O minimach dwóch ważnych funkcji. Część trzecia

Elżbieta Guziejko, I Liceum Ogólnokształcące w Olecku
Zbigniew Paprzycki, Wydział Matematyki i Informatyki UWM

Rozważane w poprzednich częściach artykułu minima funkcji

$$f(x) = \sum_{i=1}^{n} a_i |x - x_i| \quad \text{oraz} \quad f(x) = \sum_{i=1}^{n} a_i (x - x_i)^2 \tag{1}$$

o dodatnich współczynnikach $a_i > 0$ i dowolnie ustalonych wartościach rzeczywistych $x_i \in R$ dla $i = 1,...,n$ mają bardzo ważne zastosowania zarówno teoretyczne jak i praktyczne w różnych działach gospodarki.

Wyjaśnimy to na prostym przykładzie. W tym celu wyobraźmy sobie dokonywanie n pomiarów $x_1, x_2, ..., x_n$ pewnej nieznanej wielkości x, które obarczone są losowymi błędami, powiedzmy „pomiarowymi”. Pomiary $x_1, x_2, ..., x_n$ w statystyce nazywamy próbą. Mamy wtedy sytuację

$$x_1 = x + e_1, \quad x_2 = x + e_2, ..., x_n = x + e_n,$$

gdzie różnice $e_i = x_i - x$ dla $i = 1,...,n$ są nieprzewidywalnymi i nieznanymi błędami pomiarowymi.

Jak w takiej sytuacji, mając wszystkie pomiary obarczone błędami, wyznaczyć prawdziwą wartość mierzonej wielkości x? W ogólnym przypadku nie jest to możliwe. Możemy jedynie podjąć próbę minimalizacji tzw. funkcji straty, za którą z reguły przyjmuje się funkcje (1).

W najprostszej postaci dla $a_1 = a_2 = ... = a_n = 1$ funkcja $f(x) = \sum_{i=1}^{n} |x - x_i| = \sum_{i=1}^{n} |e_i|$ reprezentuje sumę wartości bezwzględnych błędów, a wartość minimalizującą tę funkcję otrzymujemy po ustawieniu wartości bezwzględnych tak, aby $x_1 \le x_2 \le ... \le x_n$ i dodawaniu kolejnych $a_i = 1$, aż osiągniemy co najmniej połowę sumy wszystkich a_i równą w tym przypadku n, co jest opisane w części pierwszej artykułu. Będzie to w tym przypadku: $x_{\left(\frac{n+1}{2}\right)}$ dla n nieparzystych lub przedział $\left\langle x_{\left(\frac{n}{2}\right)}, x_{\left(\frac{n}{2}+1\right)} \right\rangle$ dla n parzystych (w praktyce przyjmujemy środek przedziału). Tak określoną wartość nazywamy medianą z próby:

$$med = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{dla} \quad n \text{ nieparzystych} \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} & \text{dla} \quad n \text{ parzystych} \end{cases} \tag{2}$$

Twierdzenie 1. Mediana z próby określona wzorem (2) minimalizuje sumę wartości bezwzględnych błędów.

Natomiast funkcja $f(x) = \sum_{i=1}^{n} (x - x_i)^2 = \sum_{i=1}^{n} (e_i)^2$ reprezentuje sumę kwadratów błędów, a wartością minimalizującą jest średnia arytmetyczna z próby $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, co opisuje część druga artykułu.

Twierdzenie 2. Średnia arytmetyczna $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ minimalizuje sumę kwadratów błędów.

Wartość x, dla której funkcja pierwsza lub druga przyjmuje wartość najmniejszą, nazywana jest oceną prawdziwej wartości mierzonej i w przypadku pierwszym jest to ocena metodą najmniejszych wartości bezwzględnych, a w drugim - ocena metodą najmniejszych kwadratów.

Przykład. Geodeta, mierząc kilkukrotnie odległość między dwoma punktami w terenie otrzymał wyniki:

631.26, 631.18, 631.19, 631.24, 631.27, 631.20, 631.23, 631.25.

Jaka jest prawdziwa odległość między mierzonymi punktami?

Rozwiązanie

Przyjmując funkcję straty postaci $f(x) = \sum_{i=1}^{n} |x - x_i| = \sum_{i=1}^{n} |e_i|$ w celu wyznaczenia minimum, należy wyniki pomiarów ustawić rosnąco:

631.18, 631.19, 631.20, 631.23, 631.24, 631.25, 631.26, 631.27.

Ponieważ $n = 8$ jest liczbą parzystą, to zgodnie ze wzorem na minimum wyprowadzonym w pierwszej części niniejszego artykułu otrzymujemy przedział domknięty $\langle 631.23, 631.24 \rangle$. W praktyce w takim przypadku przyjmuje się środek przedziału

$$\hat{x} = \frac{x_4 + x_5}{2} = \frac{631.23 + 631.24}{2} = 631.2350$$

Zauważmy, że stosując funkcję $f(x) = \sum_{i=1}^{n} (x - x_i)^2 = \sum_{i=1}^{n} (e_i)^2$ minimum jej otrzymamy dla

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{632.82}{8} = 631{,}2275 .$$

Stosując różne reguły (funkcje straty) minimalizujące, otrzymaliśmy - czego mogliśmy się spodziewać - różne oceny mierzonej odległości. Którą z nich należy przyjąć, zależy od naszego uznania i od własności danej oceny. Analizą własności tak otrzymanych ocen zajmuje się statystyka matematyczna, a dokładniej teoria estymacji. Przy okazji warto wiedzieć, że w geodezji zwyczajowo i powszechnie stosuje się kwadratową funkcję straty, czyli ocenianie metodą najmniejszych kwadratów. Jest to metoda bardzo rozpowszechniona ze względu na prostotę matematyczną.

Obie metody posiadają również duże znaczenie teoretyczne, ale omówienie tego wykracza poza ramy tego artykułu.

Zadanie do przemyślenia

Sprawdź, że mediana z próby 1, 2, 2, 2, 3, 3, 4, 4, 5 wyliczona ze wzoru (2) jest równa wartości x minimalizującej funkcję $f(x) = |x-5| + 2|x-4| + 2|x-3| + 3|x-2| + |x-1|$.