Известия Алтайского государственного университета. 2025. № 1 (141). С. 129–134.
Izvestiya of Altai State University. 2025. No 1 (141). P. 129–134.

Научная статья
УДК: 519.688
DOI: 10.14258/izvasu(2025)1-18

Применение аналитических методов и инструментов компьютерной алгебры для расчета инвариантных характеристик случайных точечных изображений

Александр Львович Резник[1], *Александр Анатольевич Соловьев*[2]

[1]Институт автоматики и электрометрии СО РАН, Новосибирск, Россия, reznik@iae.nsk.su
[2]Институт автоматики и электрометрии СО РАН, Новосибирск, Россия, soloviev@iae.nsk.su

Original article

Application of Analytical Methods and Computer Algebra Tools for Calculating Invariant Characteristics of Random Point Images

Alexander L. Reznik[1], *Alexander A. Soloviev*[2]

[1]Institute of Automation and Electrometry SB RAS, Novosibirsk, Russia, reznik@iae.nsk.su
[2]Institute of Automation and Electrometry SB RAS, Novosibirsk, Russia, soloviev@iae.nsk.su

Аннотация. При решении целого ряда задач, связанных с обработкой изображений, важнейшим моментом является правильный выбор признаков, по которым определяется степень «аномальности» исследуемого изображения. Предложен набор классификационных признаков случайных точечных изображений для их использования в задачах выявления в обрабатываемом потоке данных аномальных сгущений или, наоборот, выявления областей и фрагментов с увеличенным межэлементным разбросом. Существенное отклонение численных значений признаков, вычисляемых по ансамблю анализируемых опытных изображений, от заранее рассчитанных теоретических значений, соответствующих случайным точечным изображениям, может указывать на наличие в массиве обрабатываемых данных регулярной компоненты или «аномальной» составляющей. Приведены теоретические и программные расчеты, выполненные с применением специализированных средств компьютерной алгебры. Представленные в статье классификационные признаки и вероятностные зависимости являются устойчивыми инвариантами, характеризующими случайное точечное изображение.

Ключевые слова: случайные точечные изображения, порядковые статистики

Abstract. When solving a number of problems related to image processing, one of the most important points is the correct choice of features that determine the degree of “abnormality” of the image under study. The paper proposes a set of several classification features of random point images for their use in the tasks of detecting anomalous clumps in the processed data stream or, on the contrary, detecting areas and fragments with increased inter-element spread. Numerical values of features are calculated over the ensemble of analyzed images. Their significant difference from the pre-calculated theoretical values that correspond to random point images can help identify the presence of a regular component or an “abnormal” component in the processed data array. Theoretical and software calculations performed using specialized computer algebra tools are presented in the paper. The discussed classification features and probabilistic dependencies are stable invariants characterizing a random point image.

Keywords: random point images, ordinal statistics

Для цитирования: Резник А.Л., Соловьев А.А. Применение аналитических методов и инструментов компьютерной алгебры для расчета инвариантных характеристик случайных точечных изображений // Известия Алтайского государственного университета. 2025. № 1 (141). С. 129–134. DOI: 10.14258/izvasu(2025)1-18

For citation: Reznik A.L., Soloviev A.A. Application of Analytical Methods and Computer Algebra Tools for Calculating Invariant Characteristics of Random Point Images. *Izvestiya of Altai State University*. 2025. No 1 (141). P. 129–134. (In Russ.). DOI: 10.14258/izvasu(2025)1-18

Финансирование: работа поддержана Министерством науки и высшего образования Российской Федерации (проект № 124041700103-1).

Funding: this study was supported by Ministry of Science and Higher Education of the Russian Federation (project no. № 124041700103-1).

Введение

Во многих научно-прикладных задачах, связанных с распознаванием образов и анализом точечных структур, исследователю требуется определить, является ли анализируемый фрагмент типичным представителем определенного класса цифровых изображений либо его по ряду признаков следует отнести к разряду аномальных и подвергнуть дополнительному более углубленному анализу [1]. Например, образование аномальных сгущений на анализируемых биомедицинских изображениях может свидетельствовать о наличии серьезных патологических отклонений [2–4]. В метеорологических задачах, связанных с прогнозированием погоды, наличие значимых динамических изменений на анализируемых снимках может свидетельствовать о зарождении опасных атмосферных явлений [5, 6]. При решении задач подобного рода важнейшим моментом является правильный выбор признаков, по которым определяется степень «аномальности» исследуемого изображения. В прикладных задачах, имеющих дело с большими объемами обрабатываемой в реальном времени информации, к требованию эффективности используемых признаков добавляется требование простоты их расчета.

Цель настоящего сообщения — представить теоретические основы для возможного применения на практике простых в расчете классификационных признаков (инвариантных характеристик) точечных изображений, которые связаны со статистическими закономерностями образования сгущений или, наоборот, фрагментов с чересчур большим разбросом элементов. Для этого нами решены две одномерные задачи: рассчитаны первые моменты для порядковых статистик при случайном разбиении интервала и выполнен расчет этих же вероятностных характеристик для случайного разбиения окружности. Особенностью проведенных теоретических расчетов является то, что часть из них выполнена с применением специально разработанных программных средств компьютерной алгебры.

1. Математическое ожидание минимального расстояния между отсчетами при случайном разбиении интервала

Как показано в [7], вероятность $P_{n,k}(\varepsilon)$ события, состоящего в том, что при случайном бросании *n* точек $x_1, x_2, \ldots, x_n$ на интервал (0,1) внутри него не будет существовать ни одного подынтервала длиной ε, содержащего более *k* точек, может быть задана соотношением

$$P_{n,k}(\varepsilon) = n! \int \underset{D_{n,k}(\varepsilon)}{\ldots} \int dx_1 \ldots dx_n, \qquad (1)$$

где область интегрирования $D_{n,k}(\varepsilon) \subset R^n$ описывается системой линейных неравенств

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

Вводя в рассмотрение единичную функцию Хевисайда $1[z] = \begin{Bmatrix} 0, & z \le 0, \\ 1, & z \le 0, \end{Bmatrix}$ преобразуем исходный кратный интеграл (1) к виду

$$\begin{aligned} P_{n,k}(\varepsilon) = n! \int \ldots \int 1[x_1] 1[x_2] \ldots 1[x_n - x_{n-1}] 1[1 - x_n] 1[x_{k+1} - x_1 - \varepsilon] \times \\ \times 1[x_{k+2} - x_2 - \varepsilon] \times \ldots \times 1[x_n - x_{n-k} - \varepsilon] dx_1 \ldots dx_n, \end{aligned} \qquad (3)$$

Здесь, в отличие от интеграла (1), индикаторная функция области $D_{n,k}(\varepsilon)$ включена в подынтегральное выражение в виде произведения единичных функций, а интегрирование ведется уже по всему пространству R^n. Если теперь к полученному интегралу (3) последовательно по всем переменным $x_1, x_2, \cdots, x_n$ применить тождественное преобразование

$$\left(\prod_{j=1}^{l}1[x_r-\alpha_j]\right)\left(\prod_{i=1}^{m}1[\beta_i-x_r]\right)=\sum_{j=1}^{l}\sum_{i=1}^{m}1[x_r-\alpha_j]1[\beta_i-x_r]\left\{1[\beta_i-\alpha_j]\left(\prod_{\substack{q=1\\q\neq j}}^{l}1[\alpha_j-\alpha_q]\right)\left(\prod_{\substack{s=1\\s\neq i}}^{m}1[\beta_s-\beta_i]\right)\right\}, \qquad (4)$$

то для случая k=1 исходная вероятность $P_{n,1}(\varepsilon)$ представится единственным и легко интегрируемым повторным интегралом

$$P_{n,1}(\varepsilon)=n!\int_{(n-1)\varepsilon}^{1}dx_n\left\{\int_{(n-2)\varepsilon}^{x_n-\varepsilon}dx_{n-1}\cdots\left\{\int_{2\varepsilon}^{x_4-\varepsilon}dx_3\left\{\int_{\varepsilon}^{x_3-\varepsilon}dx_2\left\{\int_{0}^{x_2-\varepsilon}dx_1\right\}\right\}\right\}\right\}. \qquad (5)$$

В итоге получим формулу $P_{n,1}(\varepsilon)=(1-(n-1)\varepsilon)^n, \quad 0<\varepsilon<1/(n-1)$), которая совпадает с результатом, приводимым в [8, 9].

Заметим, что выражение $P_{n,1}(\varepsilon)$ есть вероятность того, что при случайном бросании n точек $x_1, x_2, \cdots, x_n$ на интервал (0,1) все они разлетятся между собой на расстояние, не меньшее ε. Отсюда следует, что функция распределения $F_\rho(x)$ случайной величины ρ, соответствующей минимальному расстоянию между отсчетами в парах (x_i, x_{i+1}), равна

$$F_\rho(x)=P(\rho<x)=1-(1-(n-1)x)^n, \quad 0<x<1/(n-1).$$

Таким образом, плотность вероятности распределения минимального расстояния между ближайшими отсчетами есть

$$f(x)=F'_\rho(x)=n(n-1)(1-(n-1)x)^{n-1}, \ \ 0<x<1/(n-1).$$

Соответственно, математическое ожидание минимального расстояния ρ между соседними отсчетами при случайном бросании n точек на интервал (0,1) будет равно $E[\rho]=\int_0^{1/(n-1)}xf(x)dx=\int_0^{1/(n-1)}xdF_\rho(x)$, или, интегрируя по частям,

$$E[\rho]=\int_0^{1/(n-1)}xdF_\rho(x)=[xF_\rho(x)]\Big|_{x=0}^{x=1/(n-1)}-\int_0^{1/(n-1)}F_\rho(x)dx=$$

$$=\left(\frac{1}{n-1}-0\right)-\left(\frac{1}{n-1}+\left[\frac{1}{n-1}\times\frac{1}{n+1}(1-(n-1)x)^{n+1}\right]\Bigg|_{x=0}^{x=1/(n-1)}\right)=\frac{1}{(n-1)(n+1)}=\frac{1}{n^2-1}. \qquad (6)$$

С учетом простоты расчета найденная величина $E[\rho]$ может быть принята в качестве одного из классификационных признаков случайного точечного изображения.

2. Математическое ожидание длины наименьшего фрагмента при случайном разбиении интервала

Отличие этой задачи от задачи, решавшейся в предыдущем разделе 1, состоит в следующем. В ней вместо нахождения среднего (по ансамблю реализаций) значения минимального расстояния ρ между случайно выброшенными на интервал (0,1) точками требуется определить среднее значение случайной величины $\xi^{(n)}$, соответствующей длине самого короткого из (n+1) фрагментов, на которые распадается интервал (0,1), если его разрезать в n случайно выбранных точках $x_1, x_2, \cdots, x_n$. По аналогии с системой неравенств (2), описывающей область интегрирования $D_{n,k}(\varepsilon)$ для интеграла (1), система линейных неравенств $D_{n,k}^{(\xi)}(\varepsilon)$, соответствующая событию, состоящему в том, что каждый из (n+1) фрагментов, на которые разбился интервал (0,1), имеет длину более ε, запишется в виде

$$D_{n,k}^{(\xi)}(\varepsilon)=\begin{cases}0<\varepsilon<x<x_2...<x_n<1-\varepsilon<1\\ x_2-x_1>\varepsilon\\ x_3-x_2>\varepsilon\\ ...\\ ...\\ x_n-x_{n-1}>\varepsilon\end{cases}. \qquad (7)$$

С применением преобразований (3) и (4) вероятность такого события выразится интегралом

$$P(\xi^{(n)} > \varepsilon) = n! \int\limits_{\varepsilon}^{1-n\varepsilon} dx_1 \left\{ \int\limits_{x_1+\varepsilon}^{1-(n-1)\varepsilon} dx_2 \cdots \left\{ \int\limits_{x_{n-3}+\varepsilon}^{1-3\varepsilon} dx_{n-2} \left\{ \int\limits_{x_{n-2}+\varepsilon}^{1-2\varepsilon} dx_{n-1} \left\{ \int\limits_{x_{n-1}+\varepsilon}^{1-\varepsilon} dx_n \right\}\right\}\right\}\right\} = (1-(n+1)\varepsilon)^n . \qquad (8)$$

Соответственно, функция распределения $F_{\xi^{(n)}}(x)$ случайной величины $\xi^{(n)}$, соответствующей длине наименьшего из (n+1) фрагментов, задается соотношением

$$F_{\xi^{(n)}}(x) = P(\xi^{(n)} < x) = 1-(1-(n+1)x)^n, \quad 0 < x < 1/(n+1).$$

Поэтому плотность вероятности распределения величины $\xi^{(n)}$ будет иметь вид

$$f_{\xi^{(n)}}(x) = F'_{\xi^{(n)}}(x) = n(n+1)(1-(n+1)x)^{n-1}, \; 0 < x < 1/(n+1).$$

Отсюда математическое ожидание случайной величины $\xi^{(n)}$ равно

$$E[\xi^{(n)}] = \int\limits_0^{1/(n+1)} x f_{\xi^{(n)}}(x)dx = \int\limits_0^{1/(n+1)} x dF_{\xi^{(n)}}(x) = [xF_{\xi^{(n)}}(x)]\Big|_{x=0}^{x=1/(n+1)} - \int\limits_0^{1/(n+1)} F_{\xi^{(n)}}(x)dx =$$
$$= [x(1-(1-(n+1)x)^n)]\Big|_{x=0}^{x=1/(n+1)} - \int\limits_0^{1/(n+1)} [1-(n+1)x)^n]dx = \frac{1}{n+1} - \frac{1}{n+1} + \int\limits_0^{1/(n+1)} (1-(n+1)x)^n dx = \qquad (9)$$
$$= -\frac{1}{(n+1)^2}(1-(n+1)x)^{n+1}\Bigg|_{x=0}^{x=1/(n+1)} = \frac{1}{(n+1)^2}.$$

Величина $E[\xi^{(n)}]$ может быть принята в качестве еще одного из классификационных признаков случайного точечного изображения.

3. Расчет первых моментов для порядковых статистик при случайном разбиении окружности

Задача, рассматриваемая в этом разделе, формулируется следующим образом: «Окружность единичной длины с помощью n случайных точек делится на n дуг. Дуги упорядочиваются по длине: $\eta_1^{(n)} < \eta_2^{(n)} < \ldots < \eta_n^{(n)}$. Требуется определить математическое ожидание $E[\eta_i^{(n)}], \; i = \overline{1,n}$ каждой из случайных величин $\eta_i^{(n)}$». Сразу отметим, что вместо набора из n случайных дуг может рассматриваться набор из n фрагментов, на которые распадается интервал (0,1) при его разрезании в (n-1) случайных точках. Поэтому в соответствии с формулой (9) математическое ожидание длины наименьшей из n дуг равно

$$E[\eta_1^{(n)}] = E[\xi^{(n-1)}] = \frac{1}{n^2}. \qquad (10)$$

Соответственно, математическое ожидание совокупной длины оставшихся (n-1) дуг будет равно $E[\sum\limits_{i=2}^{n} \eta_i^{(n)}] = 1 - \frac{1}{n^2}$, поэтому математическое ожидание второй по длине дуги $\eta_2^{(n)}$ составит $E[\eta_2^{(n)}] = (1-\frac{1}{n^2})/(n-1)^2 = \frac{n+1}{n^2(n-1)}$.

В общем случае для математического ожидания случайной величины $\eta_i^{(n)}$, если действовать в соответствии с формулой (10), должно выполняться соотношение

$$E[\eta_i^{(n)}] = (1 - \sum_{j=1}^{i-1} \eta_j^{(n)})/(n-i+1)^2. \qquad (11)$$

Общее решение рекуррентного уравнения (11) имеет вид

$$E[\eta_m^{(n)}] = \frac{n+1}{n} \times \frac{1}{(n-m+1)(n-m+2)}, \quad m = \overline{1,n}. \qquad (12)$$

Это решение может быть получено многими способами. Детальное описание одного из таких способов можно найти, например, в работе [10].

4. Расчет инвариантных характеристик случайных точечных изображений с применением специализированных систем компьютерной алгебры

В разделе 1 настоящего сообщения был представлен общий вид интегрального выражения, описывающего вероятность $P_{n,k}(\varepsilon)$ того, что при случайном бросании *n* точек на интервал (0,1) внутри него не будет существовать ни одного подынтервала длиной *ε*, содержащего более *k* точек. Алгоритмическое выполнение описанных в разделах 1–3 вычислительных схем, связанных с расчетом инвариантных характеристик случайных точечных изображений, основано на знании вероятности $P_{n,k}(\varepsilon)$ при *k*=1. К сожалению, для случая *k*>1 получить общий вид аналитической зависимости $P_{n,k}(\varepsilon)$ не удается. Поэтому ниже приводятся результаты, полученные нами с помощью специально разработанного пакета программ компьютерной алгебры, который применялся для нахождения точных аналитических решений $P_{n,k}(\varepsilon)$ для произвольных, но фиксированных значений *n* при *k*>1. Эффективное распараллеливание вычислительного процесса и его проведение на высокоскоростном вычислительном кластере Новосибирского государственного университета позволило проводить расчет формул $P_{n,k}(\varepsilon)$ для любых значений $k, n \leq 15$. Наборы таких частных формул в последующем использовались при отыскании общих ранее неизвестных аналитических зависимостей. В частности, для диапазона $\frac{1}{m+1} < \varepsilon < \frac{1}{m}$ получена и доказана общая для всех значений *n* формула

$$P_{n,2}(\varepsilon) = \begin{cases} (C_{2m}^{m} - C_{2m}^{m-1})(1-(m-1)\varepsilon)^{2m} - \{(1-m\varepsilon)^{m+2}(1-(m-2)\varepsilon)^{m-4} \times \\ \times[C_{2m}^{m-2}(1-(m-2)\varepsilon)^2 - 2C_{2m}^{m-3}(1-m\varepsilon)(1-(m-2)\varepsilon) + C_{2m}^{m-4}(1-m\varepsilon)^2]\}, \quad n=2m; \\ \\ (1-m\varepsilon)^{m+1}(1-(m-1)\varepsilon)^{m-2} \times [C_{2m+1}^{m+1}(1-(m-1)\varepsilon)^2 - \\ -2C_{2m+1}^{m+2}(1-m\varepsilon)(1-(m-1)\varepsilon) + C_{2m+1}^{m+3}(1-m\varepsilon)^2], \quad n=2m+1, \end{cases} \quad (13)$$

соответствующая вероятности того, что при бросании *n* случайных точек на интервал (0,1) сумма никаких двух смежных интервалов между соседними отсчетами не будет превышать *ε*. Эта величина может быть использована в качестве одной из инвариантных характеристик случайного точечного изображения.

Заключение

Приведенные в статье классификационные признаки и вероятностные зависимости являются устойчивыми инвариантами, характеризующими случайное точечное изображение. В реальных научно-прикладных задачах существенное отклонение аналогичных параметров, рассчитанных по набору анализируемых изображений, от выбранных теоретических инвариантов может свидетельствовать о влиянии на опытные данные систематической либо не учтенной заранее компоненты. Список представленных в настоящей работе устойчивых инвариантов случайных точечных изображений, основанных на порядковых статистиках, может быть существенно расширен, но применение таких инвариантов в практических научно-прикладных задачах затруднено сложностью и трудоемкостью их расчета.

Библиографический список

1. Webster J., Eren H. Measurement, Instrumentation, and Sensors Handbook. Boca Raton: CRC Press, 2018. 3559 p.

2. Dougherty G. Medical Image Processing, techniques and Applications. New York: Springer, 2011, 380 p.

3. Wójcik W., Pavlov S., Kalimoldayev M. Information Technology in Medical Diagnostics. London: CRC Press, 2019. 220 p.

4. Frangi A., Prince J., Sonka M. Medical Image Analysis. Cambridge: Academic Press, 2023. 698 p.

5. Berry R., Burnell J. The Handbook of Astronomical Image Processing. Richmond: Willmann-Bell. 2005. 684 p.

6. Chen C., Zhong J., Tan Y. Multiple-Oriented and Small Object Detection with Convolutional Neural Networks for Aerial Image // Remote Sensing. 2019. No 11. P. 2176–2200. DOI: 10.3390/rs11182176

7. Резник А.Л., Тузиков А.В., Соловьев А.А., Торгов А.В. Интеллектуальная программная поддержка в задачах анализа случайных цифровых изображений // Вычислительные технологии. 2018. Т. 23. № 5. С. 70–81. DOI: 10.25743/ICT.2018.23.5.007

8. David G. Ordinal Statistics. M.: Nauka. 1979. 335 p.

9. Parzen E. Modern Probability Theory and Its Applications. New York: John Wiley and Sons. 1960. 464 p.

10. Reznik A.L., Soloviev A.A. Software-Analytical Calculation of Invariant Characteristics of Random Point Images Based on Order Statistics // Pattern Recognition and Image Analysis. 2024. Vol. 35. No 3. P. 379-385. DOI: 10.1134/S1054661824700780

References

1. Webster J., Eren H. *Measurement, Instrumentation, and Sensors Handbook.* Boca Raton: CRC Press, 2018. 3559 p.

2. Dougherty G. *Medical Image Processing, Techniques and Applications.* New York: Springer, 2011. 380 p.

3. Wójcik W., Pavlov S., Kalimoldayev M. *Information Technology in Medical Diagnostics.* London: CRC Press, 2019. 220 p.

4. Frangi A., Prince J., Sonka M. *Medical Image Analysis.* Cambridge: Academic Press, 2023. 698 p.

5. Berry R., Burnell J. *The Handbook of Astronomical Image Processing.* Richmond: Willmann-Bell, 2005. 684 p.

6. Chen C., Zhong J., Tan Y. Multiple-Oriented and Small Object Detection with Convolutional Neural Networks for Aerial Image. *Remote Sensing.* 2019. No 11. P. 2176–2200. DOI: 10.3390/rs11182176

7. Reznik A.L., Tuzikov A.V., Soloviev A.A., Torgov A.V. Intellectual Program Support for the Analysis of Random Digital Images. *Computational Technologies.* 2018. Vol. 23. No 5. P. 70–81. DOI: 10.25743/ICT.2018.23.5.007 (In Russ).

8. David G. *Ordinal Statistics.* M.: Nauka, 1979. 335 p.

9. Parzen E. *Modern Probability Theory and Its Applications.* New York: John Wiley and Sons, 1960. 464 p.

10. Reznik A.L., Soloviev A.A. Software-Analytical Calculation of Invariant Characteristics of Random Point Images Based on Order Statistics. *Pattern Recognition and Image Analysis.* 2024. Vol. 35. No 3. P. 379–385. DOI: 10.1134/S1054661824700780

Информация об авторах

А.Л. Резник, доктор технических наук, заведующий лабораторией вероятностных методов исследования информационных процессов, Институт автоматики и электрометрии СО РАН, Новосибирск, Россия;

А.А. Соловьев, кандидат технических наук, старший научный сотрудник лаборатории вероятностных методов исследования информационных процессов, Институт автоматики и электрометрии СО РАН, Новосибирск, Россия.

Information about authors

A.L. Reznik, Doctor of Sciences in Technology, Head of the Laboratory of Probability Research Methods for Information Processing, Institute of Automation and Electrometry SB RAS, Novosibirsk, Russia;

A.A. Soloviev, Candidate of Sciences in Technology, Senior Researcher at the Laboratory of Probability Research Methods for Information Processing, Institute of Automation and Electrometry SB RAS, Novosibirsk, Russia.