copies of the English version of this edition of *Delta*. It is true that participants of this conference hardly need an additional education in statistics, but we believe that everyone will find some intellectual stimulation on these pages, regardless of their scientific background and experience.

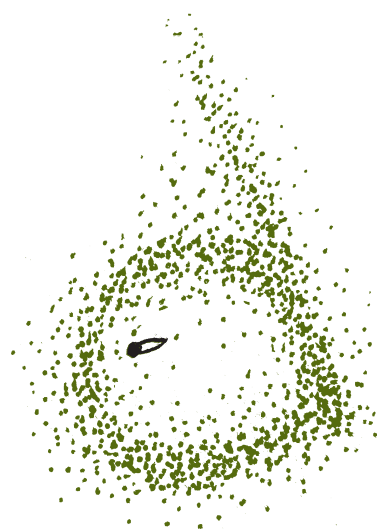
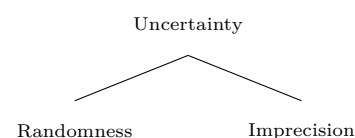
We wish you pleasant reading and, if you are a participant of the conference, a memorable event.

Editorial Board

Statistics with imprecise data

Przemysław GRZEGORZEWSKI*

* Faculty of Mathematics and Information Science, Warsaw University of Technology



Statistics might be perceived as an art of making decisions in the presence of uncertainty. It delivers tools for describing and explaining reality as well as for making predictions and verifying hypotheses. For a long time, uncertainty has been identified with randomness, and consequently, probability has been perceived as the only well-grounded theory of uncertainty. However, during the last fifty years, several approaches extending or orthogonal to the classical probability theory have appeared. A common feature of these new approaches is an attempt to soften the classical methods so that they can more easily adapt to the factual nature of the data available and deal with other types of uncertainty, such as imprecision.

It is important to remember that imprecision as a concept itself is not entirely unambiguous. Quite often the results of an experiment are imprecise due to inaccuracy of the measuring apparatus or errors made by the persons making the measurements. Sometimes the desired measurement is so difficult that its result, as a rule, should be treated as highly uncertain. It may also happen that the exact value of a variable is intentionally hidden for some confidentiality reasons. In all these situations data are often recorded as set-valued objects (e.g. as intervals) containing the exact but unknown values so a set-valued observation A delivers incomplete information about the point quantity x : we know only that A contains x but the true value of x remains unknown. Hence A represents the **epistemic** state of the subject. But there are also situations when the experimental data appear as essentially imprecise. A typical case is the analysis of perceptions collected from a human when there is no objective value behind (like the taste or mood). Another example refers to objects or phenomena with an intrinsically gradual representation subject to variability in nature, with fuzzy or changing boundaries, flexible time intervals or rating scales, etc. Each such observation represents an objective entity, even if it is vague, and hence corresponds to **ontic** imprecision.

A convenient method of mathematical modeling of imprecision was indicated by Lotfi A. Zadeh (1921–2017) who introduced **fuzzy set theory** as an extension of the classical set theory. Zadeh, one of the most outstanding thinkers of the current time, realized that although we are used to dividing everything into “yes” and “no” or to black and white, the entire world is in shades of grey. His famous statement that *everything is a matter of degree* became the main idea behind fuzzy logic and its impressive applications. It is worth noting that fuzzy logic is actually not a “fuzzy” logic, but a logic that describes and tames imprecision.

A fundamental Zadeh’s concept is a **fuzzy set**. Let \mathbb{U} be a universe of discourse. A **fuzzy set** A in \mathbb{U} is identified with a mapping, called a **membership**

L.A. Zadeh, *Fuzzy sets*, Information and Control 8 (1965), 338–353.

Note that in the set theory functions $f: X \rightarrow Y$ are usually defined as sets; more precisely, subsets of $X \times Y$ such that for all $x \in X$ there exists a unique $y \in Y$ satisfying $(x, y) \in f$. In this sense $f(x)$ is only a notational convention to denote y for which $(x, y) \in f$.

function $A: \mathbb{U} \rightarrow [0, 1]$, which assigns to each object $x \in \mathbb{U}$ a real number in the interval $[0, 1]$, so that $A(x)$ represents the degree of membership of x into A . Thus a fuzzy set A may be perceived as a (standard) subset of $\mathbb{U} \times [0, 1]$

$$A = \{(x, A(x)) : x \in \mathbb{U}, A(x) \in [0, 1]\}.$$

The interpretation of the membership function is natural: if $A(x) = 1$ then we are sure that element x belongs to A , while $A(x) = 0$ means that x does not belong to A . In all other cases, i.e. if $A(x) \in (0, 1)$, we have a partial membership (belongingness) to A . It means that if $A(x)$ is close to 1 then the degree of membership of x in A is high, while if $A(x)$ is close to 0 then the degree of membership of x in A is low. If $A(x) \in \{0, 1\}$ for all $x \in \mathbb{U}$ then A is a set in the classical meaning (each “usual” set is a fuzzy set whose membership function is its characteristic function).

Another important notion connected with a fuzzy set is the so-called α -cut. For each $\alpha \in [0, 1]$ the α -cut of a fuzzy set A , denoted as A_α , is given by

$$A_\alpha = \begin{cases} \{x \in \mathbb{U} : A(x) \geq \alpha\} & \text{if } \alpha \in (0, 1), \\ \text{cl}\{x \in \mathbb{U} : A(x) > 0\} & \text{if } \alpha = 0, \end{cases}$$

where cl stands for the closure (for now on we assume that \mathbb{U} is equipped with such operation). In other words, the α -cut is a “usual” subset of \mathbb{U} whose degree of belonging to A is not less than α . It can be shown that every fuzzy set is completely characterized by a family of all its α -cuts $\{A_\alpha\}_{\alpha \in [0,1]}$. Two α -cuts are of special interest: A_1 known as the **core**, which contains all values which are fully compatible with the concept described by A and A_0 called the **support**, which are compatible to some extent with the concept modeled by A .

An important subfamily of fuzzy sets are fuzzy numbers. We say that A is a **fuzzy number** if $A: \mathbb{R} \rightarrow [0, 1]$ such that its α -cuts for each $\alpha \in [0, 1]$ are nonempty closed intervals. An example of a fuzzy number is shown in Fig. 1.

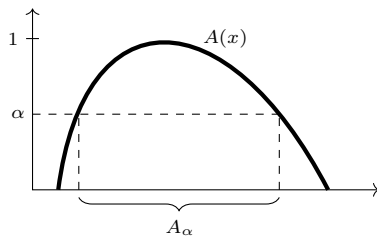


Fig. 1. A membership function $A(x)$ of a fuzzy number A .

Example. Gamonedo cheese is a kind of blue cheese produced in Asturias (northern Spain). It experiences a smoking process and later on is left to settle in natural caves or a dry place. To maintain the quality of the cheese, experts (tasters) express their subjective perceptions about different characteristics of the cheese, such as visual parameters (shape, rind, appearance), texture parameters (hardness and crumbliness), olfactory-gustatory parameters (smell intensity, smell quality, flavor intensity, flavor quality, and aftertaste) and an overall impression of the cheese. Recently tasters were asked to express their subjective perceptions about the quality of the Gamonedo cheese by using fuzzy numbers. This type of fuzzy number is the most commonly used for fuzzy descriptions both because is easy to understand by the tasters and simple in further processing. Valuation of the different features of each cheese is made over a graduate scale ranging from 0% (for lowest quality) to 100% (for highest quality). The 0-level is the set of values considered by a tester as compatible with his opinion to some extent, i.e., he thinks it is not possible that the quality is out of this set. The 1-level is the set of values considered as fully compatible with his opinion. For example, Fig. 2 illustrates a situation of a tester who believes that a given cheese meets the quality requirements in terms of the examined feature in 70–80%. At the same time, he is undoubtedly convinced that the quality requirements are satisfied with not lower than 50% but not higher than 90%.

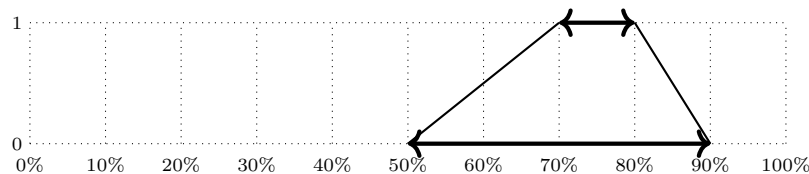


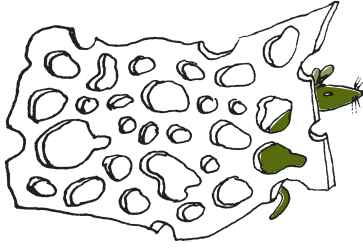
Fig. 2. An exemplary opinion of a taster expressed by a trapezoidal fuzzy set.

As is seen in the figure, both 0-level and 1-level are linearly interpolated to get the so-called trapezoidal fuzzy set used later to represent this tester’s personal

Ramos-Guajardo A.B., et al., *Applying statistical methods with imprecise data to quality control in cheese manufacturing*, In: Grzegorzewski P., et al. (Eds.), *Soft Modeling in Industrial Manufacturing*, Springer 2019, pp. 127–147.



Solution to Problem M 1750. Consider any arrangement of numbers. Note that the numbers from 1 to 2022 represent at most 2022 rows and 2022 columns. Therefore, there exists a row and a column that contain only numbers greater than 2022. The product of any two of these numbers is at least $2023 \cdot 2024$, which is greater than any number on the board. This means that there is no arrangement of numbers satisfying the conditions stated in the problem.



valuation. A sample of tester's opinions modeled with trapezoidal fuzzy sets is given in Fig. 3.

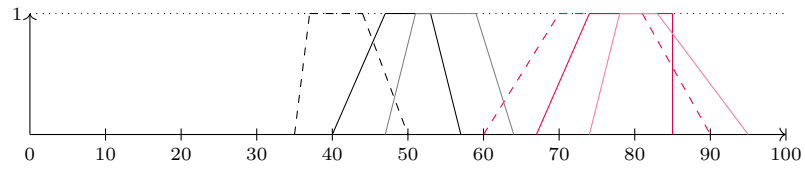


Fig. 3. A sample of tester's opinions modeled with trapezoidal fuzzy sets.

More formally, we say, that A is a **trapezoidal fuzzy number** if its membership function is given by

$$(1) \quad A(x) = \begin{cases} \frac{x-a_1}{a_2-a_1} & \text{if } a_1 \leq x < a_2, \\ 1 & \text{if } a_2 \leq x \leq a_3, \\ \frac{a_4-x}{a_4-a_3} & \text{if } a_3 < x \leq a_4, \\ 0 & \text{otherwise,} \end{cases}$$

where $a_1, a_2, a_3, a_4 \in \mathbb{R}$ such that $a_1 \leq a_2 \leq a_3 \leq a_4$. Thus, since a trapezoidal fuzzy number (1) is characterized completely by four real numbers, it is often denoted as $A = (a_1, a_2, a_3, a_4)_T$.

It is worth noting that a sum of trapezoidal fuzzy numbers is also a trapezoidal fuzzy number, i.e., if

$$A = (a_1, a_2, a_3, a_4)_T \text{ and } B = (b_1, b_2, b_3, b_4)_T \text{ then } A + B = (a_1 + b_1, a_2 + b_2, a_3 + b_3, a_4 + b_4)_T.$$

Before we take the next step we have to define some basic operations on fuzzy numbers. Although one can introduce these operations directly on membership functions it seems that it is easier to do this equivalently as α -cut-wise operations on intervals. In particular, the sum of two fuzzy numbers A and B is given by the Minkowski addition of the corresponding α -cuts (see Fig. 4), i.e. for all $\alpha \in [0, 1]$

$$(2) \quad (A + B)_\alpha = [\inf A_\alpha + \inf B_\alpha, \sup A_\alpha + \sup B_\alpha].$$

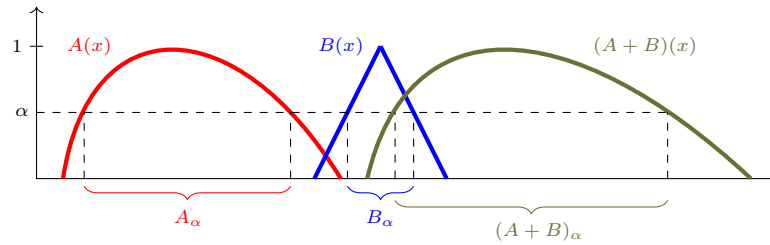


Fig. 4. Addition of fuzzy numbers A and B .

The product of a trapezoidal fuzzy number $A = (a_1, a_2, a_3, a_4)_T$ by a scalar θ is a trapezoidal fuzzy number, i.e.

$$\theta \cdot A = \begin{cases} (\theta a_1, \theta a_2, \theta a_3, \theta a_4)_T & \text{if } \theta \geq 0, \\ (\theta a_4, \theta a_3, \theta a_2, \theta a_1)_T & \text{if } \theta < 0. \end{cases}$$

Similarly, the product of a fuzzy number A by a scalar $\theta \in \mathbb{R}$ is defined by the Minkowski scalar product for intervals (see Fig. 5), i.e. for all $\alpha \in [0, 1]$

$$(3) \quad (\theta \cdot A)_\alpha = [\min\{\theta \inf A_\alpha, \theta \sup A_\alpha\}, \max\{\theta \inf A_\alpha, \theta \sup A_\alpha\}].$$

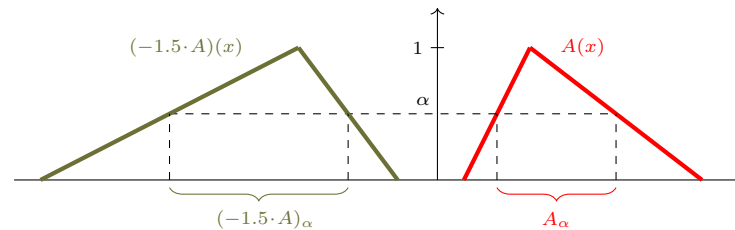


Fig. 5. The product of a fuzzy number A by a scalar.

Unfortunately, in general, $A + (-1 \cdot A) \neq \mathbb{1}_{\{0\}}$ (see Fig. 6). Consequently, the Minkowski-based difference does not satisfy, in general, the addition/subtraction property that $(A + (-1 \cdot B)) + B = A$.

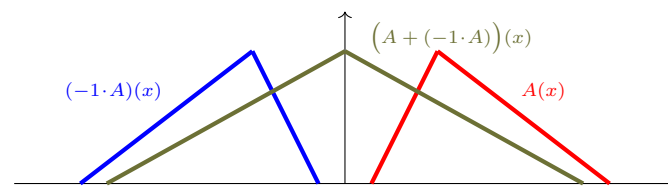


Fig. 6. Problems with subtraction of fuzzy numbers.



Solution to Problem M 1752.

Let us observe that for any $M, a, b \in [0, 1]$ such that $M \geq a, b$, the following inequality holds:

$$(M - a)(M - b)(1 - bM) \geq 0,$$

After some transformations, we obtain:

$$(1 - bM + b^2)(1 - aM + M^2) \geq 1 - ab + b^2. \quad (4)$$

To prove the inequality for $n = 2$, it is sufficient to take $a = b = x_1$ and $M = x_2$ in the above inequality.

Assume that the inequality holds for some n ; we will deduce its validity for $n + 1$. Without loss of generality, assume that $x_{n+1} = \max\{x_1, \dots, x_{n+1}\}$. Using the above inequality with $b = x_n$, $M = x_{n+1}$, and $x_1 = a$, we obtain:

$$(1 - x_n x_{n+1} + x_n^2)(1 - x_{n+1} x_1 + x_{n+1}^2) \geq (1 - x_n x_1 + x_n^2).$$

Therefore,

$$\prod_{\text{cycle}}^{n+1} (1 - x_i x_{i+1} + x_i^2) \geq \prod_{\text{cycle}}^n (1 - x_i x_{i+1} + x_i^2) \geq 1,$$

where “cycle” denotes the cyclic product. We complete the proof by invoking the principle of mathematical induction.

To overcome some of the problems associated with the lack of a satisfying difference, especially in constructing tools for statistical reasoning based on fuzzy observations, an alternative approach utilizing distances is often considered. Let us define the following distance between two fuzzy numbers A and B

$$D(A, B) = \sqrt{\int_0^1 [(\inf A_\alpha - \inf B_\alpha)^2 + (\sup A_\alpha - \sup B_\alpha)^2] d\alpha}.$$

Indeed, (4) defines a *metric* (in the sense explained in details in e.g. Jarosław Górnicki’s article from *Delta* 5/2021). It is clear that $D(A, B) \geq 0$ and $D(A, B) = 0$ if and only if $A = B$. Proving that $D(A, B) + D(B, C) \geq D(A, C)$ (triangle inequality) is slightly less trivial and we leave it as an exercise to the reader.

Suppose, we observe independently two fuzzy random samples $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_m)$ drawn from two populations (each x_i and y_i is a fuzzy number). We want to check if there is a significant difference between these two populations. To this end we measure the distance (4) between the arithmetic means of these samples. Note that we already know how to compute an arithmetic mean of fuzzy numbers (which is itself a fuzzy number) as we have tools of adding them together and multiplying by real numbers. But is a specific distance between means, like 3.14, large or small? This is where the statistics come into the picture.

In statistical jargon our goal is to verify the null hypothesis H_0 that both samples come from the same distribution, against the alternative hypothesis that the population distributions differ. If the null hypothesis holds we expect that both sample means would not differ too much. On the other hand, a significant difference between the two sample means may indicate that the samples under study come from different distributions.

To decide whether the measurement is large enough to conclude as significant statisticians often use the notion of *p-value*. In our case it is the probability, under the assumption that the null hypothesis is true, of getting at least as great distance between means as the distance observed. Intuitively, if this probability is low, we have a good reason to reject the null hypothesis. The problem is that in this case we cannot compute it exactly as the null hypothesis merely says that the two populations can be treated as one but it does not give us a specific description of this population! We need to resort to another clever idea.

Let \mathbf{v} be the concatenation of the two samples, i.e. $v_i = x_i$ if $1 \leq i \leq n$ and $v_i = y_{i-n}$ if $n + 1 \leq i \leq N$, where $N = n + m$. Now, let \mathbf{v}^* denote a permutation of the initial dataset \mathbf{v} . Then the first n elements of \mathbf{v}^* are assigned to the first sample \mathbf{x}^* and the remaining m elements to \mathbf{y}^* . In other words, it works like a random assignment of elements into two samples of the size n and m , respectively. Each permutation corresponds to some relabeling of the combined dataset \mathbf{v} . Please note that if H_0 holds, i.e. both samples come from the same distribution, then we are completely free to exchange the labels x or y attributed to particular observations – this will not change the randomness behind them. As a consequence we can *estimate* the true p-value from the data by taking a fraction of all possible permutations \mathbf{v}^* that yield a larger distance between means of \mathbf{x}^* and \mathbf{y}^* than the one observed.

Formally this can be expressed as

$$(5) \quad \text{p-value} = \frac{1}{N!} \sum_{\mathbf{v}^*} \mathbb{1}(T(\mathbf{v}^*) \geq t_0),$$

where the sum ranges over all possible permutations \mathbf{v}^* of \mathbf{v} , $T(\mathbf{v}^*)$ is the distance between means of \mathbf{x}^* and \mathbf{y}^* and t_0 is the observed distance between

Equation (5) can be treated as a definition of the *true* p-value, conditioning on the event that our samples sum up (as sets) to \mathbf{v} . This approach is somewhat standard in designing so called *nonparametric tests*, like *runs test*, described in details in an article of the same name in *Delta* 9/2017.

means of \mathbf{x} and \mathbf{y} . The value of $\mathbb{1}(\text{condition})$ is 1 if condition is met and 0 otherwise.

Formula (5) can be further simplified, using the fact that the permutations can be split into groups of size $n!m!$ each, which give the same means of \mathbf{x}^* and \mathbf{y}^* (those groups are permutations that can be obtained by each other by permuting first n and last m observations). Even in this case number of summands (which becomes $\binom{N}{n}$) grows exponentially with N (given that n/N is kept at fixed level).

Therefore, instead of considering all possible permutations we consider an approximate distribution obtained by drawing randomly a large number of samples (permutations) with replacement.

Let $\mathbf{v}_1^*, \mathbf{v}_2^*, \dots, \mathbf{v}_K^*$ be some random permutations of \mathbf{v} (where K is usually not smaller than 1000). Then the approximate p-value of our test is given by

$$(6) \quad \text{p-value} \simeq \frac{1}{K} \sum_{k=1}^K \mathbb{1}(T(\mathbf{v}_k^*) \geq t_0).$$

Example. Now we utilize some data given in Ramos-Guajardo et al. (2019) to compare the opinions of the two experts about the overall impression of the Gamonedo cheese. The trapezoidal fuzzy sets corresponding to their opinions are gathered in Table 1. There we have two observations of independent samples $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_m)$ of sizes $n = 40$ and $m = 38$, respectively. Numbers in parentheses correspond to the notation used to describe trapezoidal fuzzy numbers, e.g. $x_1 = (65, 75, 85, 85)_T$, $y_1 = (50, 50, 63, 75)_T$, etc. Our problem is to check whether there is a general agreement between these two experts. To reach the goal we verify the following null hypothesis H_0 stating there is no significant difference between experts' opinions, against that their opinions on the cheese quality differ.

Simple calculations on data from Table 1 lead to means

$$\bar{x} = (57.65, 63.20, 69.18, 73.48)_T \quad \text{and} \quad \bar{y} = (47.34, 51.21, 59.87, 66.84)_T.$$

Substituting these results into (4) we obtain a value of our test statistic $t_0 = D(\bar{x}, \bar{y}) = 7.96$. Then, after combining samples and generating $K = 1000$ random permutations and following (6) we obtain the approximation of p-value of 0.002. Its interpretation is shown in Fig. 7, where one can find the histogram of all sampled differences $D(\bar{x}^*, \bar{y}^*)$. Black dot indicates the value t_0 of the test statistic. The barely seen grey area on the right side of this dot corresponds to the probability of obtaining the distance between \bar{x}^* and \bar{y}^* not smaller than t_0 . Therefore, we can rather confidently reject the null hypothesis and conclude that there is no general agreement between experts' opinions on the overall impression of the Gamonedo cheese.

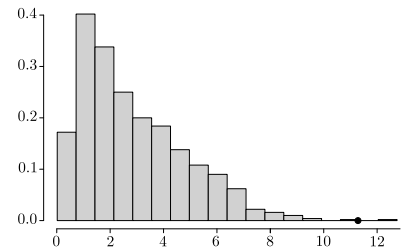


Fig. 7

The permutation agreement test considered above for two samples containing imprecise information is just one example of how fuzzy modeling can be combined with statistical inference. Although initially, some statisticians were skeptical about attempts to combine both theories, researchers realized that both statistics and fuzzy set theory should not be regarded as competitive, but that they can complement each other effectively. Moreover, expanding statistics with fuzzy sets not only solves some issues but also raises new questions. In particular, the distinction between the so-called ontic and epistemic sets yields different definitions of concepts as basic as variance and, consequently, different inferential tools. It is also worth noting that statisticians have also recognized fuzzy sets as convenient means for constructing procedures that allow the weakening of hypotheses or requirements that are excessively rigid.

Table 1. Opinions of two experts concerning the overall impression on the total of 78 samples of the Gamonedo cheese, cf. Ramos-Guajardo et al. (2019) Each entry refers to a different sample.

Expert 1	Expert 2
(65, 75, 85, 85)	(50, 50, 63, 75)
(35, 37, 44, 50)	(39, 47, 52, 60)
(66, 70, 75, 80)	(60, 70, 85, 90)
(70, 74, 80, 84)	(50, 56, 64, 74)
(65, 70, 75, 80)	(39, 45, 53, 57)
(45, 50, 57, 65)	(55, 60, 70, 76)
(60, 66, 70, 75)	(50, 50, 57, 67)
(65, 65, 70, 76)	(65, 67, 80, 87)
(60, 65, 75, 80)	(50, 50, 65, 75)
(55, 60, 66, 70)	(50, 55, 64, 70)
(60, 65, 70, 74)	(39, 46, 53, 56)
(30, 46, 44, 54)	(19, 29, 41, 50)
(60, 65, 75, 75)	(40, 47, 52, 56)
(70, 75, 85, 85)	(54, 55, 65, 76)
(44, 45, 50, 56)	(59, 65, 75, 85)
(51, 56, 64, 70)	(50, 52, 57, 60)
(40, 46, 54, 60)	(60, 60, 70, 80)
(55, 60, 65, 70)	(50, 54, 61, 67)
(80, 85, 90, 94)	(40, 46, 50, 50)
(80, 84, 90, 90)	(44, 50, 56, 66)
(65, 70, 76, 80)	(60, 64, 75, 85)
(75, 80, 86, 90)	(54, 56, 64, 75)
(65, 70, 73, 80)	(50, 50, 60, 66)
(70, 80, 84, 84)	(44, 46, 55, 57)
(55, 64, 70, 70)	(59, 63, 74, 80)
(64, 73, 80, 84)	(49, 50, 54, 58)
(50, 56, 64, 70)	(55, 60, 70, 75)
(55, 55, 60, 70)	(44, 47, 53, 60)
(60, 70, 75, 80)	(19, 20, 30, 41)
(64, 71, 80, 80)	(40, 44, 50, 60)
(50, 50, 55, 65)	(50, 50, 59, 66)
(50, 54, 60, 65)	(50, 53, 60, 66)
(65, 75, 80, 86)	(50, 52, 58, 61)
(50, 55, 60, 66)	(60, 65, 72, 80)
(40, 44, 50, 50)	(50, 50, 55, 60)
(70, 76, 85, 85)	(30, 34, 43, 47)
(44, 50, 53, 60)	(19, 25, 36, 46)
(34, 40, 46, 46)	(53, 63, 74, 80)
(40, 45, 51, 60)	
(84, 90, 95, 95)	

Magic fives

Oskar SKIBSKI*

* Faculty of Mathematics, Informatics and Mechanics, University of Warsaw

† Editor's note: The (erroneous) opinions expressed in this article are the author's own and may not necessarily coincide (and they do not) with the views of *Delta*.

We all know that magic exists. Scientists try to explain most phenomena by mathematical equations or create new definitions to pretend that what cannot be described by equations also makes sense. However, we know very well that they succeed in convincing only those already convinced. If the rabbit is not in the hat, and the magician pulls it out of the hat, then there is no equation that could describe it. One is one, and zero is zero.†

Since we in *Delta* consistently pretend that magic does not exist, we will discuss in this article an algorithm that appears to be magical, and yet is not. We can compare it with a magic trick in which the magician simply pulls a rabbit out of a hat, but does not show the audience beforehand that the hat is empty. For a moment it may surprise us (“Oh! Rabbit!”). However, with careful analysis and logical reasoning this trick can be explained: the rabbit must have been sitting on the magician's head throughout the performance, and while removing the hat the magician gently hooked the rabbit's legs and thus had something to pull out of the hat.

Magic trick: We have a hat with n rabb. . . no no, let's be at least a little serious! We have a set A containing n numbers (some numbers may repeat). One of the spectators says a number k from 1 to n . The magician in *linear time* finds the k -th largest number in the set A .

How to do the trick: For some values of k this task is very simple, e.g. for $k = 1$ our problem boils down to finding the smallest element in the set, which can be easily done in linear time. However, in general (e.g., if $k = \lceil n/2 \rceil$) it is not clear how we can do this. The case of $k = \lceil n/2 \rceil$ is, by the way, very important for statisticians, as it concerns the so-called *median*, which can have even more charm for them than the average. The natural idea is to sort all the elements and then point to the one in the k -th position. However, the fastest sorting algorithms run in $O(n \log n)$ time. Our problem seems much simpler than sorting all elements – how to solve it in linear time?

We will borrow the idea for our solution from the magic trick of a woman sawn in half, or if you prefer, from *Hoare's algorithm*. Take a random element m and divide our entire set into two non-empty sets so that the first set contains only elements less than or equal to m (set A_{\leq}), and the second set: greater than or equal to m (set A_{\geq}). In order to make the sets non-empty, we can, for example, put all the elements less than or equal to m into the first set, and if the second set turns out to be empty put m into it. Now, if A_{\leq} has at least k elements, then the element we are looking for must be in A_{\leq} – so we recursively search for it in this set. On the other hand, if the set A_{\leq} has less than k elements, then the element we are looking for is in the set A_{\geq} : we must therefore recursively find $(k - |A_{\leq}|)$ -th largest element there.

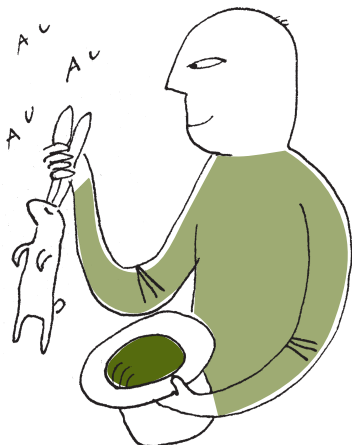
The idea is simple, but it may not be very efficient – if we are unlucky to always draw the smallest or the largest element, in every step our set A will decrease by only one element. And since each step requires linear time, the pessimistic running time of our algorithm will be quadratic: $O(n^2)$. That is even worse than with sorting!

However, our algorithm can be improved by changing the element based on which we divide the set. In the card trick, to find the card chosen by a spectator, the magician usually does not rely on fate, but carefully shuffles the deck to control where that card is. We will do the same – we will shuffle the elements a bit and draw one that guarantees that none of the parts are too big.

This is how the *median of medians* algorithm (in Polish called *the magic fives algorithm*) works. Let us divide arbitrarily our set into fives of elements (the last five can be incomplete) and for each find its median. Now, using our algorithm recursively, we find the median of the medians: we will denote it by m . Now, we divide our set into three parts: elements smaller than m (denoted by $A_{<}$), elements equal to m (denoted by $A_{=}$) and elements larger than m (denoted

When we write that the running time of $f(n)$ is $O(g(n))$ we mean that f is at most of the order of g , that is, for large values of n the function $f(n)$ grows no faster than $g(n)$. Formally: there exist constants c and n_0 that for $n \geq n_0$ it holds $f(n) \leq c \cdot g(n)$. Thus, this is an upper bound estimate, but usually not the best one that can be found

To some readers, this may resemble the QuickSort algorithm. In the QuickSort algorithm, in order to sort a set of numbers, we divide it into elements smaller and larger than a certain element, and then sort both parts. It is no coincidence that the author of the QuickSort algorithm is also Tony Hoare.





Solution to Problem F 1076.

If we neglect boundary effects, a uniform electric field perpendicular to the surface of the plates will appear between them after charging. The charge densities will also be uniform. Let's assume that the total surface area of each plate is S , then the charge densities will be: $\sigma_1 = Q_1/S$ and $\sigma_2 = Q_2/S$. Based on Gauss's law, we find that inside the capacitor, each plate is the source of an electric field with intensity:

$$E_i = \frac{\sigma_i}{2\epsilon_0}.$$

In the above formula, ϵ_0 represents the vacuum permittivity. The electric field is directed "away from the plate" if its charge is positive and "towards the plate" if it is negative. The resultant electric field inside the capacitor is the sum of the fields from both plates and is given by:

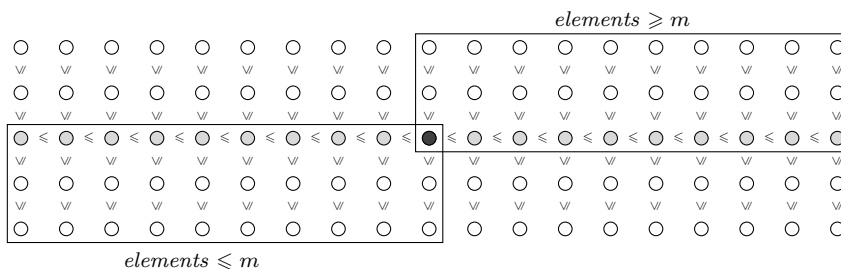
$$E = E_1 - E_2 = \frac{\sigma_1 - \sigma_2}{2\epsilon_0} = \frac{Q_1 - Q_2}{2\epsilon_0 S},$$

and the value of the potential difference is:

$$U = Ed = (E_1 - E_2)d = \frac{(Q_1 - Q_2)d}{2\epsilon_0 S} = \frac{Q_1 - Q_2}{2C}.$$

by $A_>$). Now if $|A_<| \geq k$, we recursively look for the k -th largest element in it. If $|A_<| < k$, but $|A_<| + |A_|= \geq k$, it means that m is the k -th largest element. If, on the other hand, $|A_<| + |A_|= < k$, then we have to look recursively for our element in the set $A_>$ - it is the $(k - |A_<| - |A_|=)$ -th largest there. Voila!

Explanation of the trick: Okay, but how can we be sure that none of the parts will be too big and the running time will be linear? Note that in half of the fives, the median is less than or equal to m . In each such five, at least half of the elements are less than or equal to m . It follows that at least 1/4 of all the elements are less than or equal to m . This also means that no more than 3/4 of all the elements are greater than m . Similarly, no more than 3/4 of all the elements are smaller than m . This means that no matter which case occurs, we will recursively call our algorithm on a set reduced by at least 25%.



Is this enough to make our algorithm run in linear time? It turns out that it is! Let us denote the running time of our algorithm by $T(n)$. Dividing into fives and selecting from each the median can be done in linear time - $O(n)$. Selecting the median of the medians will take us $T(\lceil \frac{n}{5} \rceil)$ time. The last step is a recursive call for any of the parts. As we have already showed, none of the parts has more than 3/4 of all elements, so the time is $T(\lceil \frac{3n}{4} \rceil)$. Thus, we get the following upper bound:

$$T(n) \leq O(n) + T\left(\left\lceil \frac{n}{5} \right\rceil\right) + T\left(\left\lceil \frac{3n}{4} \right\rceil\right).$$

As it turns out this upper bound implies the running time of our algorithm is linear! The key here is the inequality $\frac{1}{5} + \frac{3}{4} = \frac{19}{20} < 1$, which ensures that $T(n)$ does not grow too fast.

To see this, first assume that n is a power of 20 so that it divides nicely by 4 and 5. We know that the time needed to divide into fives and pick medians can be bounded by cn for some natural c . It is easy to show by induction that $T(n) \leq 20cn$:

$$T(n) \leq cn + T\left(\frac{n}{5}\right) + T\left(\frac{3n}{4}\right) \leq cn + 20c\left(\frac{n}{5}\right) + 20c\left(\frac{3n}{4}\right) = 20cn.$$

From the upper bound for powers of 20, we immediately get an upper bound for other numbers: Take any n and choose k so that $20^k < n \leq 20^{k+1}$. Since T is a non-decreasing function, we know that $T(n) \leq T(20^{k+1})$. But since $n > 20^k$ we get that

$$T(n) \leq T(20^{k+1}) \leq 20c \cdot 20^{k+1} \leq 20^2 cn.$$

Since there is a constant $20^2 c$ such that for any n there exists $T(n) \leq 20^2 c \cdot n$ this means that our algorithm runs in linear time: $O(n)$.

Finally, it remains to ask the question - why fives? It seems natural to take an odd number (so that there are middle values), but could we take threes or sevens? Or elevens?

It turns out that we could not take threes: we would first have to find the median of 1/3 of all elements, and this would only reduce the problem by 1/4. As $1/3 + 1/4 < 1$, that would result in $O(n \log n)$ time. We could instead take sevens, nines, etc: Looking for the median among 1/7 or 1/9 of all the elements would be even faster than looking for the median of 1/5. However, we would increase the cost of finding the medians (i.e., that enigmatic c in the above proof) and the implementation would be more complex. So fives are used in the algorithm, because five is the smallest odd number greater than four. And that's the whole magic.

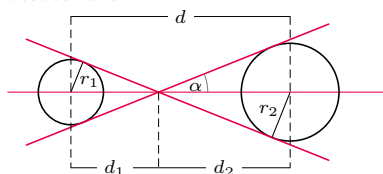


Solution to Problem F 1075.

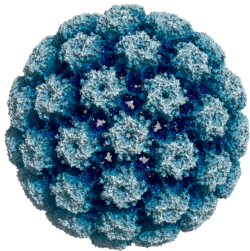
Observations from directions between the two tangents to the surface of the stars (as shown in the diagram) are associated with the occurrence of eclipses. These tangents intersect at a point dividing the segment d into segments d_1 and d_2 , $d = d_1 + d_2$. This corresponds to the range of observation angles α for which the inequality holds:

$$|\sin \alpha| \leq \frac{r_1}{d_1} = \frac{r_2}{d_2} = \frac{r_1 + r_2}{d}.$$

The model considered in the task describes the simplest case of a binary system. In general, gravitational interactions between the stars can lead to deformations from spherical to ellipsoidal shapes or even to the flow of matter between them.



“Życie na żywo” (Polish for “Live life”) is a single author column series in *Delta* that concerns life sciences.



HPV type 16 viral capsid protein.
 Attribution: Opabinia regalis, CC BY-SA 4.0 via Wikimedia Commons

Virus means poison

The history of the discovery of viruses, forms of life that require a host cell (bacterial, animal, or plant) to survive, is not too distant. It was with surprise that Louis Pasteur, in the late 19th century, observed the virus causing rabies, realizing that an infectious agent could be isolated by filtering tissue extracts through porcelain filters that retained bacteria. This was the first indication that the rabies virus must be smaller than bacteria. With the invention of electron microscopes, we were able to see various viruses and learn about their shapes. This device turned out also incredibly useful in studying the fundamental processes of large molecules necessary for life, such as proteins, nucleic acids, and their structural and metabolic interrelationships. The word “virus” is widely known, although detailed knowledge about different viruses is accessible only to specialists. From time to time we come across news in newspapers about viruses and our relationship with them (pandemics!), and then it is useful to listen to what the virologists have to say.

At the end of May, registration for non-compulsory and free vaccinations against human papillomaviruses (HPV) for 12 and 13-year-old children has begun in Poland. This vaccine is an exceptional medical achievement. HPV viruses cause morphological changes in the squamous epithelial cells of the cervix, which can eventually transform into squamous cell carcinoma. The discovery that cancer may be virus-induced led to the successful development of the vaccine. It was designed 17 years ago. Clinical safety studies were conducted under the supervision of the European Medicines Agency. There are also data on the effectiveness of the vaccine: according to a study of more than 2,000 women from Scandinavian countries, the incidence of cervical cancer decreased by 90% compared to the number of cases in the period when there were no vaccinations. During the 8 years of the study, there was not a single new case of this disease among the participants, which proves its high effectiveness. Universal vaccination was first introduced in Australia, in 2007, followed by the United Kingdom in 2008, and by 2019, almost all European Union countries had implemented such free vaccinations for both girls and boys. These vaccinations are now used in 125 countries worldwide (data as of May 2023).

There are over 150 types of HPV known, including low-risk types that cause benign genital warts and high-risk types responsible for precancerous changes, cervical cancer, and other malignancies. HPV infection usually occurs through sexual contact, most commonly in the early stages of sexual activity. Throughout their lives, 80% of sexually active women and men have been or will be infected with HPV. HPV infections can also lead to anal, vaginal, vulvar, penile, oral, and head and neck cancers. The three most common highly oncogenic types of HPV are HPV-16, HPV-18, and HPV-45, which are believed to be responsible for 80-90% of the aforementioned cancers. For every one million women infected with oncogenic HPV, 8000 develop cancer, and the mortality rate for cervical cancer is 50%.

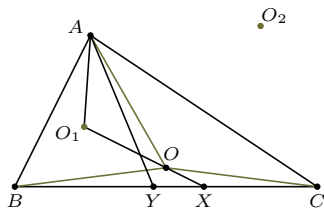
The vaccines constructed according to the same principle have the viral protein L1 as the antigen. Vaccines targeting different types of viruses consist of different (though chemically similar) L1 proteins specific to each virus type. These proteins form virus-like shells that do not contain any genetic material and do not cause infections. They are produced using genetic engineering, not from propagation of real viruses.

HPV vaccines are the first medicinal products with primary purpose to prevent cancer. For me, it’s a revelation!

Magdalena FIKUS (magda.fikus@gmail.com)



Solution to Problem M 1751.



Let O_1 and O_2 be the centers of the circumcircles Ω_1 and Ω_2 of triangles AOB and AOC respectively. Note that O_1 lies on the line OX , hence we have:

$$\sphericalangle AO_1X = 2\sphericalangle ABO = \pi - \sphericalangle AOB = \pi - 2\sphericalangle ACB = \sphericalangle AXY,$$

thus O_1 lies on the circumcircle of the triangle AXY . Similarly, we can prove that O_2 lies on that circle.

All models are wrong, so which ones are useful?

* MI2.AI, University of Warsaw and Warsaw University of Technology

G. Box, "Science and statistics", Journal of the American Statistical Association (1976)



F. Anscombe, "Graphs in Statistical Analysis", American Statistician (1973)

Przemysław BIECEK*

George Box's quote "All models are wrong, some are useful," is a well-known phrase among statisticians. It acknowledges that it is impossible to create a perfectly accurate model to describe reality but that imperfect models can still be useful in real-world applications. Over the decades, various data-driven models have been developed using this philosophy and formulated to answer a variety of questions. Does the analyzed therapy produce positive medical outcomes? Does investment in education translate into student performance? These are just two examples of research questions that can be verified with data-driven models. Such approaches are the foundation of all empirical sciences.

Despite acknowledging the validity of Box's statement, an important question remains unanswered: how do we determine which models are useful? The answer is *some* but apparently not all of the models. Choosing the appropriate model or models is a critical decision. The conventional approach is to select some model quality criterion, usually based on how well the model fits the analyzed data, and then choose the model that best satisfies this criterion. There are several model quality criteria used by statisticians, including RMSE, R^2 , AIC, BIC (we won't give their proper definitions here as they are of no importance to this article). In the machine learning community, criteria based on predictive performance on a new independent data are more common. However, the general procedure remains unaltered: start with a group of candidate models, select the best one according to a specific criterion and consider it the most accurate description of reality. From there we begin our inference.

Such approaches sometimes lead to surprises and interesting paradoxes. One of them is *Anscombe's quartet*, introduced (precisely!) 50 years ago. Anscombe created four artificial sets of data, each consisting of 11 pairs of real numbers (like height and weight of eleven newborns). All those datasets have the same best-fitted linear model (in the sense of R^2) with the same value of R^2 . Yet each set of data tells a completely different story. To understand the nature of the relationship between variables, visualization of the data is essential, as in Figure 1.

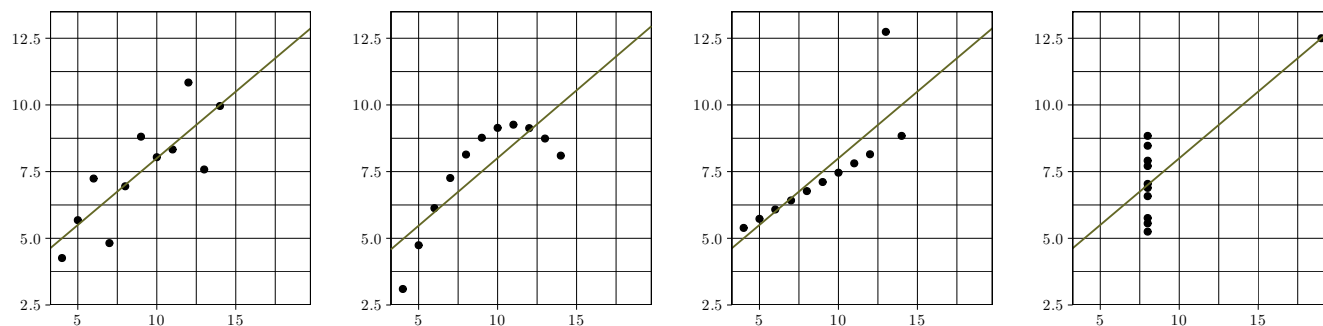


Fig. 1. Anscombe's quartet. For each dataset the model $y = x/2 + 3$ is the best linear fit and has similar fit to the data with the coefficient of determination $R^2 = 0.66$

J. Tukey, "Exploratory Data Analysis", Pearson (1977)

L. Breiman, "Statistical Modeling: The Two Cultures", Statistical Science (2001)

As a consequence, a linear model that fits the data well is not enough to infer the relationship between variables accurately. Anscombe's solution to this paradox is to visualize the data, even with basic methods such as a scatterplot. Visual analysis can complement statistical inference in such cases and many well-known statisticians have proposed new methods of data visualization, which today are referred to as Exploratory Data Analysis tools.

Anscombe demonstrated that different data sets can have the same model fit equally well but present entirely different stories. However, can the opposite be true? Can one dataset have several models with different stories that produce the same fit? Surprisingly the answer is positive. It was pointed out in 2001 by Leo Breiman in his influential paper "The Two Cultures". This quality is now

known as the “Rashomon perspectives” or “the multiplicity of good models” and it continues to exist in the foundations of statistical modeling in today’s world, which is increasingly reliant on such models. The name “Rashomon” refers to a 1950 movie by Akira Kurosawa, in which an event is described from the perspective of four witnesses, each offering a different account of what had happened. Their stories vary so much that it is impossible to tell what is the truth. Breiman used this term to describe a hypothetical scenario in which several models have an equally good fit to the data, but they offer different explanations for the data. Such a situation would call into question any inference based on the “single best” data-driven model. For instance, what should one do facing two models that claim conflicting results about the effectiveness of a medical therapy? Which of these models should be trusted if they both have an equally good fit to the data?

In order to illustrate the multiplicity of good models problem, Breiman used several linear models with the same fit to the data (hardly different from the best possible linear fit). At the same time those model lead to different conclusions regarding the dependence between variables (e.g. is the increase of one variable followed by the increase or the decrease of another variable?) To make this phenomena even more striking, in our recent paper we introduced the *Rashomon’s quartet*. The paper presents a regression tree, a random forest, a neural network, and a linear model (I will not delve into the details of these models here, for this article they are not important). All these models were fitted to the same data, resulting in the same predictive performance, but it turns out that each model is describing an entirely different story.

P. Biecek, H. Baniecki, M. Krzyżiński, and D. Cook, “Performance is not enough: a story of the Rashomon’s quartet” arxiv (2023)

... but wait! How do we know what stories are depicted by such complex models as a neural network or a random forest with hundreds of trees? Visualization techniques for predictive models developed under the name eXplainable Artificial Intelligence (XAI) or Explanatory Model Analysis (EMA) come to our aid. One of them is Partial Dependence (PD), a technique proposed by Jerome Friedman in his famous work on the boosting method. PD is a model agnostic method, meaning that it can be used to analyze any predictive model regardless of its complexity or structure. Due to this universality, this method has quickly found many applications.

J. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine”, Annals of Statistics (2000)

Let us introduce the intuition behind the Partial Dependence profile. The value of PD of any given variable S and its potential value t is equal to an average prediction of the model for the available data in which the value of the variable S is “artificially” set to t .

See also P. Biecek, T. Burzykowski “Explanatory Model Analysis” CRC Press (2021) or <https://ema.drwhy.ai/>

In order to describe it more formally let us assume that we have N observations of p input variables (e.g. for N newborns we observe weight, height, eye color...), where the i -th observation is $(x_{i,1}, x_{i,2}, \dots, x_{i,p})$. These variables are used to predict *something* (this *something* is usually called the *target variable*) and in order to do so we apply the function $f(x_1, \dots, x_p)$, which is our *model*. The Partial Dependence of the s -th variable is the function defined by the formula

$$PD^s(t) = \frac{1}{N} \sum_{i=1}^N f(x_{i,1}, \dots, x_{i,s-1}, t, x_{i,s+1}, \dots, x_{i,p}).$$

In essence, the PD shows how the predicted value of the target variable changes as the value of the predictor variable(s) of interest varies, while holding all other predictor variables constant at their observed values in the dataset.

The PD response profiles of each model in the Rashomon quartet are displayed in Figure 2, revealing distinct relationships between the variables, particularly for variables X_2 and X_3 . For example, X_2 is insignificant in the first model but has positive effect in the other models. In contrast, X_3 is insignificant in the first model, has a negative effect in the second model and a positive effect in the third and fourth model. However, when all models have an equivalent fit to the data, it is challenging to determine which description is accurate.

Which model should we trust? If we don’t know, why should we trust any of them? Since, due to random fluctuations, each of these model can be considered

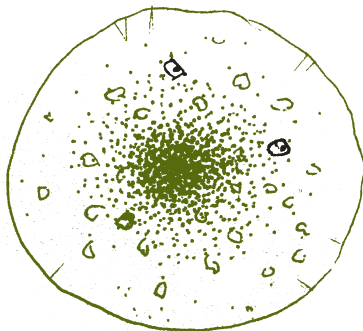
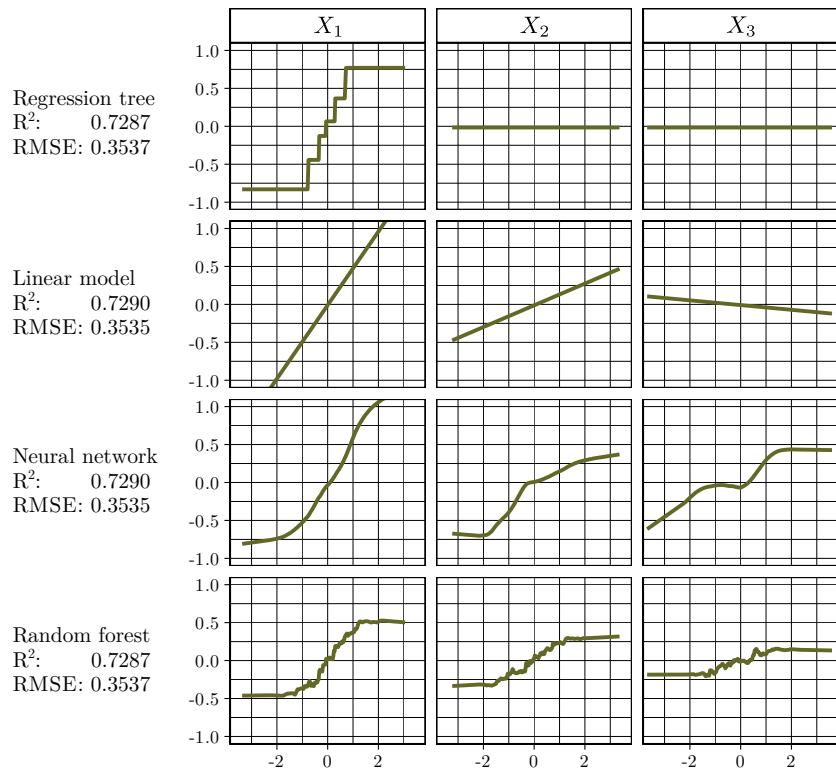




Fig. 2. Rashomon's quartet. Each row stands for a different model while columns stand for consecutive input variables. Panels show Partial Dependence profiles. These four models were fitted to a single dataset. The R^2 and RMSE fit coefficients presented show that all these models are equally well suited to the data analyzed.



as the best one, analyzing only the single best model, and disregarding the slightly inferior ones, may be a bad strategy.

PD profiles say a lot about the models, but they are not yet the ultimate solution to the model visualisation problem. Despite their universality, they have some limitations or shortcomings, e.g. correlations between variables or interactions can distort the picture presented by these profiles. Variable-by-variable analysis will be difficult if the model uses hundreds or thousands of variables.

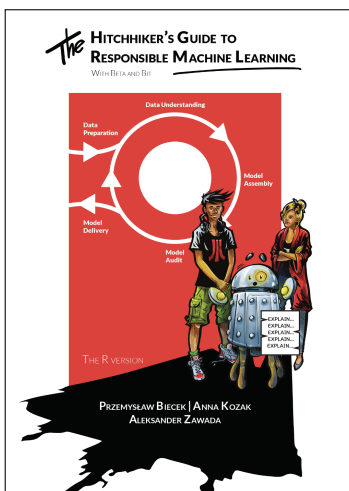
Characterization of a set of the best models is still an unresolved research problem, and different research groups are struggling with it. It is difficult but essential. To comprehend the world, we cannot depend solely on a single model's perspective, even if it has the best performance in relevant criteria. We must combine sets of models to differentiate between hypotheses supported by data and hypotheses that may result from the chosen predictive models.

The task of recognizing situations in which we can deal with the Rashomon perspective is yet to be solved. In the meantime, we can resort to thorough verification using a range of model visualization techniques. Anscombe's quartet has shown that data visualization techniques can be very useful for reasoning about nature of relations between variables. Similarly, Rashomon's quartet shows that model visualization techniques can be equally useful.

If you wish to explore the topic of model visualisation and comparison further, you may refer to resources such as the statistical comic book "The Hitchhiker's Guide to Responsible Machine Learning" available at <https://betaandbit.github.io/RML/>. Using the SARS-COV-2 mortality prediction as an example, it discusses the process of training predictive models as well as methods for exploring these models, including the best-known ones that is Partial Dependence, Shapley Values, and Ceteris Paribus. The examples presented above are for synthetic data. Similar challenges can be encountered when analyzing real-world problems as shown in the RML comic.

At the end of the day, we know that all models are wrong and we don't know which ones are useful. However, we can look at many good models at once. Looking at the world through the perspective of multiple models is essential to separate relationships truly supported by the data from relationships that are artifacts of the chosen modeling technique.

C. Rudin, Ch. Chen, Z. Chen, H. Huang, L. Semenova, Ch. Zhong "Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges" arxiv (2021)



A boy or a girl?

* Faculty of Mathematics, Informatics and Mechanics, University of Warsaw

Łukasz RAJKOWSKI*

Imagine, Dear Reader, that you are embarking on a long train journey. Just as you have comfortably settled in your seat and started reading your favourite magazine, you are interrupted by a cheeky exclamation: “Hey, what a coincidence! Long time no see, huh?” You raise your head, and your eyes meet a friendly face that, with a little effort of memory, you recognize as an old friend from elementary school.

“Yeah, long time indeed. . .” you reply, reluctantly closing your *Delta* magazine. “How have you been?” your friend courteously inquires.



Well, there comes a moment when you have to summarize several years of life in one sentence, so you respond, “Good! And how about you?” Your friend opens his mouth, and you already know that the question was a mistake. He **really** wants to summarize the past few years of his life, certainly not in one sentence. After an hour of narrating his school and professional adventures, the conversation shifts to family matters. At some point, you limit yourself to nodding politely, smiling, and the infallible “Oh boy!” while in your mind you are actually solving the Delta’s Problem Corner [in this issue on page 19, ed. note].

At some point, you realize that your friend has started talking about the difficulties of finding childcare for his child. The problem is that you missed the moment when he mentioned having a child at all.

Hmm, you think to yourself, I didn’t hear whether it’s a boy or a girl. It’s better to refer to them in a gender-neutral way; otherwise, I have about a 50% chance of an awkward mistake.

A similar issue is addressed in the song *Nie mam pojęcia* by Lona and Webber.

As the conversation continues, it turns out that your friend was seeking childcare not for one child but for two! Furthermore, you caught the sentence in his monologue, “I went for a walk with my son,” which establishes the gender of one of the children. The gender of the second child remains a mystery to you.

Ah, I once read something about this on the internet! you think. *At first glance, it would seem that the chance of the second child also being a boy is 50%. But we can look at it differently. If I forget the information I already have, the chance of my friend’s older child being a boy is 50%, and similarly, the chance of the younger one being a boy is also 50%. Therefore, the chance of having two boys is $50\% \times 50\% = 25\%$. Likewise, the chance of having two girls is 25%. There are two remaining possibilities, each with a probability of 25%: older boy, younger girl, and vice versa. I already know that my friend has a son, so out of these four equally likely situations, I can restrict attention to three (by excluding two daughters). And out of those three, only one corresponds to the situation of my friend having two sons, so the chance of that is $1/3$. With my back against the wall, I should bet on a daughter!*

	♂	♀
♂		
♀		

Fig. 1. The rows correspond to the gender of the older child, and the columns correspond to the gender of the younger child. The hatched area represents $1/3$ of the coloured region

To ensure the correctness of your reasoning, you discreetly took out a piece of paper and a pencil and sketched a convincing 2×2 table representing different combinations of the gender of the older and younger child (reproduced in Figure 1). Yes, it clearly shows that everything is correc. . .

(voiceover) No! Nothing is correct! As a matter of fact you are dealing with a situation where a randomly encountered person (the information that it is your acquaintance is irrelevant) happens to have two children and starts telling you about one of them, who turns out to be a boy. It is reasonable to assume that your acquaintance randomly selected one of his children (each with equal probability). If he is the father of two boys, he will definitely talk about a boy. If he is the father of a boy and a girl, he has a 50% chance of talking about a boy. In other words, out of all fathers of two children of both genders, only half of them would start talking to you about their son. Thus, your table should correspond to the one in Figure 2. In the end, from your perspective, the chance of your acquaintance having two sons remains 50%!

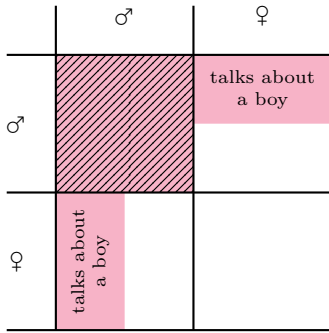


Fig. 2. The hatched area represents 1/2 of the coloured region

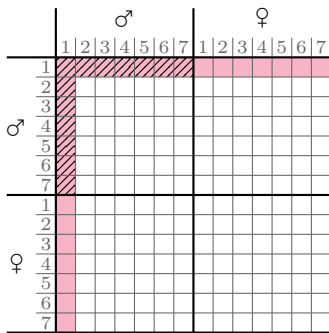


Fig. 3. The numbers correspond to the day of the week on which each child was born. The hatched area represents 13/27 of the coloured region.

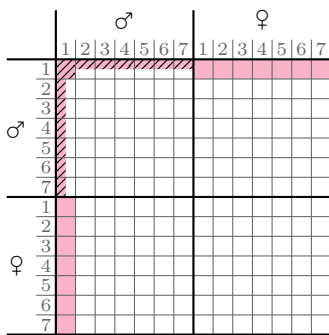


Fig. 4. The hatched area represents 1/3 of the coloured region

However, it would be a different scenario if you specifically asked your acquaintance, “Dear friend, please start telling me about one of your sons if you have at least one.” Assuming that your acquaintance would indeed begin a story (instead of quickly changing seats), the reasoning and table presented earlier in Figure 1 are correct, and the chance of him having two sons would indeed be 1/3.

For the purpose of the rest of the story, let’s assume that was the case.

The monologue of your acquaintance continues. Once again, you concentrate on something else. At some point, a piece of information breaks through your consciousness: the aforementioned son was born on a Monday, and since then, Monday has been your acquaintance’s favorite day of the week (lucky him). This news prompts you to engage in another mathematical reflection.

Should the information I just obtained change my estimation of the chances that my acquaintance has two sons? It might seem that it shouldn’t in any case since the day of the week on which his mentioned son was born cannot influence the gender of the other child in any way. However, on the other hand, this knowledge requires me to modify the previously sketched table. Now, for each child, the older and the younger, I should record not only their gender but also the day of the week on which they were born. Assuming that each day of the week is an equally likely birthday, this gives me 14 equally likely configurations of gender and day of the week for each child, resulting in a total of 196 configurations for both children ($14^2 = 196$). This time, 27 of them correspond to the occurrence of a boy born on Monday, and out of those, in 13 cases, the other child is also a boy. Therefore, my estimation of the probability that my acquaintance has two boys should increase from 1/3 to 13/27.

Because the situation intrigued you, you redrew your table, obtaining something similar to Figure 3 in the margin. And although you know from experience that tables are not to be debated, the conclusions you reached still bothered you.

(voiceover) And rightly so, as such reasoning replicates the mistake described earlier. This time, you should take into account the fact that if your acquaintance has two boys, with exactly one of them being born on Monday, he started talking about that specific child, but he could have talked about “the other one”, which effectively reduces the probability of such a configuration from your perspective. Considering this in the reasoning leads us back to the answer of 1/3 (as before), as can be seen by analyzing Figure 4. Again, if you were to ask whether your acquaintance has a son born on Monday, an affirmative answer would allow you to change the estimation of the chances of having two boys to 13/27, in line with the reasoning presented earlier (while a negative answer would decrease the estimation to exactly 30%, as we encourage you to verify independently).

After this information, the topic of children came to an end. Soon, after listening to a few more gripping stories from your acquaintance’s life, the train arrived at its destination – it did not surprise you when it turned out that both of you were getting off at the same station. On the platform, his wife was already waiting with a tousled rascal.

Question 1. How do you now assess the chances that your acquaintance has two sons?

As you were about to say your goodbyes, your old preschool friend, whom you haven’t seen in ages, appeared behind your school acquaintance! Before she could say anything, you shouted to her:

“Hey there, long time no see! Do you happen to have two children, at least one of whom is a boy born on Monday?”

Your friend was taken aback for a moment, then smiled cunningly and replied:

“Yes, I have two children, and none of them was born on Monday.”

Question 2. What are the chances that your preschool friend has two sons?

(Answers to the above questions can be found on page 14.)

New answers, new questions

Eleven years ago, in July, at the press conference at CERN, the discovery of the Higgs boson was announced. The blessing of middle age allows me to be both old enough to understand the significance of this discovery at the time, and young enough to remember the emotions that went with it. However, with important discoveries of elementary particles, it is a bit like with CSI movies – the audience is shown a (sometimes spectacular) murder, then specialists enter the action, documenting the event and trying to better understand what actually happened, beyond the reach of the cameras and the script.

Answers to the questions from page 13:

Question 1. In this situation, we should take into account that if the acquaintance had a son and a daughter, the mother would have only a 50% chance of bringing the son. Therefore, the table from Figure 4 in the article should be modified by narrowing the coloured, plain rectangles twice. This leads to an answer of 50%.

Question 2. The interpretation of this situation seems somewhat subjective, but in the author's opinion, it should be assumed that the friend did not answer our question ("Is at least one child a boy born on Monday?"), but two separate questions: "Do you have at least one son?" and "Do you have at least one child born on Monday?". If we had actually asked her those questions (in a sense we did when we chose "Monday"), we would estimate the chances of having two boys as $\frac{1}{3}$ (it is helpful to visualize this situation using Figure 3 from the article).

This is exactly what the Higgs boson has been subjected to for over a decade. Individual results, sometimes documented in these pages, confirmed with increasing accuracy that the experimentally determined properties of this particle are consistent with the predictions of the Standard Model of elementary particles. And theoretical physicists have been patiently waiting for some discrepancies to emerge that might indicate that the Standard Model needed to be extended.

Studying the Higgs boson is not an easy task. It is an uncharged particle and interacts only weakly, which means that the probability of its production is minuscule. It takes really high energies and large numbers of colliding particles to collect any statistically significant sample. Even in a collider as large as the LHC, only one such particle is produced every two seconds of the machine's operation. It is therefore not surprising that it takes years to determine the characteristics of some rare processes involving the Higgs boson.

An example of such a rare process is the decay of the Higgs boson into the Z boson, a massive, neutral particle that carries the weak force, and a photon known from electromagnetic interactions. The Standard Model predicts that in only 0.15% cases the Higgs boson decays this way. Even worse, the appearance of the Z boson also has to be detected somehow. It is not easy, because it is also an unstable particle, too short-lived to be directly observed. It is identified by its decay products - an electron-positron pair or a muon-antimuon pair, which occurs in 6.6% of decays. This means that assuming 100% efficiency of the detectors - and this is a senselessly optimistic assumption! - only one in ten thousand decaying Higgs bosons will produce a signal of interest that we could observe.

The importance of the decay of the Higgs boson into a Z boson and a photon is that, according to the Standard Model, it does not occur directly, but is – in a certain way – assisted by other particles present in nature, perhaps also those that have not yet been detected. Thus, if the measured probability of this decay were significantly different from the predictions of the Standard Model, this would be an argument for the existence of such new particles, and their masses are close to the energy scale achieved by colliding particles at the LHC.

To experimentally determine the probability in question, teams of physicists at the ATLAS and CMS detectors had to join forces. Using artificial intelligence methods to analyze the data, the researchers found it to be 0.34% with a measurement uncertainty of 0.11%. On the one hand, this may cause some concern, because the determined value is 2.2 times higher than the predictions, but on the other hand, the measurement uncertainties are still large enough that compliance with the Standard Model is not yet ruled out, although the probability of random statistical fluctuation of such an order is not very large and amounts to only 6%. We will probably have to wait a few more years to solve the riddle of what has actually been measured...

The ATLAS and CMS Collaborations. 2023. Evidence for the Higgs boson decay to a Z boson and a photon at the LHC. ATLAS-CONF-2023-025

Krzysztof TURZYŃSKI

Faculty of Physics, University of Warsaw

Searching for order

Andrzej DĄBROWSKI*

* Faculty of Mathematics and Computer Science, University of Wrocław

The full name is *Seminar on Applied Mathematics, General Section of the Department of Applied Mathematics of the State Mathematical Institute* [1].

The Ngandong skulls are a group of fossilized human skulls discovered in 1931 in the Solo River valley on the island of Java, Indonesia. These skulls are dated to approximately 100-200 thousands years ago and are considered to be remains of hominids of the species *Homo erectus*.

Jan Czekanowski (1882–1965) was an anthropologist and ethnographer. He studied anthropology, anatomy, ethnography, and mathematics at the University of Zurich. He participated in the *Seminar* in Wrocław (four times according to the Seminar’s chronicle).

In 1950, one of the autumn meetings of the *Seminar on Applied Mathematics* in Wrocław was dedicated to the Ngandong skulls. Anthropologists expected mathematicians to order these skulls chronologically based on Czekanowski tables *graciously sent to us by Prof. T. Henzel* (quotation from the chronicle of the Seminar).

Order, understood as linear, is expected in many fields. Various rankings are created (for schools, universities, political support). It is hard to imagine sports events without rankings based on measured time, distance or points (pentathlon). Indexes are constructed to compare objects that are often difficult to compare. In economics, stock market states are assessed by stock market indices (e.g. Dow Jones, WIG), and the price level is assessed by the inflation index. The health status of a population is assessed by the average life expectancy or infant mortality rate. Comparison methods based on rankings and indices involve assigning a number to objects. And as we know, numbers can easily be ordered.

It is more difficult when an object can be described by a set of many numbers, and even more challenging when assigning numbers to it is hard. Such a problem arises when ordering shells with various patterns found on an ancient landfill (this is the so-called seriation problem in archaeology). Anthropologists turned to mathematicians with a similar problem, ordering the Ngandong skulls.

Czekanowski tables, which were available to Professor Steinhaus’s team (Kazimierz Florek, Józef Łukaszewicz, Julian Perkal, and Stefan Zubrzycki), contained information on the degree of differentiation of each pair of skulls. This degree was expressed by the Euclidean distance in a seven-dimensional space of parameters representing the lengths of characteristic segments on the skull.

In general, such a table can be treated as a *discrepancy function*. This is any function $d(x, y)$ defined on pairs of elements from a given set X , which satisfies the conditions:

$$d(x, y) \geq 0, \quad d(x, y) = d(y, x) \text{ and } d(x, y) = 0 \Leftrightarrow x = y,$$

A special case of a discrepancy function is a distance function a metric. A metric is a discrepancy function that also satisfies the triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$ for all x, y, z .

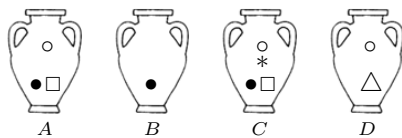


Fig. 1

Stanisław Kulczyński (1895-1975), zoologist, arachnologist, mountaineer. Rector of universities in Lviv and Wrocław.

Example 1. Vases. Four vases with decorations were discovered. Each vase is treated as a set of ornaments painted on it. The Kulczyński discrepancy function can be used as the discrepancy function, defined for two sets of ornaments R and S by the formula

$$d(R, S) = 1 - \frac{1}{2} \left(\frac{|R \cap S|}{|R|} + \frac{|R \cap S|}{|S|} \right).$$

It can be easily verified that such a function is indeed a discrepancy function. It is also quite intuitive, since $\frac{1}{2} \left(\frac{|R \cap S|}{|R|} + \frac{|R \cap S|}{|S|} \right)$ is the arithmetic mean of the fractions of common elements of R and S , contained in the set R and contained in the set S .

The resulting discrepancy matrix is presented in the margin. Note that the Kulczynski function is not a metric, since $d(B, D) = 1 > \frac{11}{12} = d(B, A) + d(A, D)$.

	A	B	C	D
A	0	1/3	1/8	7/12
B	1/3	0	3/8	1
C	1/8	3/8	0	5/8
D	7/12	1	5/8	0

Kulczyński discrepancy matrix.

Discrepancy function and linear order. How to introduce an order among objects for which we have a discrepancy matrix? Let us first consider a particular case. Assume that the objects are points distributed on the real line, and we take the discrepancy function to be the distance between them. Can we reconstruct the ordering of the points on the line from the discrepancy matrix alone? Yes, we can – we just need to choose an ordering that minimizes the sum of distances between consecutive points.

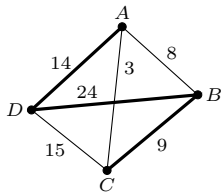


Fig. 2. Graph with Kulczyński discrepancies (multiplied by 24) and a path $ADBC$ of length 47.

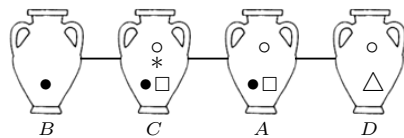


Fig. 3. Optimal path $BCAD$ in a graph with weights.

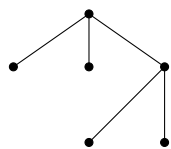


Fig. 4

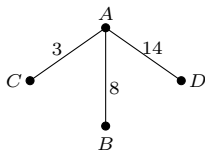


Fig. 5. Optimal dendrite for the problem of 4 weights. Its length (25) is smaller than the length of the optimal path $BCAD$ (26).

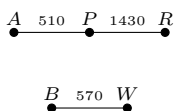


Fig. 6

It will be convenient for us to represent objects O_1, O_2, \dots, O_n with discrepancies $d_{ij} = d(O_i, O_j)$ by an undirected complete graph G with n vertices O_1, O_2, \dots, O_n , whose edge $O_i O_j$ has weight d_{ij} . An example of such graph is given in Fig. 2. We define the *length of the path* $O_{i_1} O_{i_2} \dots O_{i_n}$ as the sum $d_{i_1 i_2} + d_{i_2 i_3} + \dots + d_{i_{n-1} i_n}$.

Previous observations suggest that an order of objects O_1, \dots, O_n may be determined by the path in graph G of minimum length that passes through all its vertices (*an optimal path*).

Example 2. Five capitals. The table below shows the road distances (in kilometers) between Amsterdam, Berlin, Paris, Rome, and Warsaw.

	A	B	P	R	W
A	0	650	510	1650	1140
B	650	0	1040	1460	570
P	510	1040	0	1430	1550
R	1650	1460	1430	0	1730
W	1140	570	1550	1730	0

Out of 60 possible paths, the shortest one goes from Warsaw through Berlin, Amsterdam, Paris, and all the way to Rome, and has the length of 3160 km.

Dendrites. Linear ordering is often insufficient and even inadequate. Let us quote here Julian Perkal’s observation: *As I noticed, linear ordering is often unnatural in many cases, for example, a genealogical line often branches out.* [2]

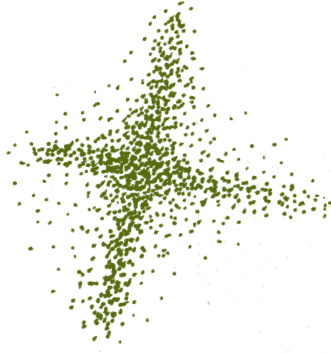
The structure that allows a nearly linear ordering of the vertices of a graph is a dendrite, more commonly known as a tree – for historical reasons, we will use the former term here. A dendrite is a graph without cycles, which is also connected. These two conditions together mean that any two vertices are connected by a uniquely determined path (Fig. 4). Any path itself is also a special case of a dendrite.

The length of a dendrite is defined as the sum of the weights of its edges. Inspired by previous observations, we assume that a dendrite of minimum length (*an optimal dendrite*) reproduces the order of n vertices of the graph in the best way (Fig. 5).

Wrocław taxonomy. In the example with weights, it was easy to identify the optimal dendrite. As the number of vertices in the graph increases, the question arises about computing it in an algorithmic way. In computer science, this problem is classical and known as the *minimum spanning tree problem*. According to [3], the earliest published solution to this problem (1926) comes from the Czech mathematician Otakar Borůvka, who was dealing with it in the context of developing an optimal electrical network in Moravia. The classical algorithms, known to participants of Olympiads in informatics, are Kruskal’s and Prim’s algorithms that were published in 1956 and 1957, respectively. It seems that mathematicians from Wrocław were the first to address this problem in the context of methods of ordering and classifying (i.e. *taxonomy*) in anthropology, biology, or linguistics. Their method, published in 1951 ([4]), is basically Borůvka’s algorithm (of which they were unaware) and is called the *Wrocław taxonomy* in Polish statistical literature.

As part of this work, the first step towards the general construction of an optimal dendrite was taken in 1949 by Kazimierz Florek. He noted that any optimal dendrite should contain segments connecting the nearest objects – that is, those connecting an object with its nearest neighbor. Such segments are called *connections of order I*.

In the example with the vases, this procedure solves the problem: connections of order I form a dendrite. However, this is not always the case. Connections of order I for 5 capitals form a disconnected subgraph, shown in the margin. Notice, however, that in the graph of connections of order I, cycles cannot occur – such a graph is called a *forest*, and it naturally breaks down into dendrites.



In 1950, the creators of taxonomy proposed an optimal dendrite construction, which is an iterative version of Florek's idea:

Method W

1. Build connections of order I for a given graph G . If they form a dendrite, the construction is finished. If not, proceed to step 2.
2. Create a new graph G_1 . The vertices of graph G_1 are sub-dendrites formed from the connections of order I of graph G . The discrepancy between sub-dendrites A and B is the discrepancy of the nearest neighbors: $d(A, B) := \min\{d(P, Q) : P \in A, Q \in B\}$. Create connections of order I for graph G_1 (these are connections of order II for G). If they form a dendrite, the construction is finished. If not, repeat step 2 for graph G_1 .
3. In this way, we create consecutive graphs G_1, G_2, G_3, \dots and for each of them, connections of the next order. These iterations must end because the number of vertices in subsequent graphs G_i decreases.
4. The construction of the final dendrite ends by connecting dendrites of successive orders with edges between objects that realize the nearest neighbor discrepancy.

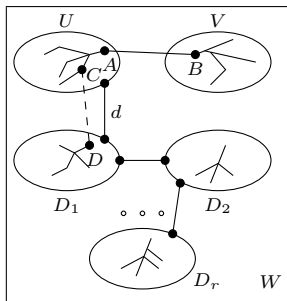


Fig. 7

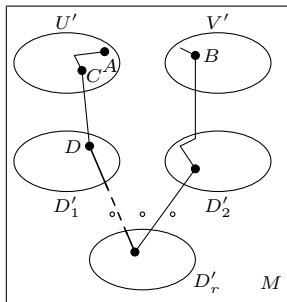


Fig. 8

Let us see how method W works for the example of 5 capitals. Connections of order I form sub-dendrites $O_1 = \{A, P, R\}$ and $O_2 = \{B, W\}$. Graph G_1 has vertices O_1 and O_2 . The minimum distance from B to the set of points $\{A, P, R\}$ is 650, and the minimum distance from W to $\{A, P, R\}$ is 1140 km. Therefore, the discrepancy of the nearest neighbor between O_1 and O_2 is equal to 650 km, which is the distance between Berlin and Amsterdam. G_1 with the connection of order II between these capitals is already a dendrite. This ends the construction, creating the final optimal dendrite, which turned out to be the path $WBAPR$.

Proof of optimality of Method W. Without loss of generality, we can assume that all nonzero discrepancies in the graph G are different. If necessary, we can replace the discrepancies d_{ij} with $d'_{ij} = d_{ij} + \varepsilon_{ij}$ (for $i < j$), where $d'_{ji} = d'_{ij}$ and $\varepsilon_{ij} > 0$ are any chosen distinct numbers smaller than $\min d_{ij}$. After the construction is finished, we need to return to the discrepancies d_{ij} .

Let W be the dendrite constructed using method W. Suppose that W is not an optimal dendrite, so it is different from some optimal dendrite M . Therefore, there is an edge in W that does not appear in M – let it connect vertices A and B .

Assume that the edge AB has weight a and order k in the dendrite W . It connects dendrites U and V of order $k - 1$ with vertex sets U' and V' , respectively. The other dendrites of rank $k - 1$, which we denote by D_1, D_2, \dots, D_r , have vertex sets D'_1, D'_2, \dots, D'_r . According to the construction of W , one of the following two cases holds:

1. a is smaller than each discrepancy between U and D_1, \dots, D_r ,
2. a is smaller than each discrepancy between V and D_1, \dots, D_r .

Without loss of generality, assume that case 1 holds.

In dendrite M , vertices A and B are connected by a path that does not contain the edge AB , and it must contain an edge CD such that C belongs to U' and D does not. Assume that D belongs to D'_1 . The weight of the edge CD is not smaller than the distance d between dendrites U and D_1 in dendrite W . Moreover, $a < d$ because case 1 holds. Therefore, we can create a new dendrite M' by replacing the edge CD in M with the edge AB . The length of M' is smaller than the length of M . This contradicts the assumption that M is an optimal dendrite.

Grouping. A set of objects might be non-homogenous: shells found in an ancient landfill or skulls found in a surveyed area may come from several distinct periods. How to divide the data so that dendrites, corresponding to the division, indicate significant differences in these groups?

A family of k dendrites D_1, D_2, \dots, D_k with sets of vertices Z_1, Z_2, \dots, Z_k is a *partition* of a given complete graph G with vertices Z when

$$Z = Z_1 \cup Z_2 \cup \dots \cup Z_k \quad \text{and} \quad Z_i \cap Z_j = \emptyset \text{ for } i \neq j$$

The *length of the partition* is the sum of the lengths $l(D_i)$ of the component dendrites. The partition is *optimal* if the length of the partition is minimal. From this definition, it immediately follows that the components of an optimal partition must be optimal, so we can assume $D_i = W(Z_i)$.

Let's go back to the example of the four vases. They can be divided into two groups in seven ways:

Z_1	$\{A, B\}$	$\{A, C\}$	$\{A, D\}$	$\{A, B, C\}$	$\{A, B, D\}$	$\{A, C, D\}$	$\{B, C, D\}$
Z_2	$\{C, D\}$	$\{B, D\}$	$\{B, C\}$	$\{D\}$	$\{C\}$	$\{B\}$	$\{A\}$
$W(Z_1)$	$A-B$	$A-C$	$A-D$	$B-A-C$	$B-A-D$	$C-A-D$	$B-C-D$
$W(Z_2)$	$C-D$	$B-D$	$B-C$	D	C	B	A
$l(W(Z_1)) + l(W(Z_2))$	23	27	23	11	22	17	24

The optimal partition $D_1 = C-A-B$, $D_2 = D$ is a subgraph of the dendrite obtained by the W method (Fig. 5). It turns out that this is always the case.

Theorem. If $\{W(Z_1), W(Z_2), \dots, W(Z_k)\}$ is an optimal partition of the graph G with vertices Z , then

$$W(Z_1) \cup W(Z_2) \cup \dots \cup W(Z_k) \subset W(Z).$$

Proof. As before, without loss of generality, we assume that all non-zero discrepancies are distinct.

Suppose that there exists an edge AB in $W(Z_1) \cup W(Z_2) \cup \dots \cup W(Z_k)$ that does not belong to $W(Z)$. There exists a path s_{AB} in $W(Z)$ connecting these vertices.

From the disjointness of Z_i , it follows that there exists an r such that the edge AB belongs to $W(Z_r)$. The elements of Z_r can be divided into subsets U and V as follows: U consists of the vertex A and all vertices in Z_r connected to A by a path in the dendrite $W(Z_r)$ that does not contain the edge AB ; V consists of the vertex B and all vertices in Z_r connected to B by a path in the dendrite $W(Z_r)$ that does not contain the edge AB .

The path s_{AB} must contain an edge CD such that $C \in U$ and $D \notin U$. Let $d(A, B) = x$ and $d(C, D) = y$. The inequality $y < x$ holds; otherwise, replacing CD with AB in the dendrite $W(Z)$ would decrease its length, which would be a contradiction.

There are two cases: $D \in V$ and $D \in Z_s$ for $s \neq r$.

1. $D \in V$ (Fig. 9). In this case, replacing AB with CD in $W(Z_r)$ yields a dendrite that is not optimal, which is a contradiction.

2. $D \in Z_s$ for $s \neq r$ (Fig. 10). We replace the dendrite $W(Z_r)$ with the dendrite $W(U) \cup CD \cup W(Z_s)$ of length $l(W(U)) + y + l(W(Z_s))$, and the dendrite $W(Z_s)$ with $W(V)$ of length $l(W(V))$. The sum of the lengths of all k dendrites after the change is smaller than the sum of their lengths before the change, which contradicts the assumption that the partition is optimal.

From the above theorem, we obtain a useful result in the context of finding an optimal grouping:

Corollary. The optimal partition of a graph G into k subdendrites involves removing from the dendrite $W(Z)$ the $k - 1$ edges with the largest discrepancies.

A bit about applications. The *Seminar on Applied Mathematics* actively promoted the idea of taxonomy, applying it in various fields. In Steinhaus's *Mathematical Snapshots*, one can read about the taxonomy of liverworts (*Hepaticae*) in the Beskidy Mountains. The dendrite corresponding to the

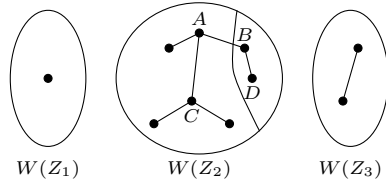
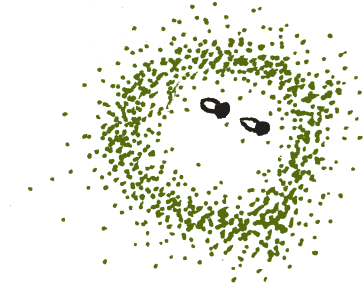
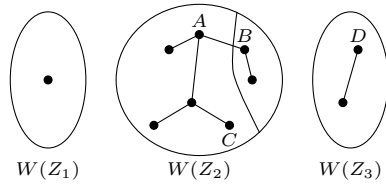
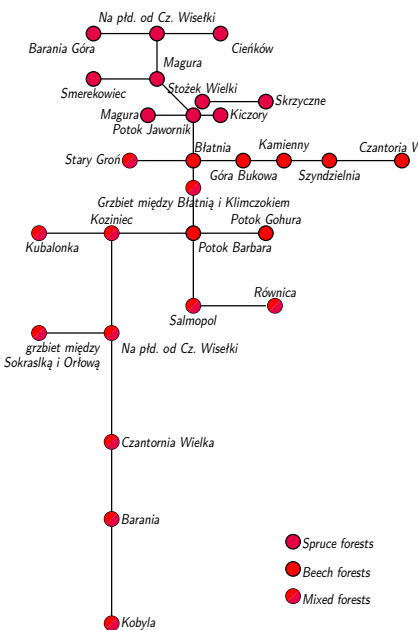


Fig. 9



Rys. 10



Rys. 11

frequency of occurrence of different liverwort species was found to be related to the type of forest – which was an interesting discovery (Fig. 11).

A characteristic feature of the work of the *Seminar* was tackling of every problem, even the most unusual ones. At a meeting in January 1952, *Julian Perkal announced that he had made a dendrite of folk songs for Professor Czekanowski's daughter* (quote from the carefully kept minutes of the *Seminar*). The classification of folk songs using the method of Wrocław taxonomy became one of the important research tools for Anna Czekanowska-Kuklińska, a professor at the University of Warsaw (d. 2021) and the head of the Ethnomusicology Department she established.

In 1953, Stefan Zubrzycki published a work [5] using the Wrocław taxonomy, which answered astronomer Włodzimierz Zonn's question of whether stars form non-random constellations (referred to by the authors as "chains") or are randomly distributed on the celestial sphere. He showed that they are randomly arranged, confirming that constellations are only a mnemonic method of remembering the position of stars.

Julian Perkal ends his work on taxonomy (op. cit.) with a warning that *"...one can construct a machine for making dendrites. This creates a danger of a mechanical approach to natural objects and of gyrating false sometimes natural bills with mathematical methods."* It is worth remembering this.

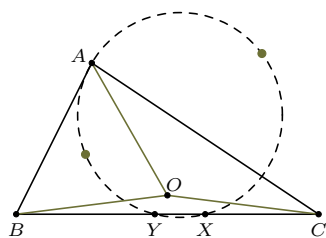


Bibliography

- [1] Szczotka, W., *Fenomen Steinhausowskiego Seminarium Matematyki Stosowanej*, *Antiquitates Mathematicae*, (2018) vol. 12(1), 197-228, (2019) vol. 13(1), 177-233
- [2] Perkal, J. (1953). *Taksonomia wrocławska*. *Przegląd Antropologiczny wrocławski*. *Przegląd Antropologiczny wrocławski*, vol. 2, nr 3-4 str. 282-285
- [3] Graham, R. L., Hell, P. (1985). *On the history of the minimum spanning tree problem*. *Annals of the History of Computing*, 7(1), 43-57.
- [4] Florek K., Łukaszewicz J., Perkal J., Steinhaus H., Zubrzycki S., *Sur la liaison et la division des points d'un ensemble fini*, *Colloquium Mathematicum* (1951), vol.2, nr 3-4 str. 282-285
- [5] Zubrzycki, F. (1953). *O łańcuskach gwiazdnych*. *Applications Mathematicae*, 1(3), 197-205.



Problems



Edited by Dominik BUREK

M 1750. Can the numbers from 1 to 2023^2 be placed in the squares of a 2023×2023 board in such a way that for any choice of a row and a column, we can find three numbers on them, where one of the numbers is the product of the other two?

Solution on page 2

M 1751. Let O be the circumcenter of triangle ABC . Points X and Y on side BC are such that $AX = BX$ and $AY = CY$. Prove that the circumcircle of triangle AXY passes through the circumcenters of triangles AOB and AOC .

Solution on page 8

M 1752. Let $x_1, \dots, x_n \in [0, 1]$. Prove that

$$(1 - x_1x_2 + x_1^2) \cdot (1 - x_2x_3 + x_2^2) \cdot \dots \cdot (1 - x_{n-1}x_n + x_{n-1}^2) \cdot (1 - x_nx_1 + x_n^2) \geq 1.$$

Solution on page 4

Edited by Andrzej MAJHOFER

F 1075. An eclipsing binary star system with radii r_1 and r_2 is observed from Earth at an angle α to the plane of the stars' mutual orbit. What is the relation between the angle α , radii r_1 and r_2 , and the diameter d of the orbit? We assume that the orbit is circular.

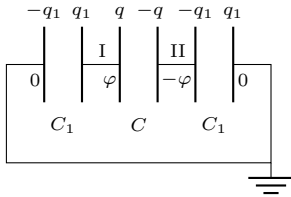
Solution on page 7

F 1076. On one of the plates of a flat capacitor with capacitance C , a charge Q_1 is placed, and on the other plate, a charge Q_2 is placed. What is the potential difference between the plates?

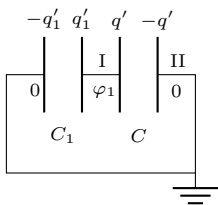
Hint: As usual in problems of this type, we neglect boundary effects.

Solution on page 7

Club 44 F



Rys. 1



Rys. 2

The rules of Club 44 F/M can be found on the webpage deltami.edu.pl (in Polish)

Edited by Elżbieta ZAWISTOWSKA

Solutions to problems from 3/2023

Let us recall problem statements:

754. The plates of a parallel plate capacitor with capacitance C are charged to potentials φ and $(-\varphi)$ relative to the ground. Each of the plates forms a capacitor with the ground, with capacitance C_1 . Find the ratio of the electric field intensities between the plates of the capacitor with capacitance C initially and after grounding one of the plates.

755. A closed container is completely filled with water. Just above the bottom of the container, there is an air bubble. How will the pressure at the bottom change when the bubble rises to the surface?

754. The equivalent system to the one described in the problem is shown in Figure 1. A capacitor with capacitance C is charged with a charge of $q = 2\varphi C$, and the charges on the plates of capacitors with capacitances C_1 are $q_1 = \varphi C_1$. The total charge on the left plate is given by

$$(1) \quad Q = q + q_1 = (2C + C_1)\varphi.$$

After grounding the right plate, the equivalent system is shown in Figure 2. The charge on the plates of the capacitor C is $q' = \varphi_1 C$, where φ_1 represents the potential of the ungrounded plate. The charge on the capacitor with capacitance C_1 is $q'_1 = \varphi_1 C_1$, and the total charge on the left plate is

$$(2) \quad Q' = q' + q'_1 = \varphi_1(C + C_1) = Q.$$

Taking into account equation (1), we obtain

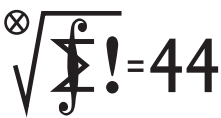
$$\varphi_1 = \frac{\varphi(2C + C_1)}{C + C_1}.$$

The electric field intensity between the capacitor plates is the ratio of the voltage to the distance between them. Therefore, the desired ratio of these intensities is

$$\frac{E}{E_1} = \frac{2\varphi}{\varphi_1} = \frac{2(C + C_1)}{2C + C_1}.$$

755. In the initial state, the air pressure inside the bubble is the same as the pressure of the water at the bottom of the container. The pressure difference between the bottom of the container and the top is given by $\Delta p = \rho gh$, where ρ is the density of water and h is the height of the container. During the ascent of the bubble, its volume remains unchanged because the liquid is practically incompressible. Therefore, the air pressure inside the bubble also remains constant. When the bubble reaches the top of the container, the pressure of the water under the surface is equal to the initial pressure at the bottom, so the pressure at the bottom has increased by Δp .

Club 44 M



Edited by Marcin E. KUCZMA

Solutions to problems from 3/2023

Let us recall problem statements:

857. Find all pairs of positive integers x, y such that $x^2 - 4y$ and $y^2 - 4x$ are squares of integers.

858. An equilateral triangle ABC of sidelength 1 and a segment DE of length 1 lie in the 3-space so that the segment has a point in common with the triangle. Show that one of the points A, B, C, D, E lies at a distance not exceeding 1 from each one of the other four points.

Problem 858 proposed by Michal Adamaszak of Copenhagen.

857. Investigated are solutions of the system of equations

$$(1) \quad x^2 = a^2 + 4y, \quad y^2 = b^2 + 4x$$

in integers $x, y \geq 1, a, b \geq 0$. Obviously, x and a must be of equal parity; (and likewise y and b).

By symmetry it will suffice to consider $x \geq y$. Then $x^2 = a^2 + 4y \leq a^2 + 4x$, i.e.

$$(2) \quad x^2 - 4x - a^2 \leq 0.$$

The first equation from (1) shows that $x^2 \geq a^2 + 4$, which (combined with the quadratic inequality (2)) yields the two-sided estimate

$$(3) \quad \sqrt{a^2 + 4} \leq x \leq 2 + \sqrt{a^2 + 4}.$$

It is easily seen that the number $a + 2$ lies in the interval $[\sqrt{a^2 + 4}, 2 + \sqrt{a^2 + 4}]$ (of length 2). If $a \geq 1$, this is the

only integer of the same parity as a (in this interval); so $x = a + 2$. The system (1) now forces $y = a + 1$ and $(a + 1)^2 = b^2 + 4(a + 2)$. Rewrite the last equation as $(a - 1)^2 = b^2 + 8$; i.e.,

$$(a - 1 - b)(a - 1 + b) = 8,$$

with the unique solution (in integers) $a = 4, b = 1$. Hence

$x = 6, y = 5$. The quadruple $(a, b, x, y) = (4, 1, 6, 5)$ satisfies the original system (1).

If, however, $a = 0$, the two-sided inequality (3) is fulfilled by two even integers $x = 2$ and $x = 4$. Plugging these into (1), the first of these two values yields contradiction, while the second one results in $(a, b, x, y) = (0, 0, 4, 4)$, which is a solution.

Taking symmetry into account (hence dismissing $x \geq y$), we obtain the following pairs (x, y) : $(4, 4)$, $(6, 5)$ and $(5, 6)$.

858. This matrix can reveal the combinatorial nature of the problem:

$$\begin{bmatrix} i(A, D) & i(B, D) & i(C, D) \\ i(A, E) & i(B, E) & i(C, E) \end{bmatrix}$$

where

$$i(X, Y) = \begin{cases} 0 & \text{when } XY \leq 1 \\ 1 & \text{when } XY > 1 \end{cases}$$

(XY is the length of the segment with endpoints X, Y).

Since $AB = AC = BC = DE = 1$, the problem comes down to showing that at least one row or one column of this matrix has only zero entries. Suppose this is not the case. Then the following pattern appears in the matrix (up to a possible permutation of the symbols A, B, C and/or D, E , which does not influence the conditions of the problem): $\begin{bmatrix} 1 & 1 & 1 \\ & & 1 \end{bmatrix}$; in terms of geometry (with notation as above):

$$AD > 1, \quad BD > 1, \quad CE > 1;$$

i.e.

$$AD > AC, \quad BD > BC, \quad CE > DE.$$

Let π be the perpendicular bisector plane of the segment CD . These inequalities imply that the points A and B lie on the same side of π as C , while E lies on the same side of π as D does (and no one of those points lies on π). Therefore π separates the triangle ABC from the segment DE , contradicting the condition that they meet. The result follows.

Are these distributed uniformly? ... that is on the V/V_{\max} method

Radosław POLESKI*

*Astronomical Observatory, University of Warsaw (rpoleski@astrouw.edu.pl)

In the XXI century, astronomers are obtaining *huge* amounts of observations with various level of specificity. Extracting knowledge from this *cosmos* of data obviously requires statistical analysis. Such analysis can be of different kinds: from extracting brightness and positions of objects from an image, through joint analysis of multiple observations in order to find a period of some phenomenon (e.g., eclipses in a stellar binary system), to analysis of data from various sources in order to determine parameters of a specific object (e.g., what is the distance to the Galactic center or what is the fraction of mass in the Universe that consists of baryons).

Astronomy is significantly different from physics with respect to how the data to be analyzed are obtained: astronomy is based on observations of phenomena that we have no control of, whereas physics is mostly based on experiments performed (and hence controlled) by the scientist. This nature of astronomical observations poses a severe difficulty – sometimes increasing the sample of objects under study is extremely expensive and may require time that is longer than the expected lifetime of the researcher. Hence, astronomers often face incomplete samples of objects, even though they very much would like it to be otherwise. There are other obstacles, e.g., information about objects studied may come from observations taken under different conditions, epochs, etc. These subtle differences have to be taken into account during the statistical analysis, which is not an easy task.

In this article, I would like to present a statistical method that is typical for astronomy and bears an

exotic name “ V/V_{\max} .” It was designed in the late 60s in order to tackle the following issue: we have a catalog of quasars with known brightness and redshift and we would like to know if quasars are distributed uniformly in space.

Quasar (from “quasi-stellar object”) is a type of active galaxy that emits extremely bright radiation.

Note that for the most luminous quasars, the sample can be considered complete for very large distances, whereas for the less luminous quasars the sample is complete for smaller distances. At the first “glance” the quasar space density may seem to be getting lower with increasing distance from Earth due to different luminosities. However, this may be an artifact of higher overall completeness for smaller distances (compare Fig. 1 and Fig. 2).

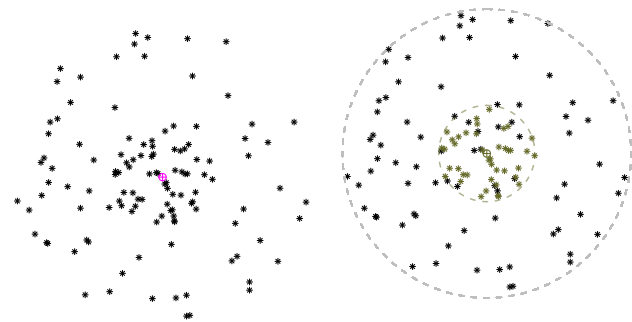


Fig. 1. Illustration of 150 quasars observed by a 2D astronomer in a 2D Universe. These quasars seem to show concentration around Earth (marked by the \oplus symbol).

Fig. 2. Illustration of the same quasars as in Fig. 1 divided into two groups with different absolute magnitudes. Quasars from each group are uniformly placed in the area in which they can be seen.

Let us start with basics and assume that all quasars have the same absolute brightness (i.e., they are equally bright when observed at a distance of 10 parsecs). If all quasars are uniformly distributed in space, then the *observed* quasars are uniformly distributed in the part of space, a ball, in which we can see them (see black points in Fig. 2). The volume of this ball is labeled V_{\max} . For each quasar K we determine the

distance and calculate the volume of the ball V_K with radius equal to that distance. For quasars uniformly distributed in space, the distribution of V/V_{\max} values should be uniform in $[0, 1]$ interval (Fig. 3), hence, the mean of V/V_{\max} should be approximately equal to $1/2$. If the distribution of V/V_{\max} does not satisfy these conditions, then we can reject the hypothesis of uniform distribution of quasars in space.

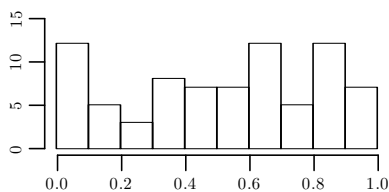


Fig. 3. Histogram of V/V_{\max} for “black quasars” from Fig. 2 assuming fixed value of V/V_{\max} . The height of each bar corresponds to the number of quasars with V/V_{\max} value in a range defined by the base of the bar.

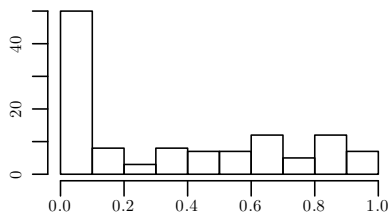


Fig. 4. Histogram of V/V_{\max} for all quasars from Fig. 2 assuming fixed value of V_{\max} .

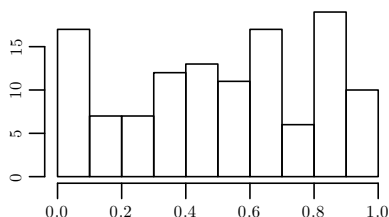


Fig. 5. Histogram of V/V_{\max} for all quasars from Fig. 2 with V_{\max} estimated for each quasar separately.

What about quasars with different absolute brightness (i.e. as with real quasars)? First, let us assume (for simplicity) that for each quasar, the absolute brightness has one of the two values (like in Fig. 1 and 2). As already noted, it will affect the sample of observed quasars because the more luminous quasars will be seen from larger distances than the less luminous ones. If we use the same V_{\max} value for each quasar, then we will get a histogram that is similar to Fig. 4, which is significantly different from the one in Fig. 3. We have to modify our calculations and determine the V_{\max} value separately for each quasar, based on the absolute brightness (labeled $V_{\max,K}$ for quasar K). Nonetheless, if all quasars are distributed uniformly in space, then for each quasar the value of $V_K/V_{\max,K}$ can be treated as randomly drawn from $[0, 1]$ interval, independently from other quasars. In other words, the distribution of $V_K/V_{\max,K}$ should still be uniform in $[0, 1]$ interval (see Fig. 4 and 5) and its mean should be close to $1/2$. These conditions are fulfilled both by all quasars in Fig. 2, as well as only the black ones and only the colored ones.

Now we are ready for a generalization – each quasar has different absolute brightness. For each quasar we can calculate corresponding V_K and $V_{\max,K}$. Once more we expect that for a uniform quasar distribution the V/V_{\max} distribution will be uniform in $[0, 1]$ range and with mean close to $1/2$. If we have observational data, then it is enough to calculate V/V_{\max} , check if the distribution is consistent with expectations and we know whether or not quasars are uniformly distributed!

The V/V_{\max} method has been further developed in various directions. For Curious Readers (that already have some familiarity with these matters) I would like to briefly present an extension that is most important in my opinion. It is possible to analyze how parameters of populations change with time. Let us imagine the Universe at the age corresponding to a redshift z . Let $\rho(z)$ be the ratio of volume density of studied objects then and now (i.e., $z = 0$). Instead of considering volume V we consider generalized volume:

$$V'(z) = \int_0^z \rho(z') dV(z'),$$

where $V(z)$ is volume in co-moving coordinates and is calculated based on the assumed cosmological model. For each quasar we know its redshift z_K , hence, we can calculate the corresponding $V'_K(z_K)$ and $V'_{\max,K}(z_K)$. If we assume that quasars are distributed uniformly and we assume $\rho(z)$ function that is close to the true one, then the distribution of V'/V'_{\max} should be... (I do not think that I need to repeat myself). How we can know what $\rho(z)$ we should assume? The easiest approach is to check many different possibilities, for each of them calculate V'/V'_{\max} , and at the end select those that give the results expected.

It is worth noting that the V/V_{\max} method and its variants are still used in scientific papers. For example, a few years ago this method was used for a completely different problem: studying exoplanet frequency $f(q)$ as a function of star-planet mass ratio q for very small values of q , i.e., for cases for which our knowledge is rather poor ([1], [2]). The problem was how to include planets found in data collected in heterogeneous way. Instead of $\rho(z)$, the searched function was $f(q)$ and instead of $V(z)$ researchers used probability of finding in given system a planet with a different mass ratio. Detailed description of these studies and their results is a topic for a separate article...

- [1] Udalski et al. *OGLE-2017-BLG-1434Lb: Eighth $q < 1 \times 10^{-4}$ Mass-Ratio Microlens Planet Confirms Turnover in Planet Mass-Ratio Function*. Acta Astronomica 68.1 (2018): 1-42.
- [2] Jung et al. *KMT-2017-BLG-0165Lb: A Super-Neptune-mass Planet Orbiting a Sun-like Host Star*. The Astronomical Journal 157.2 (2019): 72.

Straight from Heaven: Wooden clock

Archaeologists use the fact that the radioactive element is present in studied samples to determine their age. Once produced, carbon-14 turns back into nitrogen, with a half-decay time of about 5750 years. Living organisms maintain roughly constant levels of radioactive elements due to continuous uptake and excretion of matter, while a dead organism contains a certain initial amount of carbon-14, which becomes less and less with time.

amount of cosmic radiation in the past dating back as far as thousands of years. The carbon-14 content also depends on other factors, such as the magnetic field of the Earth and the Sun, which acts as a shield for the Earth's surface against cosmic radiation coming from outside the Solar system (more particles reach Earth when these magnetic fields are weaker and fewer when they are stronger). Changes in carbon-14 levels in growth rings store a history of changes in Earth's magnetization as well as the 11-year cycle of the solar dynamo, which is related to the solar magnetic field.

The wood, however, contains the data that we cannot explain. In 2012, a Japanese physicist Fusa Miyake discovered a significant jump in the carbon-14 content of tree rings dating back to the year 774. The difference was so large that it must have been caused by cosmic radiation many times larger than average. Subsequent "Miyake events" were in the years 993 and 663 BCE, as well as even earlier events of 5259, 5410 and 7176 BCE. The well-located events in the wood (and in time) make it possible to precisely date specific events to the exact year, e.g., the event of 993 allowed to pinpoint the timing of the the establishment of the first European settlement in America, a Viking village in New Fundland, to the year 1021.

How does such massive and short-lived radiation occur? Among the "suspects" are nearby supernovae, gamma-ray bursts, emission from magnetized neutron

"Modelling cosmic radiation events in the tree-ring radiocarbon record", Qingyuan Zhang i inni, Proc. R. Soc. A. 478 2022.0497, 2022.

The Night Sky in July

Throughout the month, the Sun will lower its altitude by over 4.5° , reducing its time above the horizon in central Poland by over an hour. On July 24th, the Sun will cross $+20^\circ$ declination and thus the period of the longest days and shortest nights will end. As every year at the beginning of July, the Earth is at the aphelium of its orbit, which means that the Sun has the smallest visible disc in the year. Therefore, it is easier for it to be covered by the Moon during a potential eclipse – and such eclipses also can last longer.

In July, the solar system's brightest planets are not well visible on the sky. *Mercury* will begin the month with an upper conjunction with the Sun and will move toward maximum eastern elongation, which it will

When high-energy cosmic radiation collides with the upper layers of the Earth's atmosphere, some collisions lead to the formation of neutrons, leading to the neutron-proton $n-p$ reaction, which is the conversion of nitrogen nuclei ${}^7_{14}\text{N}$ into radioactive carbon-14, ${}^6_{14}\text{C}$: ${}^7_{14}\text{N} + n \rightarrow {}^6_{14}\text{C} + p$. Carbon-14 falls to the Earth's surface, taking part in the normal biochemistry of living organisms, including becoming bound during tree growth.

By examining the wood, and in particular the differences in carbon-14 content in individual annual growth rings (wood grain), it is possible to determine the initial amount of radioactive material and therefore study the evolution of the

stars, and even comets. Currently, the best explanation is that Miyake events are associated with solar superstorms. These (hypothetical) eruptions from the Sun are 50-100 times more energetic than the largest recorded in the modern era: the solar storm observed by Richard C. Carrington and Richard Hodgson in 1859.

In a paper by Q. Zhang and colleagues, the authors analyse available material from the tree rings finding evidence that events can occur at any moment of the Sun's 11-year activity cycle (which we wrote about in e.g. Δ_{21}^1). On the other hand, solar flares tend to occur near the maximum of the cycle. Several of the recorded spikes in of radioactivity appear to last longer than would be suggested by a model of a single solar superstorm. This suggests that these events can sometimes last longer than a year, which is not expected for a single giant solar flare, i.e., that we would be dealing with a long-lasting solar superstorm solar weather.

Such an event happening today would destroy power grids, telecommunications and most satellites. If they occur randomly, for example, once every thousand years, there is about a 1% chance per decade, which is a non-negligible probability.

Michał BEJGER

Nicolaus Copernicus Astronomical Center of the Polish Academy of Sciences, Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Ferrara, Włochy

reach in the first ten days of August, moving away then by 27° from the Sun. Unfortunately, at this time of year the ecliptic is tilted unfavourably to the horizon, making the planet set less than an hour after the Sun, and it is invisible from high latitudes. Notably, with each successive day the planet's brightness lowers, from -0.4^m on July 19th to $+0.1^m$ on July 31st. A pity, because on July 28th the planet will pass less than $20'$ from Regulus, the brightest star of Leo constellation.

Observational conditions in July for *Venus* are even worse. The second planet from the Sun, after June's maximum elongation, is rapidly moving toward August's lower conjunction with the Sun. This means that the planet is rapidly approaching us, while

increasing the size of the disc and decreasing the phase. At the beginning of July, Venus will present a disc of brightness -4.4^m , diameter $34''$ and phase of 31%. On the last day of the month, the corresponding magnitude, diameter and phase will be: -4.2^m , $53''$ and 5%. Unfortunately, the planet will dive several degrees below the weakly inclined ecliptic in the process, disappearing into the evening aurora at the end of the month. The disc of Venus in July is therefore an attractive observational target for owners of binoculars and telescopes, however not for those residing far north of the equator.

In early July after dusk, you can also try to spot *Mars*. The Red Planet originally travels less than 4° from Venus, but then Mars will travel further southeast and Venus will turn back toward the Sun, so by July 20th the distance on the sky between the planets will more than double. On July 10th, Mars will pass the aforementioned Regulus at a distance of just over 0.5° . Mars is the most difficult to see through the evening aurora, as it is already far from Earth and shines with a brightness of $+1.7^m$, so not much greater than the nearby brightest star Leo.

All of these planets in the last ten days of the month will be visited by the Moon in the waxing crescent phase. The new Moon falls on July 17th and in the following days it will move to the evening sky, but will also suffer from the low ecliptic. This situation will be somewhat improved by the fact that it will spend almost all of its time, until the first quarter, which falls on July 25th, north of the ecliptic. As early as July 18th, it is possible to try to spot the Moon's very thin sickle at a dusk, but this is a difficult task, requiring a very clear atmosphere and a low exposed skyline. Thirty minutes after sunset, the Moon's disc in phase of just 1% will take up a position at a height of 3° . At the same time 7° to the left of the Moon will be the planet Mercury, while another 15° further away – the planet Venus. In turn, 3.5° above Venus will show Regulus, 5° from Regulus, at 10 o'clock relative to it will be the planet Mars. The Moon, Mercury and Venus will set very quickly, Regulus and Mars a little later. To find all these celestial bodies, however, binoculars may be necessary. On July 19th, the Moon in phase will pass 4° over Mercury, while 24 hours later, with its phase increased to 8%, the Silver Globe will pass 3° over Regulus and at the same time 7° over Venus. At 6° to the left of the Moon will show Mars. On July 21st, the phase of the lunar disk will grow to 14%, and Mars should then be sought at a distance of 6° at 4 o'clock relative to the Moon.

After passing the planets, on July 24th and 25th, the Earth's natural satellite, in the first quarter, will meet Spika, the brightest star of Virgo. It is still worth mentioning the very close encounter between the Moon and Antares, on July 28. At the time of the Sun's setting, the Moon's disc in phase of 79% will show 0.5° from Scorpio's brightest star. Until the end of the month, the Moon will remain south of the ecliptic, wandering low over the horizon. The closer we get to

the end of July, however, the fuller the lunar disk will become, as the Moon will pass through a full Moon on the evening of August 1st our time.

The beginning of July will also be influenced by the light of the Silver Globe's disc. July's full moon will fall on July 3rd in the constellation of Sagittarius. Before it, on the first two nights of the month, the Moon will visit Scorpio, shining first 6° to the right and then 8° to the left of Antares. On 7th July, presenting a disc illuminated in 83%, Earth's natural satellite will approach at 5° from Saturn. The ringed planet in August will move southwesterly less than 1° from the 5th magnitude double star *sigma* Aqarii. The planet's disk will exceed a diameter of $18''$, shining with a brightness of $+0.7^m$. Saturn will peak at dawn, rising to an altitude of over 25° .

The Silver Globe will meet with *Neptune* on July 9th, approaching it at a distance of 5° . The planet will pass through opposition to the Sun in September and also move in retrograde motion. This year Neptune will trace a loop on the border of the Pisces and Aquarius constellations, not far from the distinctive, miniature Ursa Minor-like system of 5th and 6th magnitude stars, which is formed by the stars 30, 33, 27, 29, 24 and 20 Piscium. In July, Neptune is located about 1° north of 24 Psc, shines with a brightness of $+7.8^m$ and around 2 o'clock rises to a height of more than 20° above the southeastern horizon.

The Moon will pass through the last quarter on July 10th, and then proceed towards the faintly visible planets *Jupiter* and *Uranus*. On the morning of July 12th, the phase of the lunar disk will drop below 30% and it will ascend just after midnight 2.5° from Jupiter and at the same time 9° west of Uranus. Both planets in the coming observing season will circle against the background of the constellation Aries at a distance of a few degrees from each other and in November both will pass through opposition to the Sun. For the time being, they are roughly 10° apart and at around 2 o'clock they rise 15° above the eastern part of the sky. Jupiter shines with brightness of -2.3^m , presenting a disc of diameter $38''$, so there is no trouble with spotting it. The situation is different for Uranus, whose brightness is $+5.8^m$ and because of its low altitude in the dark sky, its image depends on the quality of the atmosphere.

The new Moon falls on July 17th in the evening, our time, and thanks to the fact that its orbit is now almost maximally tilted north of the ecliptic, its thin crescent along with the so-called ash light will remain visible for the next 4 days. On the 13th day of the month, the Moon's crescent in phase of 20% will approach the Pleiades at 3° , 24 hours later its phase will drop to 13% and it will pass 8° over Aldebaran, while on July 15th, in phase of 7%, it will pass at a distance of just over 2° from El Nath, two bright stars of Taurus. On July 16th at dawn, the Moon in phase at just 3% will show itself at a height of 7° over 20° under Capella.

Ariel MAJCHER



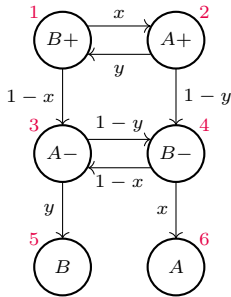
Markov Chains – part 1

Bartłomiej BZDEGA

Adam Mickiewicz University in Poznań

Let us consider the following problem:

Two players, A and B , are playing chess, with A starting. Each move can be either strong or weak. The first player to respond with a strong move to the opponent's weak move wins the game. Player A makes a strong move with probability x and a weak move with probability $1 - x$. Similarly, player B has probabilities y and $1 - y$. We assume that $0 < x < 1$ and $0 < y < 1$. For which values of x and y are the chances of winning for A and B equal?



Solution. Note that as long as nobody has won, the situation depends only on the last move. There are four possibilities: a strong/weak move by player A/B . Let us denote them by $A+$, $A-$, $B+$, and $B-$. The moment when player A starts can be considered separately, but there is no need for that because it is equivalent to the situation $B+$. So, we can take $B+$ as the starting point. Additionally, we include states A (indicating that A has won) and B (indicating that B has won). The diagram shows all six situations, with assigned numbers from 1 to 6, and connected by arrows representing transition probabilities.

We call such an object a finite *Markov chain*. It is described by a set of states S_1, S_2, \dots, S_n and transition probabilities $p_{i,j}$ for moving from state S_i to state S_j in one step for all $i, j \in 1, 2, \dots, n$. This means that when in state S_i , the chain will transition to state S_1 with probability $p_{i,1}$, to state S_2 with probability $p_{i,2}$, and so on. For each i , the equation $p_{i,1} + p_{i,2} + \dots + p_{i,n} = 1$ holds. States S_i from which there are no outgoing transitions ($p_{i,i} = 1$) are called *absorbing*.

We are interested in the probability of player A winning, which is the probability that the process ends in state S_6 , given that it started in state S_1 . To approach this problem, let q_i denote the probability that the process ends in state S_6 , given that it started in state S_i . Note

that $q_6 = 1$ and $q_5 = 0$. From state S_1 , we can transition to S_2 (with probability x) or S_3 (with probability $1 - x$). This implies that $q_1 = xq_2 + (1 - x)q_3$. Similarly, we obtain the equations:

$$q_2 = yq_1 + (1 - y)q_4, \quad q_3 = (1 - y)q_4, \quad q_4 = (1 - x)q_3 + x.$$

By solving this system of four equations with variables q_1, q_2, q_3, q_4 , we obtain:

$$\frac{1}{2} = q_1 = \frac{x(1-y)}{(1-xy)(x+y-xy)} \iff x-y = xy(1-x)(1-y),$$

which means that for x and y satisfying the last equality, both players have an equal chance of winning. Note that it must hold that $x > y$, since the right-hand side of the last equality is positive.

Problems

- The kitten wanders between the house, kindergarten, garden, field, and forest. It starts in the garden and always chooses one of the remaining four places. It always plays in the house and then goes to the field or back to the garden. After a walk in the forest, the kitten always goes either to the garden or to the kindergarten. If the kitten reaches the kindergarten, it never leaves from there (children, you know...). On the field, however, the kitten catches a mouse and ends its wanderings. Calculate the probability of ending up in the field. (We assume that the kitten's choices are random and equally likely.)
- A student wants to buy their favorite energy drink, which costs 5 zlotych. Unfortunately, the student has only 2 złote. So, they decide to go to a casino where, with a probability of p , they can win 3 złote for each zloty bet, or lose the bet with a probability of $1 - p$. The student stops playing when they have enough money to buy the drink or when they run out of money. Depending on the value of p , determine the chances of the student achieving their goal. (Note: The author of this column does not endorse energy drinks or gambling in any way.)
- Hansel and Gretel toss a fair coin. If the sequence HTT (where H denotes heads and T denotes tails) appears in three consecutive tosses, Hansel wins the game. If the sequence HHT appears, Gretel wins. Determine the probability of Hansel winning.
- Seven children are standing in a circle, playing with a ball. Each child who currently has the ball throws it to the child standing immediately to their left (with probability $p < \frac{1}{2}$) or to the child standing immediately to their right (also with probability p), or takes the ball and goes back home (with probability $1 - 2p$). Depending on the value of p , calculate the probability that the same child who brought the ball home will return with it.

Hints and solutions
 1. Let S_1, S_2, \dots, S_7 denote the cat's presence at home, kindergarten, garden, field, and forest respectively. Let q_i denote the probability that the cat finishes its journey in the field (state S_4), given that it started in state S_i . Of course, $q_4 = 1$ and $q_2 = 0$. The value of q_3 is determined by solving the system of equations: $q_1 = \frac{1}{2}q_3 + \frac{1}{2}$, $q_3 = \frac{1}{2}q_1 + \frac{1}{2} + \frac{1}{2}q_5$, $q_5 = \frac{1}{2}q_3$. State S_i for $i = 0, 1, 2, 3, 4$ can be identified with the amount of money the student has, and state S_5 is the one in which the student has 5 or 6 zlotys, i.e., enough to buy a drink.
 3. In this problem, we can identify the states as follows: start, H and T (the outcome of the first toss and those following tails), LHL, LHT, HHL, HHT (Gretel's win), and LHL, LHT, HHL, HHT (Hansel's win). Gretel can choose a different configuration that gives her an advantage in such a game (Rozwiązania O Rzeszce i OOH, Δ15).
 4. Let us number the children from 1 to 7 in the order they stand in the circle. Let S_i be the state in which the i -th child has the ball, and let S_{7+i} be the state in which the i -th child takes the ball home. Assuming that the first child came with the ball, the initial state is S_1 , and we are interested in q_6 . Calculations are significantly simplified by the equalities $q_2 = q_7, q_3 = q_6$, and $q_4 = q_5$, which follow from symmetry.

