

Grzegorz Harańczyk

Uniwersytet Jagielloński, StatSoft Polska

Andrzej Sokołowski

Uniwersytet Ekonomiczny w Krakowie

ROZWAŻANIA O ROZKŁADZIE ODLEGŁOŚCI OD PUNKTÓW KRATOWYCH W ROZKŁADZIE RÓWNOMIERNYM

1. Wstęp

Zauważmy, że rozkład równomierny ma wiele ciekawych własności. Między innymi każdy ucięty lub warunkowy rozkład k -wymiarowego rozkładu równomiernego jest też rozkładem równomiernym. Rozkład równomierny łatwo sprowadzić do przedziału $[0,1]$ (ogólnie – wielowymiarowej kostki o boku jednostkowym). W literaturze w odniesieniu do takiego rozkładu znajdujemy określenie „zmienna lilipucia” (*lilliputian variable*); np. [Kowalczyk, Pleszczyńska, Ruland 2004].

Pociecha i Sokołowski [1989] zaproponowali szereg statystyk przeznaczonych do testowania wielowymiarowej jednostajności. Założyli, że jeżeli rozkład generujący jest jednostajny, to punkty empiryczne powinny być rozłożone mniej więcej „równomiernie”, bez skupień i dziur. Zaproponowali, aby strukturę punktów opisać dendrytem. Wiązadła takiego dendrytu powinny mieć podobną (statystycznie) długość w przestrzeni o zadanym wymiarze. Wiązadła dendrytu można łatwo znormalizować tak, aby ich suma była równa 1. We wspomnianej pracy zaproponowano pięć statystyk testowych; są nimi: miara podobieństwa struktury dendrytu do struktury, w której wszystkie elementy są tej samej wielkości, najkrótsze wiązadło dendrytu, najdłuższe wiązadło dendrytu, rozstęp oraz wariancja długości wiązadła. Podano wartości krytyczne uzyskane drogą symulacji komputerowych.

Denkowska i Sokołowski [2008] pokazali wielowymiarowy rozkład o strukturze grupowej, jednostajny przedziałami, na który „nie reagują” wymienione statystyki.

Jeżeli uznamy, że pewnego rodzaju modelem wielowymiarowego rozkładu równomiernego jest krata regularna, to warte rozważenia wydaje się poszukiwanie rozkładu odległości punktów empirycznych od najbliższego punktu kratowego.

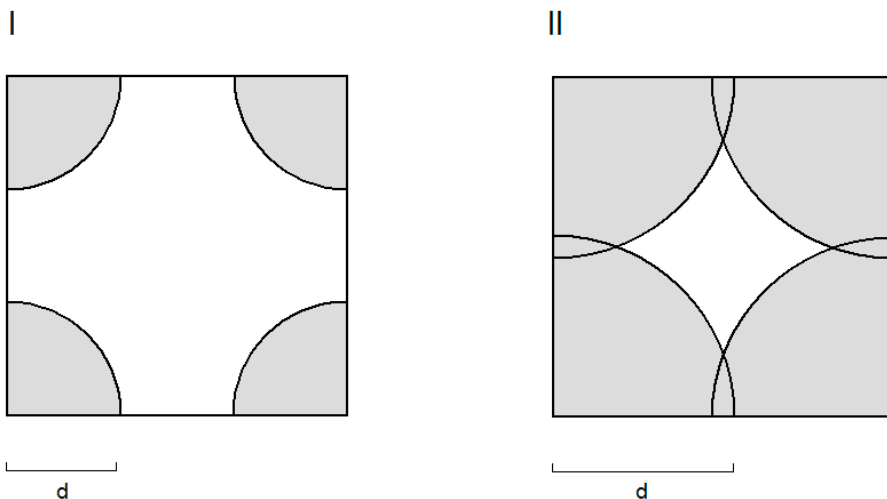
1. Rozkład odległości od punktów kratowych w przypadku jednowymiarowym

Zauważmy, że do badania rozkładu odległości wystarczy rozważanie pojedynczej „celi” kraty, gdyż w uogólnionej kratce regularnej cele są takie same. Bez zmniejszania ogólności rozważań zakładamy, że bok celi jest jednostkowy. W przestrzeni jednowymiarowej cele (okienka karty) są odcinkami o jednakowej długości. Ponieważ rozważamy odległość do najbliższego węzła, punktów odległych od najbliższego końca o mniej niż d jest $2d$. Mamy zatem

$$F(d) = P(X < d) = 2d \quad d \in \left[\frac{0,1}{2}, \right]. \quad (1)$$

2. Rozkład odległości od punktów kratowych w przypadku dwuwymiarowym

Modelem pojedynczego okna w uogólnionej regularnej kratce dwuwymiarowej jest kwadrat o boku jednostkowym. Na rysunku 1 pokazano model ilustrujący mechanizm konstrukcji funkcji gęstości rozkładu odległości od punktów kratowych. Prawdopodobieństwo jest równe polu zaznaczonemu na schematach. Dla odległości nie większych od 0,5 pole składa się z czterech ćwiartek koła. Dla $d > 0,5$ te rosnące ćwiartki zaczynają się pokrywać, zatem od pola koła należy odjąć cztery jednakowe części wspólne.



Rys. 1. Zbiory punktów odległych od węzłów kraty o mniej niż d ; I – gdy $d \in \left[\frac{0,1}{2}, \right]$,

$$\text{II – gdy } d \in \left[\frac{1}{2}, \frac{\sqrt{2}}{2} \right]$$

Źródło: opracowanie własne.

Dystrybuanta ma postać

$$F(d) = P(X < d) = \begin{cases} \pi d^2 & d \in \left[0, \frac{1}{2}\right] \\ \pi d^2 - 4K_1 & d \in \left[\frac{1}{2}, \frac{\sqrt{2}}{2}\right], \end{cases} \quad (2)$$

gdzie

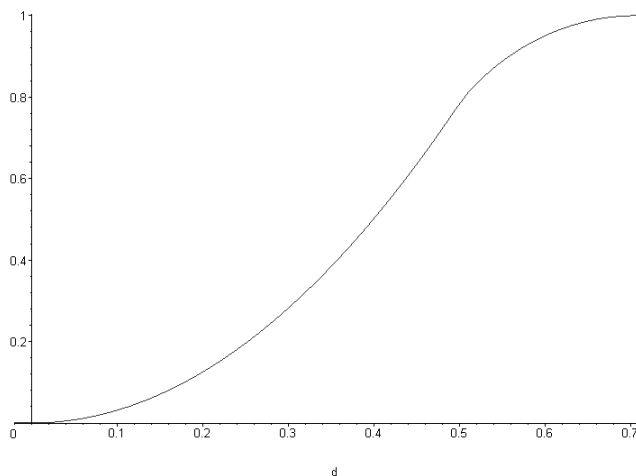
$$K_1 = 2 \int_{\frac{1}{2}}^d \sqrt{d^2 - x^2} dx = d^2 \frac{\pi}{2} - d^2 \arcsin\left(\frac{1}{2d}\right) - \frac{1}{2} \sqrt{d^2 - \frac{1}{4}}. \quad (3)$$

Po podstawieniu (3) do (2) otrzymujemy

$$F(d) = P(X < d) = \begin{cases} \pi d^2 & d \in \left[0, \frac{1}{2}\right] \\ 4d^2 \arcsin\left(\frac{1}{2d}\right) + 2\sqrt{d^2 - \frac{1}{4}} - \pi d^2 & d \in \left[\frac{1}{2}, \frac{\sqrt{2}}{2}\right]. \end{cases} \quad (4)$$

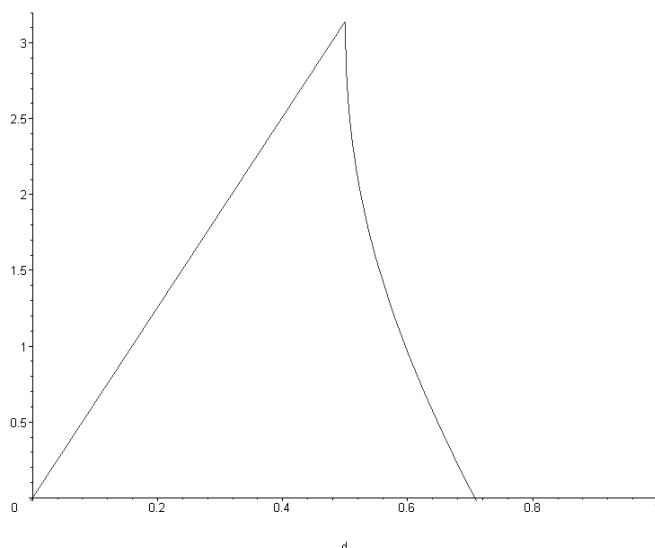
Oczywiście widać, że

$$F(0) = 0, \quad F\left(\frac{\sqrt{2}}{2}\right) = 1.$$



Rys. 2. Dystrybuanta rozkładu odległości od punktów kratowych dla dwuwymiarowego rozkładu równomiernego

Źródło: opracowanie własne.



Rys. 3. Funkcja gęstości rozkładu odległości od punktów kratowych dla dwuwymiarowego rozkładu równomiernego

Źródło: opracowanie własne.

Na rysunku 2 przedstawiono postać dystrybuanty rozkładu odległości od punktów kratowych w dwuwymiarowym rozkładzie równomiernym, a na rys. 3 – jego gęstość.

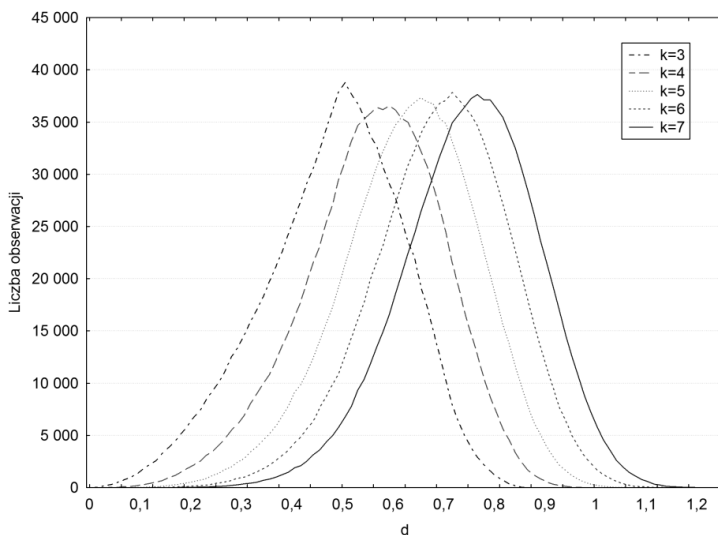
2. Rozkład odległości od punktów kratowych w przypadku k -wymiarowym

W przypadku trójwymiarowego rozkładu równomiernego modelem pojedynczej celi kraty regularnej jest sześcian. Przy konstrukcji rozkładu rozważa się osiem kul (z każdej kuli jej $1/8$ objętości) o środkach położonych w wierzchołkach sześcianu oraz promieniu d . Dla $d < 1/2$ dystrybuanta jest równa objętości kuli. Dla $d > 1/2$ sytuacja komplikuje się o tyle, że najpierw zaczynają pokrywać się kule o środkach leżących w wierzchołkach będącymi wierzchołkami tej samej ściany sześcianu. Spotkanie z kulą wychodzącą z przeciwnego wierzchołka nastąpi, gdy d będzie równe połowie przekątnej sześcianu.

W przypadku k -wymiarowym modelem pojedynczej celi regularnej kraty jest hipersześcian o liczbie wierzchołków równej 2^k . Dystrybuanta i funkcja gęstości będą wieloczęściowe, co wynika z częściowego pokrywania się hiperkul o środkach w tych wierzchołkach. Pierwsza część dystrybuanty (dla $d < 1/2$) jest równa objętości hiperkuli o promieniu d i wynosi

$$F(d) = \frac{\pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2} + 1\right)} d^k. \quad (5)$$

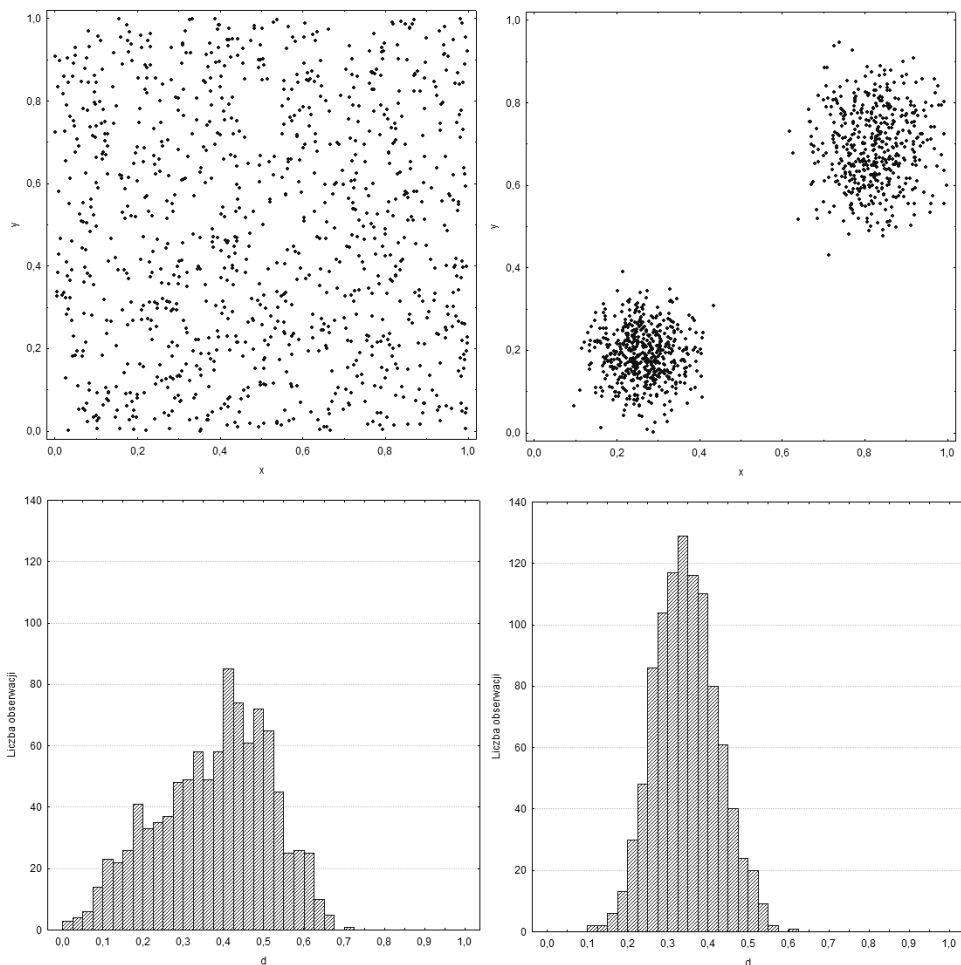
Jak widać, przy wzroście wymiaru k rośnie skomplikowanie funkcji rozkładu. Wydaje się, że w takim przypadku wygodniejsze jest szacowanie kształtu rozkładu przez badania symulacyjne. Na rysunku 4 przedstawiono przykłady oszacowanych funkcji gęstości odległości od punktów kratowych dla k od 3 do 7. Można zauważyć, że oprócz przesuwania się położenia rozkładu – co jest związane ze znanym efektem „rozdymania się” przestrzeni klasyfikacji – wydaje się on asymptotycznie przybliżać do rozkładu symetrycznego. Zagadnienie to wymaga jednak dalszych badań symulacyjnych.



Rys. 4. Funkcja gęstości rozkładu odległości od punktów kratowych dla przypadku wielowymiarowego oszacowana na podstawie symulacji (1 mln przypadków)

Źródło: opracowanie własne.

Testowanie wielowymiarowej równomierności (czyli braku struktury grupowej) można przeprowadzać przez porównywanie empirycznego rozkładu od punktów kratowych uzyskanego z analizowanych danych z takim samym rozkładem generowanym przy założeniu wielowymiarowej równomierności. Na rysunku 5 pokazano porównanie w przypadku $k = 2$, gdy zbiorowość badana składa się z dwóch grup. Przedstawione rozkłady odległości od punktów kratowych są zdecydowanie różne.



Rys. 5. Wyniki symulacji dla przypadku dwuwymiarowego.

Wykresy rozrzutu i poniżej odpowiadające im rozkłady odległości od punktów kratowych
 Źródło: opracowanie własne.

3. Podsumowanie

Jak pokazano, potencjalnym zastosowaniem proponowanego rozkładu jest testowanie wielowymiarowej równomierności przez porównanie zgodności empirycznego rozkładu odległości od punktów kratowych z rozkładem teoretycznym. Dalszy kierunek badań to poszukiwanie rozkładu odległości od swoistych punktów kratowych. Każdą oś w przedziale zmienności cechy dzielimy na $n-1$ równych odcinków. Otrzymujemy n równomiernie rozłożonych punktów. Są one podstawą wielowymiarowej kraty równomiernej. Jeżeli oryginalne wartości cech zamienimy

na rangi, to swoistym punktem kratowym nazywamy wierzchołek kraty regularnej wyznaczony przez te współrzędne rangowe.

Literatura

- Denkowska S., Sokołowski A. (2008), *O pewnym wielowymiarowym rozkładzie prawdopodobieństwa*, II Konferencja Naukowa im. Prof. A. Zeliasia, Zakopane (w druku).
- Kowalczyk T., Pleszczyńska E., Ruland F. (2004), *Grade models and methods for data analysis*, Springer.
- Pociecha J., Sokołowski A. (1989), *Empirical tests of multidimensional uniformity*, "Control and Cybernetics", vol. 18, no 1, s. 81-86.

THE DISTRIBUTION OF DISTANCE FROM LATTICE POINTS IN UNIFORM DISTRIBUTION

Summary

A search for probability distribution of distance to lattice points in the uniform distribution is presented in the paper. Probability functions for one and two dimensional cases have been derived analytically. For higher dimensions, Monte Carlo analysis is advised in order to approximate the distribution.