Krzywa Gaussa i pomarańcze Fidela Castro

Rafał SZTENCEL

– Bo wie pan, jakby zmierzyć dużo razy ciśnienie, to zawsze wyjdzie krzywa Gaussa.

Doktor Konstanty R. do dociekliwego pacjenta

Mało kto nie słyszał o krzywej Gaussa, i co gorsza prawie wszyscy wierzą święcie, że po zbadaniu dowolnego zjawiska losowego jakoś tam w końcu ona wyjdzie, czyli pojawi się wykres funkcji

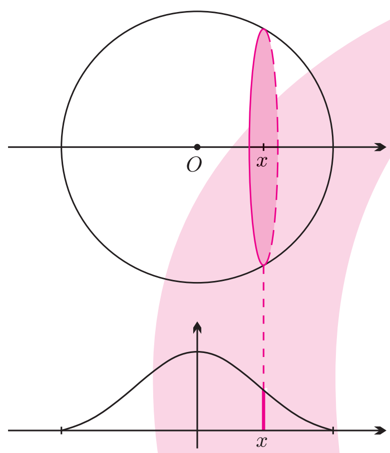
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Jest to zadziwiający fenomen społeczny, skoro 99% populacji (szczególnie absolwenci szkół średnich z ostatnich lat) nie potrafiłoby powiedzieć, co znaczą robaczki składające się na powyższy wzór, tym bardziej że zawiera on pierwiastek kwadratowy, koncept nieznany większości Europejczyków [1].

Dlaczego najważniejszy rozkład prawdopodobieństwa ma tak dziwaczną postać? Pokażemy, jaki jest jego związek z bardzo naturalnym rozkładem, a mianowicie jednostajnym na n-wymiarowej kuli o promieniu $\sqrt{n}$, a dokładniej – zbiorze opisanym nierównością

$$x_1^2 + x_2^2 + \ldots + x_n^2 \leqslant n.$$

Rozpatrzmy rzut tego rozkładu na średnicę kuli, czyli np. odcinek $[-\sqrt{n}, \sqrt{n}]$ na osi x_1. Jest to rozkład współrzędnej x_1. Nietrudno obliczyć jego gęstość.



Przekrój kuli płaszczyzną o równaniu $x_1 = x$ jest wyznaczony przez nierówność

$$\sqrt{x_2^2 + \ldots + x_n^2} \leqslant \sqrt{n - x^2} = \sqrt{n} \cdot \sqrt{1 - \frac{x^2}{n}} = r_{n-1}(x), \quad |x| \leqslant \sqrt{n},$$

jest więc $(n-1)$-wymiarową kulą o promieniu $r_{n-1}(x)$. Jak można się ostatnio dowiedzieć od studentów matematyki, objętość kuli jest funkcją wielomianową promienia; prośba o sprecyzowanie postaci wielomianu nierzadko kończy się sukcesem. Możemy zatem napisać gęstość rozkładu współrzędnej x_1:

$$f_n(x) = \begin{cases} c_n \left(1 - \frac{x^2}{n}\right)^{n/2} & |x| < \sqrt{n}, \\ 0 & |x| \geqslant \sqrt{n}. \end{cases}$$

Jeśli $n \to \infty$, to

$$f_n(x) \to f(x) = c \cdot e^{-\frac{1}{2}x^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Jest to może mniej oczywiste, niż się wydaje. Należy zauważyć, że stałe c_n można wyznaczyć z zależności

$$c_n \cdot \int_{-\sqrt{n}}^{\sqrt{n}} \left(1 - \frac{x^2}{n}\right)^{n/2} dx = 1,$$

jako że f_n jest gęstością i udowodnić jakoś (na przykład z tw. Lebesgue'a o zbieżności zmajoryzowanej), że

$$\int_{-\sqrt{n}}^{\sqrt{n}} \left(1 - \frac{x^2}{n}\right)^{n/2} dx \to \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}.$$

Rozkład Gaussa jest więc rozkładem granicznym współrzędnej x_1.

Bibliografia

[1] *Gazeta Wyborcza*, losowo wybrany numer z lata 2007.

Jeśli zastąpimy rozkład jednostajny na kuli rozkładem jednostajnym na sferze, wynik będzie ten sam. Nie trzeba powtarzać rachunków, wystarczy zdać sobie sprawę, że w wysokim wymiarze prawie cała masa kuli skupiona jest tuż pod powierzchnią sfery. Autor przekonał się o tym naocznie, gdy pod koniec lat siedemdziesiątych sprzedano mu w sklepie na Grochowie pomarańcze w kolorze szarozielonym, pochodzące z Kuby. Okazało się, że mają one niezwykle grubą skórę. Prosty rachunek pokazuje, że gdy grubość skóry wynosi 20% promienia, to objętość miąższu stanowi tylko 51,2% całości. W przestrzeni o większej liczbie wymiarów miąższu jest żałośnie mało.

Mało tego, kulę możemy zastąpić kostką (może to być kostka cukru z Kuby, nawet pusta w środku) o jednostkowej krawędzi i rzutować rozkład jednostajny na najdłuższą przekątną. Taka operacja na kwadracie da gęstość w kształcie trójkąta, ale już na kostce trójwymiarowej doprowadzi do gęstości sklejonej z wielomianów drugiego stopnia, której kształt będzie przypominać krzywą Gaussa. A jak udowodnić, że gęstość graniczna jest gaussowska?

Rozkład jednostajny na kostce $[-\frac{1}{2}, \frac{1}{2}]^n$ jest rozkładem wektora losowego $X = (X_1, \ldots, X_n)$, gdzie X_i są niezależnymi zmiennymi losowymi o rozkładzie jednostajnym na $[-\frac{1}{2}, \frac{1}{2}]$. Rzutowanie na najdłuższą przekątną jest równoważne z pomnożeniem X skalarnie przez wektor jednostkowy, wyznaczający kierunek przekątnej, czyli przejściem do zmiennej losowej

$$(X_1, \ldots, X_n) \circ \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right) = \frac{X_1 + \ldots + X_n}{\sqrt{n}},$$

która, zgodnie z centralnym twierdzeniem granicznym, ma asymptotycznie rozkład normalny o zerowej średniej i wariancji $D^2 X_1 = \frac{1}{12}$.