Co to są przedziały ufności

Doc. dr Ryszard ZIELIŃSKI

„W celu wyznaczenia ładunku elektronu wykonano pięć niezależnych pomiarów tego ładunku i otrzymano wyniki (w absolutnych jednostkach elektrostatycznych): $4{,}781 \cdot 10^{-10}$; $4{,}795 \cdot 10^{-10}$, $4{,}769 \cdot 10^{-10}$; $4{,}792 \cdot 10^{-10}$ i $4{,}779 \cdot 10^{-10}$. Pytanie: ile wynosi ładunek elektronu? Co w ogóle można na ten temat powiedzieć mając takie wyniki pomiarów?"

Stały Czytelnik Delty zauważył być może, że już kiedyś rozważaliśmy zadania tego typu: w artykule „O metodzie najmniejszych kwadratów" (Delta 8/1978) podaliśmy pewną zasadę postępowania przy rozwiązywaniu takich zadań. Zasada ta prowadziła do odpowiedzi, że ładunek elektronu jest równy średniej arytmetycznej otrzymanych wyników (a więc w naszym przypadku wynosi $4{,}783 \cdot 10^{-10}$). Można się spodziewać, że inna seria pięciu pomiarów dałaby inne wyniki i że wobec tego średnia tych nowych wyników byłaby różna od przed chwilą obliczonej. Mielibyśmy więc dwie różne odpowiedzi na to samo pytanie. Nie ma w tym nic dziwnego, bo przecież już pierwsza seria pomiarów dała nam pięć różnych odpowiedzi na interesującą nas wielkość, ale fakt, że odpowiedź ciągle zmienia się w miarę wykonywania coraz to nowych pomiarów może być irytujący.

Pewnym wyjściem z tej „niestabilnej" sytuacji byłoby zaokrąglenie wszystkich wyników pomiarów do dwóch cyfr znaczących; wtedy każdy wynik brzmiałby: $4{,}8 \cdot 10^{-10}$ i odpowiedź byłaby pozornie jednoznaczna. Co więcej, przy takich różnicach pomiędzy wynikami poszczególnych pomiarów, jakie obserwowaliśmy do tej pory, można by z dużą dozą pewności oczekiwać, że dalsze pomiary nie zmienią tej odpowiedzi. Odpowiedź taka jest jednak tylko pozornie jednoznaczna, bo powinniśmy jakoś zaznaczyć fakt zaokrąglania wyników pomiarów, co można zrobić pisząc $(4{,}8 \pm 0{,}05) \cdot 10^{-10}$ zamiast po prostu $4{,}8 \cdot 10^{-10}$. To postępowanie prowadzi do *oceny przedziałowej* interesującej nas wielkości (oceny za pomocą przedziału), zamiast poprzednio stosowanej *oceny punktowej* (oceny za pomocą pojedynczej liczby).

Jeżeli bardziej wnikliwie przyjrzymy się naszemu zadaniu, zauważymy dalsze problemy. Przede wszystkim powstaje pytanie czy jeżeli już decydujemy się na ocenę przedziałową, to musi to być ocena uzyskiwana — tak jak wyżej — za pomocą zaokrąglania wyników. Może „lepszy" byłby na przykład przedział, którego dolnym końcem byłby najniższy, a górnym najwyższy wynik z naszej serii pomiarów, a więc w rozważanym przypadku przedział $(4{,}769 \cdot 10^{-10}; 4{,}795 \cdot 10^{-10})$? Ten nowy przedział jest krótszy od poprzedniego, daje więc bardziej precyzyjne oszacowanie interesującej nas wielkości, ale...

Otóż istnieje pewne „ale", związane z następującymi okolicznościami. Wyobraźmy sobie, że decydujemy się wykonać jeszcze jeden, szósty, pomiar ładunku elektronu. Czy możemy być pewni, że wynik tego pomiaru znajdzie się w wyznaczonym przedziale $(4{,}769 \cdot 10^{-10}; 4{,}795 \cdot 10^{-10})$? Na pewno możemy być bardziej pewni, że nowy wynik znajdzie się w szerszym przedziale $(4{,}8 \pm 0{,}05) \cdot 10^{-10}$, ale czy z kolei tego możemy być pewni „na 100%"? Jak zbudować przedział, który byłby „dostatecznie pewny"? Pewność ta zresztą powinna dotyczyć nie tyle tego, że dalsze wyniki pomiarów znajdą się w tym przedziale, ile tego, że dany przedział zawiera szacowaną wielkość, w naszym przypadku „prawdziwą" wartość ładunku elektronu.

Oto pewne rozwiązanie zagadnienia estymacji przedziałowej. Oznaczmy wielkość szacowaną przez q (w naszym zadaniu jest to ładunek elektronu), wyniki pomiarów oznaczymy przez $x_1, x_2, \ldots, x_n$ (w naszym zadaniu mamy $n = 5$) i przez $I(x_1, x_2, \ldots, x_n)$ oznaczmy przedział zbudowany na podstawie tych wyników. Niech γ będzie ustaloną liczbą z przedziału $(0, 1)$.

Definicja. Przedział $I(x_1, x_2, \ldots, x_n)$ nazywamy *przedziałem ufności* na poziomie ufności γ, jeżeli

$$P\{q \in I(x_1, x_2, \ldots, x_n)\} = \gamma,$$

tzn. jeżeli przedział $I(x_1, x_2, \ldots, x_n)$ z prawdopodobieństwem γ zawiera szacowaną wielkość q.

Teoria przedziałów ufności stanowi obszerny fragment statystyki matematycznej. Zilustrujemy na bardzo ważnych dla zastosowań przykładach typowe rozwiązania podawane przez tę teorię, a na zakończenie wrócimy do naszego wyjściowego zadania szacowania ładunku elektronu.

Przykład 1 (rozkład wykładniczy). Za pomocą licznika Geigera-Müllera badamy intensywność pewnego źródła promieniowania. Intensywność źródła mierzymy średnią liczbą cząsteczek wypromieniowanych (dokładniej: zarejestrowanych przez licznik) w ciągu sekundy. Liczne badania zagadnienia promieniotwórczości pozwalają na sformułowanie prawa, że jeżeli intensywność promieniowania jest λ, to odstęp czasu pomiędzy dwoma kolejnymi „tyknięciami" licznika (oznaczymy ten odstęp przez T) jest zmienną losową o *rozkładzie wykładniczym*:

$$P\{T \leqslant t\} = 1 - e^{-\lambda t}, \quad \lambda > 0.$$

Wzór ten można również napisać w postaci:

$$(1) \qquad P\{\lambda T \leqslant t\} = 1 - e^{-t}.$$

Dla danej liczby γ, powiedzmy $\gamma = 0{,}95$, znajdźmy takie t_γ, żeby $1 - e^{-t_\gamma} = \gamma$.

Otrzymujemy (log oznacza logarytm naturalny), że $t_\gamma = -\log 0{,}05 = 2{,}996$.

Jeżeli więc w pewnym eksperymencie zaobserwujemy, że badany odstęp czasu wynosi T, to na mocy wzoru (1) mamy

$$P\left\{\lambda \leqslant \frac{2{,}996}{T}\right\} = 0{,}95$$

i przedział $\left(0, \dfrac{2{,}996}{T}\right)$ jest przedziałem ufności na poziomie ufności 0,95 dla interesującej nas intensywności λ. Innymi słowy: z prawdopodobieństwem 0,95 intensywność badanego źródła promieniowania mieści się w tak obliczonym przedziale.

Przykład 2 (rozkład normalny). W teorii pomiarów uzasadnia się, że w wielu typowych sytuacjach błąd pomiaru jest zmienną losową o *rozkładzie normalnym* (*rozkładzie Gaussa*). Oznacza to, że jeżeli X jest wynikiem pomiaru wielkości μ, to

$$(2) \qquad P\{X \leqslant x\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt,$$

gdzie σ jest parametrem rozkładu charakteryzującym dokładność pomiarów (σ nazywa się odchyleniem średnim lub odchyleniem standardowym, natomiast σ^2 nazywa się wariancją. Jest to dokładnie taka sama wariancja zmiennej losowej X, z jaką Czytelnik spotyka się w podręczniku W. Szlenka dla szkoły średniej).

Wzór (2) można oczywiście zapisać w postaci $\quad P\{X \leqslant x\} = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}t^2} dt$

lub w postaci

$$(3) \qquad P\left\{\frac{X-\mu}{\sigma} \leqslant x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt.$$

Prawa strona tego wzoru, funkcja $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt$

jest stablicowana, a jej tablice są łatwo dostępne; nazywa się ona dystrybuantą rozkładu normalnego. Jak w poprzednim przykładzie, możemy dla danej liczby $\gamma \in (0, 1)$ znaleźć takie x_γ, żeby $\Phi(x_\gamma) = \gamma$.

Jeżeli znamy wartość parametru σ, to ze wzoru (3) otrzymujemy natychmiast „jednostronny" przedział ufności dla μ na poziomie ufności γ, mamy bowiem

$$P\left\{\frac{X-\mu}{\sigma} \leqslant x_\gamma\right\} = \gamma$$

czyli $\quad P\{\mu \geqslant X - x_\gamma \cdot \sigma\} = \gamma$, co prowadzi do przedziału ufności $(X - x_\gamma \cdot \sigma, +\infty)$.

Ze wzoru (3) łatwo otrzymuje się wzór

$$(4) \qquad P\left\{\left|\frac{X-\mu}{\sigma}\right| \leqslant x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-\frac{1}{2}t^2} dt = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1.$$

Jeżeli z_γ jest taką liczbą, że $2\Phi(z_\gamma) - 1 = \gamma$, to otrzymujemy wzór

$$P\left\{\left|\frac{X-\mu}{\sigma}\right| \leqslant z_\gamma\right\} = \gamma,$$

który natychmiast daje „dwustronny" przedział ufności dla μ na poziomie ufności γ: $(X - z_\gamma \cdot \sigma, X + z_\gamma \cdot \sigma)$.

Przykład 3 (rozkład t Studenta). Jeżeli nie znamy wartości parametru σ, podane wzory stają się bezużyteczne i musimy postępować inaczej.

Przypuśćmy, że wykonaliśmy serię pomiarów pewnej nie znanej wielkości μ i że otrzymaliśmy wyniki $x_1, x_2, \ldots, x_n$. Niech

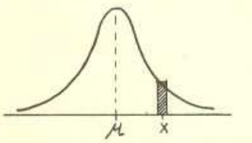
$$(5) \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Wielkość $\bar{x}$ jest po prostu średnią arytmetyczną naszych wyników, natomiast s^2 jest wielkością, za pomocą której szacuje się nieznane σ^2. W rachunku prawdopodobieństwa dowodzi się, że ma miejsce następujący wzór (analogiczny do wzoru (3)):

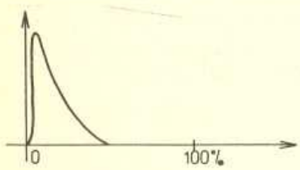
$$P\left\{\frac{\bar{x}-\mu}{s}\sqrt{n} \leqslant x\right\} = c_{n-1} \int_{-\infty}^{x} \frac{dt}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}},$$

gdzie c_{n-1} jest pewną stałą. Funkcja występująca po prawej stronie tego wzoru — oznaczymy ją przez $H_{n-1}(x)$ — nazywa się dystrybuantą *rozkładu t Studenta*. Funkcja ta, podobnie jak dystrybuanta rozkładu normalnego, jest stablicowana i tablice jej są łatwo dostępne.

Błąd pomiaru może mieć rozkład normalny tylko wtedy, gdy pomyłki w kierunku nadmiaru są tak samo prawdopodobne, jak pomyłki w kierunku niedomiaru. Na poniższym rysunku przedstawiono tzw. krzywą gęstości rozkładu normalnego; μ jest wielkością mierzoną, a rzędna w punkcie x jest proporcjonalna do prawdopodobieństwa, że wynik pomiaru będzie równy $x \pm \frac{1}{2}\,dx$.

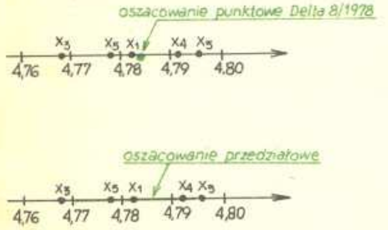


Jeżeli mierzymy np. zawartość (w procentach) jakiegoś składnika w wodzie rzecznej i ta zawartość jest bardzo niska, a nasz pomiar niezbyt precyzyjny, to krzywa opisująca rozkład prawdopodobieństwa błędu będzie miała raczej kształt taki, jak na poniższym rysunku, a więc bardzo różny od kształtu odpowiedniej krzywej rozkładu normalnego.



W takiej sytuacji budowa przedziału ufności wymaga zupełnie innego postępowania niż opisane w artykule.

γ	x_γ	z_γ	t_γ (dla $n = 5$)
0,9	1,282	1,645	2,132
0,95	1,645	1,960	2,776
0,99	2,326	2,576	4,604



Według najnowszych pomiarów ładunek elektronu wynosi $1{,}6021892 \cdot 10^{-19}$ kulomba, co w absolutnych jednostkach elektrostatycznych równa się $4{,}803242 \cdot 10^{-10}$.

Odpowiednikiem wzoru (4) jest teraz wzór

$$P\left\{\left|\frac{\bar{x}-\mu}{s}\sqrt{n}\right| \leqslant x\right\} = 2H_{n-1}(x)-1,$$

jeżeli więc znajdziemy taką liczbę t_γ, że $2H_{n-1}(t_\gamma)-1 = \gamma$, to na mocy wzoru

$$P\left\{\left|\frac{\bar{x}-\mu}{s}\sqrt{n}\right| \leqslant t_\gamma\right\} = \gamma \qquad (6)$$

otrzymujemy przedział ufności dla μ

$$\left(\bar{x}-\frac{s}{\sqrt{n}}t_\gamma,\quad \bar{x}+\frac{s}{\sqrt{n}}t_\gamma\right).$$

Przykład 4 (rozkład χ^2). W sytuacjach, gdy parametr σ nie jest znany, możemy być zainteresowani w jego oszacowaniu za pomocą przedziału ufności. Odpowiedni wzór, analogiczny do wzorów (1), (3) i (6), ma teraz postać

$$P\left\{\frac{(n-1)s^2}{\sigma^2} \leqslant x\right\} = a_{n-1}\int_0^x t^{\frac{n-3}{2}} e^{\frac{t}{2}}\, dt,$$

gdzie a_{n-1} jest pewną stałą. Funkcja występująca po prawej stronie tego wzoru jest dystrybuantą *rozkładu* χ^2 (*rozkładu chi-kwadrat*); tablice tej funkcji są również łatwo dostępne. Bardziej szczegółowe uwagi na temat wszystkich omawianych wyżej sposobów konstrukcji przedziałów ufności znajdzie Czytelnik w mojej książeczce „*Rachunek prawdopodobieństwa z elementami statystyki matematycznej*", Wydawnictwa Szkolne i Pedagogiczne 1976.

Na zakończenie powróćmy do naszego zadania szacowania ładunku elektronu. Przypuśćmy, że zgodnie z ogólną teorią błędów pomiarów mamy tu do czynienia z rozkładem normalnym i że nie znamy precyzji metody, za pomocą której otrzymaliśmy nasze wyniki, tzn. nie znamy parametru σ. W tej sytuacji powinniśmy skorzystać ze wzoru (6). Niech np. $\gamma = 0{,}9$. Z tablic otrzymujemy $t_\gamma = 2{,}132$. Na podstawie wzorów (5) mamy $\bar{x} = 4{,}783 \cdot 10^{-10}$ oraz $s = 0{,}011 \cdot 10^{-10}$. Jako rozwiązanie otrzymujemy przedział $(4{,}772 \cdot 10^{-10}; 4{,}794 \cdot 10^{-10})$, a więc znacznie krótszy od proponowanych na wstępie.

Rozwiązanie zadania M 178

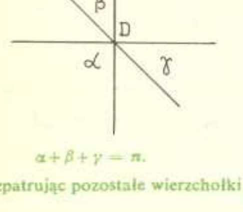
Do dowodu trzykrotnie wykorzystamy pożyteczną nierówność $xy+yz+zx \leqslant x^2+y^2+z^2$ (jest ona równoważna nierówności $\frac{1}{2}[(x-y)^2+(y-z)^2+(z-x)^2] \geqslant 0$).

Przyjmiemy kolejno $(x, y, z) = (a^4, b^4, c^4)$, (a^2b^2, b^2c^2, c^2a^2), (ab^2c, abc^2, a^2bc):

$a^8+b^8+c^8 \geqslant a^4b^4+b^4c^4+c^4a^4 \geqslant a^2b^4c^2+b^2c^4a^2+c^2a^4b^2 \geqslant a^2b^3c^3+b^2c^3a^3+c^2a^3b^3 = (bc+ca+ab)a^2b^2c^2$.

Rozwiązanie zadania M 179.

Niech A, B, C, D będą wierzchołkami czworościanu. Niech α będzie kątem, pod którym przecinają się okręgi opisane na ścianach mających wspólną krawędź AD, podobnie β niech odpowiada krawędzi BD, γ — CD, α' — BC, β' — CA, γ' — AB. Kąty α, β, γ przy wierzchołku D utworzone są przez styczne w punkcie D do okręgów opisanych na trójkątach BCD, ACD i ABD. Styczne te leżą w jednej płaszczyźnie, a mianowicie w płaszczyźnie stycznej w punkcie D do sfery opisanej na czworościanie. Jest więc (por. rysunek)



$$\alpha+\beta+\gamma = \pi.$$

Podobnie, rozpatrując pozostałe wierzchołki, otrzymujemy

$$\alpha+\beta'+\gamma' = \pi,$$
$$\alpha'+\beta+\gamma' = \pi,$$
$$\alpha'+\beta'+\gamma = \pi.$$

Dodając stronami równości pierwszą i drugą oraz trzecią i czwartą dostajemy $2\alpha+\beta+\beta'+\gamma+\gamma' = 2\alpha'+\beta+\beta'+\gamma+\gamma' = 2\pi$, skąd $\alpha = \alpha'$.

Rozwiązanie zadania M 180

Niech x_i będzie liczbą punktów zdobytych przez zawodnika, który zajął i-te miejsce w turnieju. Mamy $7 \geqslant x_1 > x_2 > x_3 > x_4 > x_5 > x_6 > x_7 > x_8$, a więc $x_2 \leqslant 6{,}5$. Ostatni czterej zawodnicy rozegrali między sobą 6 partii, punkty za te partie zostały przyznane tym zawodnikom, zatem $x_5+x_6+x_7+x_8 \geqslant 6$, skąd $x_2 \geqslant 6$. Musi więc być $x_2 = 6{,}5$ lub $x_2 = 6$. W pierwszym przypadku drugi zawodnik musiałby sześć partii wygrać i jedną zremisować, pierwszy z drugim by najwyżej zremisował i $x_1 \leqslant 6{,}5 = x_2$, co przeczy nierówności $x_1 > x_2$. Musi więc być $x_2 = 6$, zatem $x_5+x_6+x_7+x_8 = 6$, tzn. każdy z ostatnich czterech szachistów uzyskał punkty tylko z partii rozgrywanych z pozostałymi trzema z tej grupy, a więc przegrał z każdym z zawodników, którzy zajęli pierwsze cztery miejsca. W szczególności czwarty zawodnik wygrał z szóstym.

Rozwiązanie zadania F 60a

Zakładamy, że zmienna losowa X: „liczba rozpadów na minutę w objętości 1 litra" ma rozkład Poissona. Rzeczywiście stężenie J^{131} w atmosferze laboratorium daje przeciętnie $2{,}252 \cdot 10^{-12} \cdot 3{,}7 \cdot 10^{10} \cdot 60 \approx 5$ rozpadów na minutę i litr, zatem prawdopodobieństwo $P(X = k)$ tego, że w pomiarze kontrolnym zaobserwuje się k rozpadów, wynosi w przybliżeniu

$$\frac{5^k}{k!}e^{-5}.$$

Dopuszczalne stężenie J^{131} odpowiada $2{,}7 \cdot 10^{-12} \cdot 3{,}7 \cdot 10^{10} \cdot 60 \approx 6$ rozpadom na minutę i litr. Pomiar wykaże więc stężenie większe od dopuszczalnego, jeśli zaobserwuje się w nim co najmniej 7 rozpadów, wobec czego szukanym prawdopodobieństwem jest

$$P(X \geqslant 7) = 1-e^{-5}\sum_{i=0}^{6}\frac{5^i}{i!} \approx 0{,}24.$$

Rozwiązanie zadania F 60b.

Z danych zadania wynika, że proces rejestracji kwantów γ pochodzących z pojedynczego rozpadu mezonu K_S^0 na cztery kwanty γ możemy traktować jako schemat Bernoulliego złożony z czterech prób o prawdopodobieństwie sukcesu $\alpha = 0{,}4$. Liczba X zarejestrowanych kwantów γ jest wtedy ilością sukcesów w tym schemacie, i wobec tego prawdopodobieństwa zaobserwowania k kwantów γ wynoszą

$$P(X = k) = \binom{4}{k}\alpha^k(1-\alpha)^{4-k}, \quad k = 0, 1, 2, 3, 4.$$

Do uzyskania odpowiedzi na postawione pytania możemy teraz zastosować następujące (uproszczone!) wnioskowanie:

Interpretujemy prawdopodobieństwa jako względne częstości. Ponieważ z nieznanej ilości M rozpadów omawianego typu zarejestrowaliśmy 1000 rozpadów z czterema kwantami γ, a względna częstość ich występowania wynosi $\binom{4}{4}(0{,}4)^4(0{,}6)^0 = (0{,}4)^4 = 0{,}0256$, to

$$0{,}0256\, M = 1000,$$

skąd

$$M \approx 39062.$$

Interpretując dalej prawdopodobieństwa jako częstości znajdujemy:

oszacowanie liczby rozpadów nie zarejestrowanych

$$N(0) \approx 39062 \cdot (0{,}6)^4 \approx 5062$$

oraz oszacowania liczb rozpadów omawianego typu zarejestrowanych odpowiednio jako rozpady z 1, 2 i 3 kwantami γ:

$$N(1) = 39\,062 \cdot 4 \cdot (0{,}6)^3 \cdot 0{,}4 \approx 13\,500,$$
$$N(2) = 39\,062 \cdot 6 \cdot (0{,}6)^2 \cdot (0{,}4)^2 \approx 13\,500,$$
$$N(3) = 39\,062 \cdot 4 \cdot 0{,}6 \cdot (0{,}4)^3 \approx 6\,000.$$