Вычислительные технологии, 2023, том 28, № 1, с. 61–71. © ФИЦ ИВТ, 2023 ISSN 1560-7534
Computational Technologies, 2023, vol. 28, no. 1, pp. 61–71. © FRC ICT, 2023 eISSN 2313-691X

DOI:10.25743/ICT.2023.28.1.007

Symbolic computation programs and generalized Catalan numbers in problems of analysis of random point structures

A. L. Reznik[1], A. V. Tuzikov[2], A. A. Soloviev[1,*], A. V. Torgov[1]
[1]Institute of Automation and Electrometry of the Siberian Branch of the Russian Academy of Sciences, 630090, Novosibirsk, Russia
[2]United Institute of Informatics Problems, 220012, Minsk, Belarus
*Corresponding author: Soloviev Alexander A., e-mail: solowey@rambler.ru
Received April 30, 2022, accepted May 16, 2022.

New algorithms for calculating exact analytical dependences arising in the solution of various theoretical and applied problems related to the analysis of random point structures are proposed, substantiated and implemented. New scientific results presented in this work were obtained using original research methods based on computer analytics programs using generalized multidimensional Catalan numbers.

Keywords: random image, computer analysis, random local groupings, generalized Catalan numbers.

Citation: Reznik A.L., Tuzikov A.V., Soloviev A.A., Torgov A.V. Symbolic computation programs and generalized Catalan numbers in problems of analysis of random point structures. Computational Technologies. 2023; 28(1):61–71. DOI:10.25743/ICT.2023.28.1.007.

Introduction

A distinctive feature of this work is that the solution of the presented problems is based on a specially developed toolkit. This toolkit includes two main components: a) symbolic computer-analytical calculations based on multidimensional integration over convex polyhedra in n-dimensional space; b) a three-dimensional generalization of the classical Catalan numbers. This approach is dictated by the fact that the set of traditional tools usually used in the field of digital image processing is not enough to solve time-consuming tasks of analyzing random point fields.

There are many tasks that require such a non-standard approach. They arise in many branches of computer science, for example, in processing of aerospace images, when it is necessary to localize hidden objects for their more detailed study [1, 2], in computer processing of biomedical images at the stage of an operational search for dangerous disease-causing abnormalities [3–5], in mathematical theory of communication [6], in construction of multidetector search systems for pulse-point sources [7, 8]. Mathematically similar problems are encountered in information theory and technical diagnostics (in particular, when troubleshooting with an intermittent failure discipline) [9, 10], and when studying the process of registering random point fields using a scanning aperture with a limited number of threshold levels [11].

In [11], it was shown that many of the above problems of registration, processing and analysis of digital images and point fields directly or indirectly lead to the need to solve the following one-dimensional problem: "Find the probability that after random dropping of n points on the interval $(0, 1)$, no ε-grouping will be formed, which includes more than k points that can be completely covered by some subinterval $\Omega_\varepsilon \subset (0, L)$ of length ε". Hereafter, without loss of generality (this is an obvious consequence of the standard normalization), we will assume that $L = 1$, $0 < \varepsilon < 1$, and the notation $P_{n,k}(\varepsilon)$ will be used instead of the probability $P_{n,k}(\varepsilon, L)$.

The purpose of this paper is to present specialized analytical methods and software algorithms designed to find exact probability formulas using computer analytics and generalized multidimensional Catalan numbers.

1. Computer-based analytical calculation of specific solutions for the problem

The simplicity of the problem posed in the introduction is illusory, and its analytical solution is known only for the simplest case $k = 1$ [12, 13]:

$$P_{n,1}(\varepsilon) = (1 - (n-1)\varepsilon)^n, \qquad 0 \le \varepsilon \le 1/(n-1). \tag{1}$$

Formula (1) describes the probability of an event that if n points are randomly dropped onto the interval $(0, 1)$, not a single ε-group will be formed containing at least 2 points, i. e. that all dropped points will be at a distance exceeding ε. The classical way to obtain solution (1) is to represent the required probability in the form of an easily integrable iterated integral [11]:

$$P_{n,1}(\varepsilon) = n! \int\limits_{(n-1)\varepsilon}^{1} dx_n \left\{ \int\limits_{(n-2)\varepsilon}^{x_n-\varepsilon} dx_{n-1} \cdots \left\{ \int\limits_{2\varepsilon}^{x_4-\varepsilon} dx_3 \left\{ \int\limits_{\varepsilon}^{x_3-\varepsilon} dx_2 \left\{ \int\limits_{0}^{x_2-\varepsilon} dx_1 \right\}\right\}\right\}\right\}.$$

Solution (1) can be obtained in a variety of ways. For example, a simple probabilistic-geometric method was proposed that allows calculating the probability (1) without resorting to the procedure of multidimensional integration. Thus, it is not difficult to find a solution to the main problem when $k = 1$. But for $k > 1$, the problem becomes much more complicated.

To advance in solving this problem for $k > 1$, we have created several specialized algorithms using analytical transformations. These developments formed the basis for the unique software systems "APP-MNIT" and "M-READ16" [14], developed by us, which are highly specialized, focused on solving a specific probabilistic-combinatorial problem described above in this article. These software systems were created to run on specialized clusters using parallel computing. The cluster "NUSC NSU" was used to carry out the necessary calculations. That helped to ensure a significant progress in solving the problem. Next, we will focus on the mathematical component of the developed methods.

First, let us note that for arbitrary fixed values of n and k, the desired solution can be represented in the form of an n-fold integral

$$P_{n,k}(\varepsilon) = n! \int\limits_{D_{n,k}(\varepsilon)} \cdots \int dx_1 \ldots dx_n, \tag{2}$$

where the domain $D_{n,k}(\varepsilon)$ of integration is given by the system of linear inequalities

$$\begin{cases} 0 < x_1 < x_2 < \ldots < x_{n-1} < x_n < 1, \\ x_{k+1} - x_1 > \varepsilon, \\ x_{k+2} - x_2 > \varepsilon, \\ \qquad \vdots \\ x_n - x_{n-k} > \varepsilon. \end{cases}$$

To calculate integral (2), we proposed a method of successive dimensionality reduction based on multiple cyclic replacement of the initial n-fold integral by a set of repeated integrals of reduced dimension with already defined integration limits for each variable. On the basis of this recursive-cyclic algorithm, a system for the analytical calculation of probabilities was designed and implemented, which calculate the required polynomial dependencies in the form of functions of the continuous parameter ε.

The calculations made it possible to find a complete set of partial formulas in all ranges of variation of the continuous parameter ε for all values of the integer parameters n and k up to $n = 14$. It should be noted that the calculations are associated with the need to perform a large amount of routine operations on setting the limits of integration, checking intermediate systems of inequalities for consistency and direct integration in n-dimensional space, which is almost impossible to do “manually” even for $n = 4$. Therefore, all the necessary software calculations were carried out on high-performance computing clusters.

2. Generalizing formulas $P_{n,k}(\varepsilon)$ found using software, analytical and discrete-combinatorial algorithms

At the next stage, we tried to establish the general laws governing the formation of probability formulas $P_{n,k}(\varepsilon)$ for the case $k > 1$ using the analysis of software-calculated particular solutions of the problem. A number of such analytical regularities were indeed discovered and subsequently rigorously proved. So, for $k = n - 1$ it was revealed, and later proved a simple dependence

$$P_{n,n-1}(\varepsilon) = 1 - n\varepsilon^{n-1} + (n-1)\varepsilon^n. \tag{3}$$

Let us recall that formula (3) describes the probability that if n points are randomly dropped on the interval $(0, 1)$, they will not all “collapse” into one compact ε-grouping.

For $k = n - 2$, the relationship is more complex:

$$P_{n,n-2}(\varepsilon) = \begin{cases} 1 - 2C_n^2\varepsilon^{n-2}(1-\varepsilon)^2 - 2\varepsilon^n, & 0 \le \varepsilon \le 1/2, \\ 1 - 2\varepsilon^n + (2\varepsilon - 1)^n - 2C_n^2\varepsilon^{n-2}(1-\varepsilon)^2, & 1/2 \le \varepsilon \le 1. \end{cases} \tag{4}$$

For $k = n - 3$, the dependence becomes so complicated that its reconstruction by analyzing particular software solutions is an independent laborious task:

$$\underset{n>6}{P_{n,n-3}(\varepsilon)} = \begin{cases} 1 - 2\varepsilon^n + C_n^1(6\varepsilon^n - 4\varepsilon^{n-1}) + C_n^2(-3\varepsilon^n + \varepsilon^{n-2}) + C_n^3(9\varepsilon^n - 18\varepsilon^{n-1} + \\ +12\varepsilon^{n-2} - 3\varepsilon^{n-3}), & 0 \le \varepsilon \le 1/2, \\ 1 - 2\varepsilon^n + (2\varepsilon - 1)^n + C_n^1(1-\varepsilon)(-2\varepsilon^{n-1} + 2(2\varepsilon - 1)^{n-1}) + \\ +C_n^2(1-\varepsilon)^2(\varepsilon^{n-2} + (2\varepsilon - 1)^{n-2}) - 3C_n^3\varepsilon^{n-3}(1-\varepsilon)^3, & 1/2 \le \varepsilon \le 1. \end{cases} \tag{5}$$

For $\varepsilon \to 0$, an asymptotic formula, common for arbitrary n, is established

$$P_{n,2}(\varepsilon) = C_n^0 + C_n^2(-n+2)\varepsilon^2 + C_n^3(4n-10)\varepsilon^3 + C_n^4(3n^2-37n+86)\varepsilon^4 + C_n^5(-40n^2+ +394n-922)\varepsilon^5 + C_n^6(-15n^3+625n^2-5171n-12\,086)\varepsilon^6 + C_n^7(420n^3-10\,724n^2+ +79\,996n-18\,7002)\varepsilon^7 + C_n^8(105n^4-10\,570n^3+20\,5499n^2-142\,6841n+3\,336\,406)\varepsilon^8+ +C_n^9(5040n^4-155\,708n^3+226\,7664n^2-17\,317\,506n+52\,315\,558)\varepsilon^9 + C_n^{10}(-945n^5+ +189\,000n^4-15\,794\,625n^3+389\,687\,181n^2-3\,798\,029\,823n+12\,998\,966\,646)\varepsilon^{10} + o(\varepsilon^{10}). \quad (6)$$

The above formulas (3)–(6) are confirmed by both software calculations and direct analytical integration.

3. Generalized Catalan numbers in problems of analysis of random point fields

Software algorithms for fast calculation of formulas were created primarily to calculate specific analytical ratios for fixed values n and k. With their help, we also tried (by analogy with formula (1) for the probability that is valid for all values of n for a fixed $k = 1$) to find a general solution that is valid for all values of n for a fixed $k = 2$. Unfortunately, this task turned out to be difficult. This is due primarily to the fact that, in contrast to the case $k = 1$, the probability consists not of one, but of several piecewise-homogeneous fragments, continuously connected at the points of “connection”. Secondly, finding general patterns for each of the ranges of parameter’s ε variation required the creation of individual schemes for the transfer (reduction) of continuous tasks corresponding to these specific ranges to individual and very complex discrete-probabilistic subtasks. In our reduction scheme, in all of such subproblems (i. e., in all ranges of the parameter’s ε variation), multidimensional Catalan numbers arose (this is not unusual, Catalan numbers and their extensions often appear when analyzing random sequences [15, 16]).

Knowledge of their explicit form was required when ordering interdependent random number sequences. Most of these probabilistic-combinatorial problems turned out to be more convenient to formulate in a dictionary-linguistic form. As an example, below we give the formulation and the solution for one of these combinatorial problems, which led to a multidimensional generalization of the classical Catalan numbers. Its distinctive feature is that it is formulated in a dictionary-linguistic form, solved by purely geometric means, and is used to rank the interdependencies of sequences in the problems of analyzing random point fields.

3.1. Example of a combinatorial problem leading to a generalization of the classical Catalan sequence to the three-dimensional case

So, when proving one of the relations describing the probability of error-free reading of a random discrete image by an integrator with two threshold levels, we need to solve the following combinatorial problem:

Different words of length $(l+m+n)$ are composed of l symbols “a”, m symbols “b” and n symbols “c”. It is necessary to determine the total number of words $Q_{l,m,n}$ such that for each of them two conditions are simultaneously fulfilled:

– when viewing a word from left to right, the number of “b” characters encountered never exceeds the number of “a” characters encountered;
– when viewing a word from right to left, the number of “c” characters encountered never exceeds the number of “a” characters encountered.

Naturally, it is considered that $0 \le m,\, n \le l$.

3.2. Solution. Finding the formula $Q_{l,m,n}$ for the case $l > m + n - 2$

We searched for a formula in relation to problems in which the condition $l > m+n-2$ was a priori fulfilled (more precisely, for the applied problems of analysis of random point images that we solved, a more strict inequality was fulfilled: $l \geq m+n$). The formulated problem is easier to solve if we give it a geometric interpretation. So, we will consider various monotone paths on a three-dimensional discrete lattice in the coordinate system (X, Y, Z), which lead from the point $(0, 0, 0)$ to the point (l, m, n). Let us assign one of these paths to each word. In this case, the symbol "a" will correspond to the movement from the current point (i, j, k) to the neighbouring point $(i+1, j, k)$, the symbol "b" will correspond to the movement to the point $(i, j+1, k)$, and the "c" symbol — to the point $(i, j, k+1)$. Our task will be to find the total number of monotone paths from point $(0, 0, 0)$ to point (l, m, n) that do not go beyond the half-spaces $X \geq Y$ and $Z \geq X+n-1$ (i. e. the desired paths are limited not only by the original parallelepiped with sides l, m and n, but also by the planes $P_1 : X = Y$ and $P_2 : Z = X + n - l$).

We will solve this reformulated problem as follows: from the total number of monotone paths Q_0 leading from the point $(0, 0, 0)$ to the point (l, m, n), which is equal to

$$Q_0 = \frac{(l+m+n)!}{l!m!n!}$$

subtract the number of paths that go beyond at least one of the half-spaces $X \geq Y$ or $Z \geq X+n-1$. To do this, add the number of paths Q_1 that go beyond the bounding plane P_1 with the number Q_2 of paths that go outside the bounding plane P_2, and subtract from the resulting sum the number of paths Q_{12} that go beyond the bounding plane P_1 and the bounding plane P_2 (since they are counted twice in the sum).

Let us first find the number of paths Q_1 that go beyond the bounding plane P_1 at least once. To do this, we will use a standard technique that often helps in solving problems of this kind. First, note that in order for any monotonic path leading from the point $(0, 0, 0)$ to the point (l, m, n) to go beyond the half-space $X \geq Y$, it is necessary and sufficient to have a link in it of the form $(u, u, v) > (u, u+1, v)$, that is, a link in which the first point belongs to the plane P_1 and therefore enters the half-space $X \geq Y$, and the next point of the path already lies outside of it. This statement does not require proof.

Based on this, for any path that goes beyond the half-space $X \geq Y$ at least once, we will find the first link of the form $(u, u, v) > (u, u+1, v)$. Further, returning back from the point $(u, u+1, v)$ to the starting point of the path $(0, 0, 0)$, we will perform the following mirror transformation: each movement back along the X axis will be replaced by a movement back along the Y axis, and vice versa, each move backward along the Y axis will be replaced with a move backward along the X axis. Move backward along the Z axis will be left unchanged. As a result, the "corrected" path will start not at the point $(0, 0, 0)$, but at the point $(-1, +1, 0)$. Leaving the final section of the original path from the point $(u, u+1, v)$ to the point (l, m, n) unchanged, we get a combined corrected path leading from the point $(-1, +1, 0)$ to the point (l, m, n). That is, to each monotonic path leading from the point $(0, 0, 0)$ to the point (l, m, n) and at least once going beyond the half-space $X \geq Y$, we have assigned a unique and quite definite path leading from the point $(-1, +1, 0)$ to the point (l, m, n).

The converse is also true: for each monotonic path connecting the points $(-1, +1, 0)$ and (l, m, n), you can put in a one-to-one correspondence a certain path that leads from the point $(0,0,0)$ to the point (l,m,n) and at least once goes beyond the half-space $X \geq Y$. To do this, it is enough to note that any monotonic path from the point $(-1,+1,0)$ to the point (l,m,n) necessarily intersects the plane $P_3 : Y = X+1$, since the start and end points of such a path lie on different to the sides of the plane P_3. Let's find the point of the first tangency of the selected path and the plane P_3. Obviously, the link leading to tangency must have the form $(u-1,u+1,v) > (u,u+1,v)$. Now let us mirror the initial segment of the original path, ending at the point $(u,u+1,v)$, relative to the plane $Y = X+1$, and its final segment from the point $(u,u+1,v)$ to the point (l,m,n) will be left unchanged (note that such mirroring and the previously performed operation of replacing the X and Y directions during the return motion from the point $(u,u+1,v)$ to the origin are equivalent).

As a result, we get a completely accurate corrected path, starting at the point $(0,0,0)$ and ending at the point (l,m,n). Since under mirror reflection, the link of the original path $(u-1,u+1,v) > (u,u+1,v)$ goes into the link $(u,u,v) > (u,u+1,v)$, into at this point, the corrected path will go beyond the half-space $X \geq Y$. Thus, for any path connecting the points $(-1,+1,0)$ and (l,m,n), we have constructed an exact path from the point $(0,0,0)$ to the point (l,m,n), going beyond the half-space $X \geq Y$ at least once. Thus, a one-to-one correspondence between the set of monotone paths connecting the points $(-1,+1,0)$ and (l,m,n) and the set of monotone paths connecting the points $(0,0,0)$ and (l,m,n) and at least once outside the half-space $X \geq Y$ is established. That's why

$$Q_1 = \frac{((m-1)+(l+1)+n)!}{(m-1)!(l+1)!n!} = \frac{(l+m+n)!}{(l+1)!(m-1)!n!}.$$

By analogy, using the absolute symmetry of the problem with respect to m and n for the total number of paths that go beyond the half-space $Z \geq X+n-1$ at least once, we obtain the expression

$$Q_2 = \frac{(l+m+n)!}{(l+1)!m!(n-1)!}.$$

Now it remains to calculate the number of monotone paths Q_{12}, each of which, without going beyond the boundaries of the original parallelepiped, has sections that go both outside the half-space $X \geq Y$ and outside the half-space $Z \geq X+n-1$. Below we will show that between the set of paths Q_{12} and the set of monotone paths from the point $(-1,+1,0)$ to the point $(l+1,m,n-1)$, we can establish a one-to-one correspondence, which means that these two sets have the same cardinality.

First, we will demonstrate how to any monotone path leading from the point $(0,0,0)$ to the point (l,m,n) and going beyond both half-spaces $X \geq Y$ and $Z \geq X+n-1$, to put in a one-to-one correspondence a well-defined monotonic path from the point $(-1,+1,0)$ to the point $(l+1,m,n-1)$. Then, on the contrary, to any monotonic path leading from the point $(-1,+1,0)$ to the point $(l+1,m,n-1)$, we will put in a one-to-one correspondence a monotonic path leading from the point $(0,0,0)$ to the point (l,m,n) and going beyond both half-spaces $X \geq Y$ and $Z \geq X+n-1$. Let's start with the first geometric build, i. e. any path from the point $(0,0,0)$ to the point (l,m,n) and going beyond both half-spaces $X \geq Y$ and $Z \geq X+n-1$, we assign a single monotone path from the point $(-1,+1,0)$ to the point $(l+1,m,n-1)$. To do this, note that any monotone path leading from the point $(0,0,0)$ to the point (l,m,n) and going beyond both of these half-spaces must have the following property: if you build all the exits of this path outside the half-spaces $X \geq Y$ and $Z \geq X+n-1$ in their order, then the first exit outside the half-space $Z \geq X+n-1$

will occur only after the last exit outside the half-space $X \geq Y$ ends. Let us prove this statement by contradiction.

Suppose that there is a point on the path (x_1, y_1, z_1) that goes beyond the half-space $Z \geq X + n - 1$, which precedes some point on the path (x_2, y_2, z_2) that lies outside the half-space $X \geq Y$. Then, for the coordinates of the point (x_1, y_1, z_1) lying outside the half-space $Z \geq X + n - 1$, the two-sided inequality $0 \leq z_1 < x_1 + n - l$ is true, whence it follows that $l - n < x_1$. Similarly, if a point (x_2, y_2, z_2) lies outside the limits of the half-space $X \geq Y$, then the two-sided inequality $x_2 < y_2 \leq m$ must hold, which implies that $x_2 < m$. Since, according to our assumption, the point (x_1, y_1, z_1) precedes the point (x_2, y_2, z_2), then $x_1 \leq x_2$. Combining the three obtained inequalities, we get $l - n < x_1 \leq x_2 < m$. Due to the presence of two strict inequalities in this system, we will have $l \leq m + n - 2$, which contradicts the initial condition $l > m + n - 2$. Hence, our assumption that in the case $l > m + n - 2$ there may be some monotone path from the point $(0, 0, 0)$ to the point (l, m, n), in which one of the exits outside the half-space $Z \geq X + n - 1$ precedes the exit of the path beyond the half-space $X \geq Y$ is false. Therefore, all such paths should schematically look like this:

$$(0,0,0) > \cdots > (x_1, x_1, z_1) > (x_1, x_1 + 1, z_1) > \cdots >$$
$$> (x_2, y_2, x_2 + n - l) > (x_2 + 1, y_2, x_2 + n - l) > \cdots > (l, m, n).$$

The first mirror reflection of this path is carried out with its initial segment starting at the point $(0, 0, 0)$ and ending at the point $(x_1, x_1 + 1, z_1)$, which is the first point of the path lying outside the half-space $X \geq Y$. Coming back from the point $(x_1, x_1 + 1, z_1)$ to the origin of coordinates $(0, 0, 0)$, we will each time instead of a negative step along the X axis take a negative step along the Y axis, and vice versa — instead of a negative step along the Y axis we will make a negative step along the X axis. It is easy to see that with such a transformation, the corrected path will start not at the point $(0, 0, 0)$, but at the point $(-1, +1, 0)$. The second mirror reflection will be carried out with the final section of the path starting at the point $(x_2 + 1, y_2, x_2 + n - l)$ and ending at the point (l, m, n). For definiteness, as before, we will assume that the point $(x_2 + 1, y_2, x_2 + n - l)$ is the first point of the path going beyond the half-space $Z \geq X + n - 1$ (generally speaking, there can be several points with coordinates having the form $(x_2 + 1, y_2, x_2 + n - l)$).

The transformation of the final section of the path should be as follows: each step along the X axis should be replaced with a step along the Z axis, and each step along the Z axis should be replaced with a step along the X axis. It is easy to see here that the end point of the corrected path instead of the point (l, m, n) will be the point $(l + 1, m, n - 1)$. Thus, to any monotone path leading from the point $(0, 0, 0)$ to the point (l, m, n) and going beyond both half-spaces $X \geq Y$ and $Z \geq X + n - l$, we have assigned a well-defined monotonic path from the point $(-1, +1, 0)$ to the point $(l + 1, m, n - 1)$.

The converse statement that any monotone path leading from the point $(-1, +1, 0)$ to the point $(l + 1, m, n - 1)$ can be put in a one-to-one correspondence with a certain path leading from the point $(0, 0, 0)$ to the point (l, m, n) and going beyond both half-spaces $X \geq Y$ and $Z \geq X + n - l$, is proved as follows. Note that the path from the point $(-1, +1, 0)$ to the point $(l + 1, m, n - 1)$ necessarily intersects the plane $P_3 : Y = X + 1$, since these points lie on different sides of the plane P_3. Let us find the point of the first tangency of this path and plane $P_3 : Y = X + 1$. The link immediately preceding the tangency obviously has the form $(u - 1, u + 1, v) > (u, u + 1, v)$. Conducting a mirror reflection relative to plane P_3 of the initial section of the path from the point $(-1, +1, 0)$ to the point $(u, u + 1, v)$, we get the first correction of the original path.

It will now start not at the point $(-1, +1, 0)$, but at the point $(0, 0, 0)$, then pass through the point $(u, u+1, v)$ and still end at the point $(l+1, m, n-1)$. Since, as a result of mirroring, the link of the original path $(u-1, u+1, v) > (u, u+1, v)$ goes into the link $(u, u, v) > (u, u+1, v)$, then at this point the corrected path passes from the half-space $X \geq Y$ to the half-space $X < Y$. Next, let's notice that the start and end points of the new corrected path $(0, 0, 0)$ and $(l+1, m, n-1)$ lie on opposite sides of the plane $P_4 : Z = X + n - l - 1$. This means that the corrected path must cross the plane $Z = X + n - l - 1$. Find the point at which the first tangency of this original path and the plane P_4 occurs: $Z = X + n - l - 1$. The coordinates of this point, obviously, have the form $(x, y, x+n-l-1)$, i.e. it is the first point of the path lying outside the half-space $Z \geq X + n - l$. If to the previously obtained with the help of mirroring the initial section of the path, starting at the point $(0, 0, 0)$ and ending at the point $(u, u+1, v)$, add the central section of the original path from the point $(u, u+1, v)$ to the point $(x, y, x+n-l-1)$, then we get a corrected path consisting of two segments and connecting points $(0, 0, 0)$ and $(x, y, x+n-l-1)$, which contains at least one point from the half-spaces $X \geq Y$ and $Z \geq X + n - l$.

Now it remains to carry out the second transformation of the original path, namely: the final segment from the point $(x, y, x+n-l-1)$ to the end point $(l+1, m, n-1)$ is mirrored relative to the plane $P_4 : Z = X + n - l - 1$. In fact, as noted, this mirroring is equivalent to each X step being replaced by a Z step on the final leg of the path, and vice versa, each Z step being replaced by an X step. As a result, the final segment of the corrected path will end not at the point $(l+1, m, n-1)$, but at the point (l, m, n). Putting three segments together, we get a corrected combined path from point $(0, 0, 0)$ to point (l, m, n), which at least once goes beyond the half-spaces $X \geq Y$ and $Z \geq X + n - l$.

Thus, a one-to-one correspondence between the set of monotone paths connecting the point $(-1, +1, 0)$ with the point $(l+1, m, n-1)$ and the set of monotone paths connecting the point $(0, 0, 0)$ with a point (l, m, n) and going out at least once both outside the half-space $X \geq Y$ and outside the half-space $Z \geq X + n - l$, is established. That's why

$$Q_{12} = \frac{(l+m+n)!}{(l+2)!(m-1)!(n-1)!}$$

and the solution of the problem on the number of monotone paths from the point $(0, 0, 0)$ to the point (l, m, n) that do not go beyond the half-spaces $X \geq Y$ and $Z \geq X + n - l$ (as well as the solution of the problem presented in the statement of the article about the number of three-character words), in the case $l > m + n - 2$ is given by the formula

$$Q_{l,m,n} = Q_0 - Q_1 - Q_2 + Q_{12} = \frac{(l+m+n)!}{l!m!n!} - \frac{(l+m+n)!}{(l+1)!(m-1)!n!} - \\ - \frac{(l+m+n)!}{(l+1)!m!(n-1)!} + \frac{(l+m+n)!}{(l+2)!(m-1)!(n-1)!}. \quad (7)$$

The obtained relation for Q_l, m, n, considered in the area $\{l > m + n - 2; m, n \leq l\}$ is a three-dimensional extension of the classical Catalan numbers. Moreover, in the case $\{n = 0\}$ we have a two-dimensional extension, and in the case $\{n = 0; l = m\}$ we have the classical Catalan sequence.

4. Generalizing formulas $P_{n,2}(\varepsilon)$ found using generalized Catalan numbers

In the case $k = 2$ for the probabilities $P_{n,2}(\varepsilon)$, we failed to find a general compact analytic relation similar to formula (1) for the probability for $k = 1$. However, using all the above computer-based and discrete-combinatorial tools, including software-analytical calculations and generalized Catalan numbers, we established and subsequently proved a number of new and previously unknown relations. For even values of $n = 2m$ on the segment $1/m < \varepsilon < 1/(m-1)$, the previously stated hypothesis formula is rigorously proved

$$P_{2m,2}(\varepsilon) = \frac{1}{m+1} C_{2m}^{m} (1 - (m-1)\varepsilon)^{2m}. \tag{8}$$

For even values of $n = 2m$ on the segment $1/(m+1) < \varepsilon < 1/m$, the formula is established

$$P_{2m,2}(\varepsilon) = C_{2m}^{m}(1-(m-1)\varepsilon)^{2m} - C_{2m}^{m-1}(1-(m-1)\varepsilon)^{2m} - C_{2m}^{m-2}(1-m\varepsilon)^{m+2}(1-(m-2)\varepsilon)^{m-2} + \\ + 2C_{2m}^{m-3}(1-m\varepsilon)^{m+3}(1-(m-2)\varepsilon)^{m-3} - C_{2m}^{m-4}(1-m\varepsilon)^{m+4}(1-(m-2)\varepsilon)^{m-4}. \tag{9}$$

For odd values $n = 2m+1$ on the segment $1/(m+1) < \varepsilon < 1/m$, the formula is established

$$P_{2m+1,2}(\varepsilon) = C_{2m+1}^{m+1}(1-m\varepsilon)^{m+1}(1-(m-1)\varepsilon)^{m} - 2C_{2m+1}^{m+2}(1-m\varepsilon)^{m+2}(1-(m-1)\varepsilon)^{m-1} + \\ + C_{2m+1}^{m+3}(1-m\varepsilon)^{m+3}(1-(m-1)\varepsilon)^{m-2}. \tag{10}$$

It should be noted that the multiplier $\frac{1}{m+1} C_{2m}^{m}$ on the right side of our expression (8) exactly corresponds to the classical Catalan sequence. In the proof of formula (9), we directly used relation (7) given in the previous section, which specifies the explicit form of the generalized Catalan numbers. But to obtain and prove formula (10), the use of relation (7) turned out to be insufficient, therefore, in this case, we additionally used the technique of finding monotonic paths in Weyl chambers [17]. All the software calculations carried out confirm the exact analytical formulas obtained (8)–(10).

Conclusion

The results presented in this article were obtained with the help of specialized tools of machine analytics, as well as with the use of generalized Catalan numbers, which made it possible to transfer the inherently continuous problem of finding probabilistic formulas to the category of discrete-combinatorial ones.

The efficiency of the proposed discrete-combinatorial methods allows us to hope for further progress in solving the described “continuous” problem, up to finding a general analytical dependence that is valid for arbitrary values of the integer parameters n and k in all ranges of variation of the continuous parameter ε. The availability of such a generalized analytical solution will provide researchers with an important tool for assessing the degree of randomness of the analyzed point images.

Acknowledgements. This work was partially supported by the Russian Foundation for Basic Research (project № 19-01-00128) and the Ministry of Science and Higher Education of the Russian Federation (project № AAA-A17-117052410034-6).

References

[1] **Kosykh V., Gromilin G., Yakovenko N.** Small objects detection in two-color images with spatially non-stationary background. CEUR Workshop Proceedings. 2020; (2534):284–287.

[2] **Cyganek B.** Object detection and recognition in digital images: theory and practice. N.Y.: Wiley; 2013: 516.

[3] **Ablameiko S.V., Anischenko V.V., Lapitsky V.A., Tuzikov A.V.** Medical information technologies and systems. Minsk: OIPI NAS Belarus; 2007: 135.

[4] **Mcphee S., Papadakis M., Rabow M.** Current medical diagnosis and treatment, 57th ed. N.Y.: McGraw-Hill Education; 2018: 1952.

[5] **Feder C.** Computerized medical diagnosis: a novel solution to an old problem. West Conshohocken: Infinity Publishing; 2010: 172.

[6] **Shannon C.** The mathematical theory of communication, 16th ed. Illinois: The University of Illinois Press; 1971: 144.

[7] **Reznik A.L., Soloviev A.A., Torgov A.V.** Algorithms for optimal localization of a random point-pulse source uniformly distributed over a search interval. Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications). 2018; 28(2):354–361. DOI:10.1134/S1054661818020177.

[8] **Reznik A.L., Tuzikov A.V., Soloviev A.A., Torgov A.V., Kovalev V.A.** Time-optimal algorithms focused on the search for random pulsed-point sources. Computer Optics. 2019; 43(4):605–610. DOI:10.18287/2412-6179-2019-43-4-605-610.

[9] **Gnedenko B.V., Belyayev Y.K., Solovyev A.D.** Mathematical methods of reliability theory. N.Y.: Academic Press; 1969: 506.

[10] **Czichos H.** Handbook of technical diagnostics: fundamentals and application to structures and systems. Berlin: Springer; 2013: 575.

[11] **Reznik A.L., Tuzikov A.V., Soloviev A.A., Torgov A.V.** Analysis of random point images with the use of symbolic computation codes and generalized Catalan numbers. Optoelectronics, Instrumentation and Data Processing. 2016; 52(6):529–536. DOI:10.3103/S8756699016060017.

[12] **Wilks S.** Mathematical statistics, 2nd edition. N.Y.: Wiley; 1962: 644.

[13] **Parzen E.** Modern probability theory and its applications. N.Y.; London: John Wiley and Sons, Inc.; 1960: 480.

[14] **Reznik A.L., Torgov A.V., Soloviev A.A.** Certificate of registration of the computer program "M-READ16" (ROSPATENT, No. 2016662343). 2016.

[15] **Saračević M., Adamović S., Maček N., Selim A., Pepíć S.** Source and channel models for secret-key agreement based on Catalan numbers and the lattice path combinatorial approach. Journal of Information Science and Engineering. 2021; 37(2):469–482. DOI:10.6688/JISE.202103_37(2).0012.

[16] **Gardner M.** Catalan numbers: an integer sequence that materializes in unexpected places. Scientific American. 1976; 234(6):120–125.

[17] **Gessel I.M., Zeilberger D.** Random walk in a Weyl chamber. Proceedings of the American Mathematical Society. 1992; 115(1):27–31.

Вычислительные технологии, 2023, том 28, № 1, с. 61–71. © ФИЦ ИВТ, 2023 ISSN 1560-7534
Computational Technologies, 2023, vol. 28, no. 1, pp. 61–71. © FRC ICT, 2023 eISSN 2313-691X

ВЫЧИСЛИТЕЛЬНЫЕ ТЕХНОЛОГИИ

DOI:10.25743/ICT.2023.28.1.007

Программы символьных вычислений и обобщенные числа Каталана в задачах анализа случайных точечных структур

А. Л. Резник[1], А. В. Тузиков[2], А. А. Соловьев[1,*], А. В. Торгов[1]

[1]Институт автоматики и электрометрии СО РАН, 630090, Новосибирск, Россия
[2]Объединенный институт проблем информатики НАН Беларуси, 220012, Минск, Беларусь
*Контактный автор: Соловьев Александр Анатольевич, e-mail: `solowey@rambler.ru`
Поступила 30 апреля 2022 г., принята в печать 16 мая 2022 г.

Аннотация

Предложены, обоснованы и программно реализованы новые алгоритмы расчета точных аналитических зависимостей, возникающих при решении различных теоретических и прикладных задач, связанных с анализом случайных точечных структур. Отличительной особенностью работы является то, что достигнутые в ней новые научные результаты получены с помощью оригинальных методов исследования на основе программ компьютерной аналитики с использованием обобщенных многомерных чисел Каталана.

Ключевые слова: случайное изображение, компьютерные аналитические выкладки, случайные локальные группировки, обобщенные числа Каталана.

Цитирование: Резник А.Л., Тузиков А.В., Соловьев А.А., Торгов А.В. Программы символьных вычислений и обобщенные числа Каталана в задачах анализа случайных точечных структур. Вычислительные технологии. 2023; 28(1):61–71. DOI:10.25743/ICT.2023.28.1.007. (на английском)

Благодарности. Исследование выполнено при частичной финансовой поддержке РФФИ (проект № 19-01-00128) и Министерства науки и высшего образования Российской Федерации (проект № AAA-A17-117052410034-6).