HIGH PERFOMANCE COMPUTING AND GENERALIZED CATALAN NUMBERS IN IMAGE PROCESSING TASKS

A.L.Reznik, A.V. Torgov, A.A. Soloview

Institute of Automation and Electrometry, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russia

The paper describes the use of high performance computing and the generalized Catalan numbers to find the complex mathematical formulas that are responsible for the reliability of the error-free digital images reading.

Keywords: high performance computing, generalized Catalan numbers, image processing

Introduction

There are many examples how solution of simply formulated task leads to serious scientific problems. This work is clear example of that: it required the development of special research techniques, based on huge volume of analytical calculations, creation of new concept of multi-dimensional generalized Catalan numbers and solving other complex problems.

Formulation of the problem

Our research in the field of image processing, and in particular, the determination of the reliability of error-free digital images reading using multiple-threshold-level integrators, led us to «seeming» simple, but extremely difficult probability task:

«Let ***n*** points $x_1, x_2, ..., x_n$ be randomly dropped on an interval (0,1), i.e., there are ***n*** independent tests of a random variable uniformly distributed in the interval (0,1). Find the probability $P_{n,k}(\varepsilon)$ of an event that there is no subinterval of length $\Omega_\varepsilon \subset (0,1)$ containing more than ***k*** points».

The solution for this problem was found in 1960 [1], but only for k = 1 (1):

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

Also, there are some several known asymptotic relations [2].

Solution of the problem

Problem's solution may be represented as a multivariate integral

$$P_{n,k}(\varepsilon) = n! \int \dots \int_{D_{n,k}(\varepsilon)} dx_1 \dots dx_n, \qquad (2)$$

where domain $D_{n,k}(\varepsilon) \subset R^n$ is described by a set of linear inequalities (3)

$$\begin{cases} 0 < x_1 < x_2 < \dots < x_{n-1} < x_n < 1, \\ x_{k+1} - x_1 > \varepsilon, \\ x_{k+2} - x_2 > \varepsilon, \\ \vdots \\ x_n - x_{n-k} > \varepsilon. \end{cases} \qquad (3)$$

For simplest case (k=1), multivariate integral (2) calculation over domain (3) can be reduced to repeated integral (4)

$$P_{n,1}(\varepsilon) = n! \int_{(n-1)\varepsilon}^{1} dx_n \left\{ \int_{(n-2)\varepsilon}^{x_n-\varepsilon} dx_{n-1} \dots \left\{ \int_{2\varepsilon}^{x_4-\varepsilon} dx_3 \left\{ \int_{\varepsilon}^{x_3-\varepsilon} dx_2 \left\{ \int_{0}^{x_2-\varepsilon} dx_1 \right\} \right\} \right\} \right\}. \qquad (4)$$

Successive integration over variables x_1, x_2,..., x_n, gives a known formula (1).
Unfortunately, in case of k>1 it is not possible to represent integral (2) in domain (3) as a single repeated integral. Because of that, problem solution is so complicated that there is no exact analytic solution even for k=2.
So, firstly, we created specialized software based on analytical calculations for high performance computing cluster, which helped us to calculate particular solutions for fixed n and k for $P_{n,2}(\varepsilon)$ up to n = 14 [3].
Using particular computer's solutions, we was able to establish (first stage) and mathematically prove (second stage) new analytical relations [4].
Strict mathematical prove required the creation the new concept of multi-dimensional generalized Catalan numbers.

Extensions of Catalan numbers

Here are examples of Catalan Numbers and theirs extensions.

Task 1
N symbols «a» and N «b» symbols are involved in the formation of the words. How many of such words, if they are viewed from left to right – number of met symbols «b» is never more than the number of met symbols «a»?
Solution:

$$WORD1(N) = \frac{(2N)!}{N!N!} - \frac{(2N)!}{(N+1)!(N-1)!} = \frac{1}{N+1}\binom{2N}{N} \quad (5)$$

Equation (5) had been found by Leonard Euler (many applications lead to the problem), but has the name of Belgian mathematical Catalan, who lived a century later.

Task 2
Na symbols «a» and Na symbols «b» are involved in the formation of the words. How many of such words, if they are viewed from left to right - number of met symbols «b» is never more than the number of met symbols «a» (assuming Na ≥ Nb)?
Solution:

$$WORD2(N_a, N_b) = \frac{(N_a + N_b)!}{N_a!N_b!} - \frac{(N_a + N_b)!}{(N_a + 1)!(N_b - 1)!} \quad (6)$$

This is the simplest extension of the Catalan numbers. When Na = Nb we have classical Catalan sequence.

Task 3
Na symbols «a» and Na symbols «b» are involved in the formation of the words. How many of such words, if they are viewed from left to right - number of met symbols «b» is never more than the number of met symbols «a» more then k (assuming k ≥ 0 and Na+k ≥ Nb).
Solution:

$$WORD3_k(N_a, N_b) = \frac{(N_a + N_b)!}{N_a!N_b!} - \frac{(N_a + N_b)!}{(N_a + k + 1)!(N_b - k - 1)!} \quad (7)$$

This is further Catalan number's expansion. Note that for k = 0, task №2 is equivalent to the task №3, and if Na = Nb and k = 0, we obtain the problem №1.
We have also obtained another three new Catalan number's extensions.

New formulas for error-free image reading

Using this new concept of multi-dimensional Catalan numbers, we managed to prove new formulas, responsible for error-free digital images reading [5].
For even n=2m in area 1/m<ε <1/(m-1) true is the relation:

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

For odd n=2m+1 in area 1/(m+1)<ε <1/m, the probability Pn,k(ε) is represented as:

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

Conclusion

The work is a prime example of interdisciplinary research that requires knowledge from various scientific fields. Our successes give us hope for further progress in resolving the main task (finding its general analytical solutions). Developed methods of using specialized software for analytic transformation and new concept of multi-dimensional generalized Catalan numbers were key tools in our research of error-free digital images reading, but they can also “advance” many other scientific problem to a new level.

References

1. Parzen, E., Modern Probability Theory and Its Applications // John Wiley and Sons, Inc., New York-London, 1960.
2. S.Wilkes. Mathematical statistics. M., Mir, 1967, c.632.
3. Reznik A.L., Efimov V.M., Solov’ev A.A., Torgov A.V. Reliability of readout of random point fields with a limited number of threshold levels of the scanning aperture // Optoelectronics, Instrumentation and Data Processing. 2015. T. 50. № 6. C. 582-588.
4. Reznik A.L., Efimov V.M., Solovev A.A. Computer-analytical calculation of the probability characteristics of readout of random point images // Optoelectronics, Instrumentation and Data Processing. 2011. T. 47. № 1. C. 7-11.
5. Reznik A.L., Efimov V.M., Solovev A.A., Torgov A.V. Generalized Catalan numbers in problems of processing of random discrete images // Optoelectronics, Instrumentation and Data Processing. 2011. T. 47. № 6. C. 533-536.