

O średnich i wariancjach

Marek W. GUTOWSKI



Wyniki pomiarów

W szkole uczymy się, że lepszy wynik uzyskamy, gdy zamiast pojedynczego pomiaru wykonamy całą ich serię, a za końcowy rezultat uznamy średnią arytmetyczną indywidualnych pomiarów. Dodatkową zaletą tej procedury jest możliwość oszacowania niepewności pomiarowych (*rozrzutu*), zwanych dawniej błędami pomiarów. W szkole, ale i w prawdziwych laboratoriach badawczych, zwykle wykonujemy pomiary w kilkusobowych zespołach. Okazuje się, że odczytywanie tego samego wyniku przez różne osoby niekoniecznie daje takie same rezultaty. Procedura uśredniania, jak można podejrzewać, powinna w jakiś sposób przyczynić się do tego, aby rezultat końcowy był możliwie obiektywny, tj. możliwie najbardziej zbliżony do wartości prawdziwej. Pachnie to jednak trochę poznawaniem praw Przyrody za pomocą głosowania.

Zmienne losowe i centralne twierdzenie graniczne

Wyobraźmy sobie następujące zadanie: policzyć karpie w stawie hodowlanym. Żeby wykonać je rzetelnie, należałoby spuścić wodę ze stawu, „odcedzić” karpie, policzyć je, a w końcu przywrócić stan poprzedni. Takie postępowanie mogłoby bardzo zaszkodzić naszej hodowli, nie mówiąc już o kosztach jej przeprowadzenia (no i gdzie przechować tę wodę?). Musimy więc uciec się do sztuczki: najpierw wylawiamy pełną sieć ryb, a te, które są karpiami, specjalnie znakujemy. Wszystkie złowione ryby wpuszczamy z powrotem do stawu, a po pewnym czasie, np. na drugi dzień, ponawiamy połów i sprawdzamy, jaki ułamek wylowionych powtórnie karpie stanowią te oznakowane. Na podstawie tych danych oraz posiłkując się hipotezą, że ryby znaczone dobrze wymieszały się z pozostałymi, możemy już oszacować liczbę karpie w naszym stawie. Jeśliby jednak przyszło nam do głowy wykonać podobny pomiar kilka dni później, a nawet w tym samym czasie, lecz w innej części stawu, to otrzymalibyśmy inny wynik. Jeśli tak, to który z nich byłby poprawny? Kłopotliwe pytanie.

Wynik pomiaru jest *zmienną losową*. Nawet, jeśli uwierzymy, że od czasu znakowania do połowu kontrolnego nie zmieniła się ilość karpie (żaden nie zdechł, nic nie padło łupem kłusowników), to i tak za każdym razem możemy uzyskać inny rezultat. Składa się na to szereg czynników, będących poza naszą kontrolą. Czujemy jednakże, czysto intuicyjnie, że średnia z kilku pomiarów da nam wynik bliższy prawdy niż pomiar pojedynczy. Przez chwilę darujemy sobie uzasadnioną, lecz kąśliwą uwagę, że ta średnia może okazać się liczbą niecałkowitą...

Naszą intuicję znakomicie wzmacnia **centralne twierdzenie graniczne**, będące wybitnym i niekwestionowanym osiągnięciem gałęzi matematyki o nazwie *rachunek prawdopodobieństwa*. Orzeka ono, że jeśli mamy ciąg niezależnych zmiennych losowych X_1, X_2, X_3, \dots o jednakowych rozkładach, których

wartości średnie i wariancje istnieją (to znaczy mają skończoną wartość), to ciąg nowych zmiennych losowych

$$Y_k = \frac{1}{k} \sum_{i=1}^k X_i$$

(średnie arytmetyczne) ma granicę. Zmienna losowa

$$Y = \lim_{k \rightarrow \infty} Y_k$$

ma przy tym wartość średnią równą wartości średniej zmiennej X_1 . Co więcej, rozkład, do którego zdąża ciąg zmiennych Y_k , jest gaussowski (normalny). Zamiast X_1 moglibyśmy wypisać nazwę którejkolwiek ze zmiennych z rodziny X_i , gdyż mają one identyczne rozkłady, a więc także wartości średnie i wariancje. Autor usilnie nalega, aby zwrócić uwagę na założenie: zmienne losowe X_i mają mieć **skończone wartości średnie i wariancje**. Założenia twierdzenia można osłabić, nie wymagając, aby zmienne X_i miały identyczne rozkłady.

Przekład centralnego twierdzenia granicznego na język zrozumiały dla eksperymentatorów jest łatwy: średnia arytmetyczna serii pomiarów coraz lepiej odtwarza rzeczywistą wartość średnią wielkości mierzonej w miarę wzrostu liczby pomiarów. Są jednak dwa haczyki:

- twierdzenie nie mówi nic o szybkości zbieżności,
- zbieżność jest w sensie prawdopodobieństwa. Oznacza to, że w miarę wzrostu k – liczby indywidualnych pomiarów – prawdopodobieństwo dużych odchyłeń od prawdziwej wartości średniej maleje do zera. Nie oznacza to w żadnym wypadku, że *każdy* nowy pomiar zbliża nas do prawdziwej wartości średniej albo że to zbliżanie się następuje stale „z jednej strony”, tj. od dołu albo od góry; mogą, a nawet powinny, wystąpić zakłócenia.

Mamy więc problem: średnia z , powiedzmy, 10 pomiarów wcale nie musi być bliższa rzeczywistości,

niż np. średnia z 6 pomiarów. To ile właściwie pomiarów powinniśmy wykonać? Dobrą wskazówkę daje nam tu drugie ważne twierdzenie rachunku prawdopodobieństwa, mianowicie **nierówność Czebyszewa**, prawdziwa dla dowolnej zmiennej losowej X , której wartość średnia $\langle X \rangle$ i wariancja σ^2 są skończone:

$$P(|X - \langle X \rangle| \geq a) \leq \frac{\sigma^2}{a^2} \quad \text{dla } a > 0.$$

Kładąc $a = n\sigma(X_1)$ oraz przepisując nierówność dla zmiennej losowej Y_k , której wariancja wynosi

$$\sigma^2(Y_k) = \frac{\sigma^2(X_1)}{k},$$

otrzymamy

$$P(|Y_k - \langle Y \rangle| \geq n\sigma) \leq \frac{1}{kn^2} \quad \text{dla } n > 0.$$

Niestety, ocena ta jest zwykle mocno pesymistyczna i, co gorsza, nie daje się poprawić, gdyż istnieją takie zmienne losowe, dla których powyższa nierówność staje się równością. Jeśli jednak pracujemy ze zmiennymi X_i o rozkładzie normalnym (gaussowskim), to sytuacja jest o wiele lepsza. Najczęściej jednak wcale nie mamy pewności, że nasze zmienne losowe mają rozkład normalny...

W naszym przykładzie z karpami założenia twierdzenia są spełnione, prawdziwa musi więc też być teza. Ilość ryb jest tak naprawdę ustalona, choć nieznaną, jest to więc niezupełnie „prawdziwa” zmienna losowa (takie „patologiczne” zmienne losowe nazywamy *deterministycznymi*). My jednakże wcale nie badamy bezpośrednio ilości ryb, lecz pewne zmienne losowe będące *estymatorami* tej wielkości – i do nich stosujemy centralne twierdzenie graniczne.

Estymatory

Wynik pomiaru, jak stwierdziliśmy wcześniej, jest zmienną losową. Możemy tę zmienną nazwać *estymatorem* badanej wielkości fizycznej. Inna metoda pomiaru tej samej wielkości fizycznej albo inna procedura przetwarzania surowych wyników to jednocześnie inny estymator. Chciałoby się, żeby estymator używany przez nas był możliwie najlepszy, cokolwiek to znaczy. Tymczasem znane są trzy cechy, które charakteryzują jakość estymatorów. Dobry estymator powinien być *zgodny*, *nieobciążony* i *efektywny*.

Omówimy te cechy kolejno:

- estymator jest *zgodny*, gdy w miarę zwiększania liczby pomiarów jego wartość średnia coraz lepiej przybliża (w sensie prawdopodobieństwa) prawdziwą wartość. Średnia arytmetyczna jest (zazwyczaj) takim estymatorem.
- estymator jest *nieobciążony*, gdy szacuje badaną wielkość raz z dołu, raz z góry – w sposób losowy. Możemy też powiedzieć, że estymator nieobciążony

nie wprowadza niepewności (dawniej: *błędów*) systematycznych.

- estymator jest *efektywny*, kiedy jego wariancja jest niewielka. Estymator o najmniejszej możliwej wariancji nazywa się *najefektywniejszy*.

Okazuje się, że znalezienie najefektywniejszego estymatora, który jednocześnie byłby zgodny i nieobciążony, nie zawsze jest możliwe, można nawet powiedzieć, że zdarza się raczej wyjątkowo.

Wykonując dowolne pomiary fizyczne, jesteśmy skazani na posługiwanie się różnymi estymatorami. Nazywając wartość średnią estymatora wynikiem pomiaru, i uznając jednocześnie, że jego wariancja jest jakąś miarą rozrzutu (albo jak kto woli: skupienia) indywidualnych pomiarów wokół tejże wartości średniej, podajemy końcowy rezultat pomiarów w formie (dyspersja σ , zwana również *odchyleniem standardowym*, jest pierwiastkiem z wariancji, σ^2):

$$\text{wynik} \pm \text{dyspersja}.$$

Przypomina to do złudzenia formę rekomendowaną przez 7 międzynarodowych organizacji, w tym ISO:

$$\text{wynik} \pm \text{niepewność pomiarowa}.$$

Zbieżność ta nie jest przypadkowa, rzeczywistość przez niepewność pomiarową, jeśli nie zaznaczono inaczej, rozumie się właśnie odchylenie standardowe.

Pora na komentarze

Widząc (cudzy) wynik w podanej wyżej postaci, należy go rozumieć w sensie probabilistycznym, czyli mniej więcej tak: wyniki (dostatecznie wielu) pomiarów gromadzą się przede wszystkim wokół liczby podanej jako „wynik”, a ich rozrzut scharakteryzowany jest przez „dyspersję”. W żadnym wypadku nie oznacza to, że *każdy* indywidualny wynik pomiaru mieścił się w przedziale

$$[\text{wynik} - \text{dyspersja}, \text{wynik} + \text{dyspersja}]$$

Często spotyka się stwierdzenie, że w tym przedziale mieści się ok. 68% wszystkich pomiarów, co generalnie także nie jest prawdziwe. Osoba, która tak twierdzi, zakłada bowiem, że zmienna losowa opisująca wyniki pomiarów ma rozkład normalny, co najczęściej mija się z prawdą. Wracając do naszego przykładu ze zliczaniem karp: jeśli przypuścimy, że ilość karp w stawie podlega rozkładowi normalnemu, to tym samym godzimy się na następujące nonsensy:

- prawdopodobieństwo tego, że ilość karp mieści się w przedziale $[1470,2, 1470,7]$ jest różne od zera – a powinno być dokładnie równe zero,
- szansa, że w stawie pływa ujemna ilość karp, także jest różna od zera,
- najśmieszniejsze jest to, iż prawdopodobieństwo tego, że w stawie znajduje się jakakolwiek całkowita ilość karp, jest zerowe!

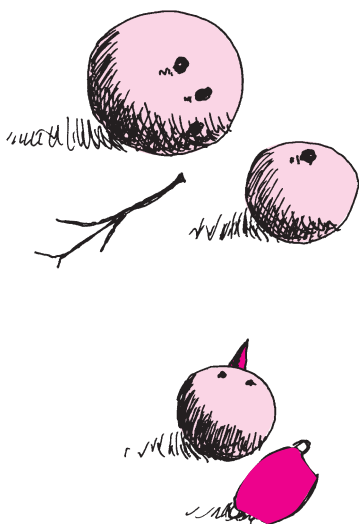
Widzimy, że model probabilistyczny wyników pomiarów jest daleki od ideału. Wartość liczbową, uznana przez nas za „wynik”, *nie* jest prawdziwą wartością badanej wielkości fizycznej, lecz jedynie jej oszacowaniem (estymatą), a konkretnie wartością średnią ze (skończonej) serii wykonanych pomiarów. Podobnie „dyspersja” jest dyspersją wyników w, jak mówimy, pobranej próbie, a *nie* dyspersją badanej wielkości fizycznej.

Używając modelu probabilistycznego potrafimy odpowiadać na pytania w rodzaju: *jaka jest szansa, że badana wielkość naprawdę mieści się w przedziale $[x_1, x_2]$* ? Powiedzmy, że ta szansa jest równa akurat 0,986. Czy taka odpowiedź rzeczywiście nas zadowala? Czy potrafimy w sposób obiektywny powiedzieć, czy ta szansa jest duża czy mała? Pomyślmy tylko, jaka w XIII wieku była szansa spotkania człowieka, który zgodziłby się z twierdzeniem, że Ziemia jest płaska? Z pewnością większa niż 0,986, a jednak my dziś wiemy *na pewno*, że Ziemia *nie* jest płaska.

Fizyk powinien dobrze wiedzieć, co mierzy i jaka procedura pomiarowa (i opracowania danych) jest odpowiednia do postawionego zadania. Jeśli chce zastosować standardową procedurę uśredniania wyników, to najpierw powinien się upewnić, czy ma to sens, to znaczy czy badana zmienna losowa spełnia założenia centralnego twierdzenia granicznego. Nawet jeśli tak jest, to mimo wszystko można się natknąć na nieoczekiwane komplikacje. Przypuśćmy, że w próbce piasku należy określić średnią wielkość ziaren, ale przeznaczona do badania próbka została przygotowana w sposób złośliwy, przez staranne wymieszanie dwóch wyraźnie różnych rodzajów piasku, w nierównych ilościach. Jeśli nie weźmiemy tego faktu pod uwagę, a jedynie mechanicznie wykonamy mrówczą pracę, to otrzymamy wynik o niewytłumaczalnie wielkiej dyspersji, np. pięciokrotnie większej niż wartość średnia.

Czy to zawsze działa?

Najgorsze zostawiliśmy na koniec. Niestety, zdarzają się przypadki, kiedy procedura uśredniania wielu pomiarów najzwyczajniej zawodzi. Gdyby ktoś np. chciał badać średnią długość trasy pokonywanej przez albatrosa pomiędzy lądowaniami, spotkałby się z czymś niezwykłym. Zwykle albatros żeruje w pewnej okolicy, ale od czasu do czasu wykonuje bardzo długie przeloty do odległych wysp, przynajmniej, że niekoniecznie dobrowolnie. Śledząc średnią długość trasy, wraz ze wzrostem numeru kolejnego przelotu zaobserwowalibyśmy w miarę przyzwoite „ustalanie się” jej wartości średniej. Jednorazowy długi przelot potrafiłby jednak wywrócić całą tę statystykę do góry nogami. Oczywiście, czas życia konkretnego albatrosa jest ograniczony, i chociażby z tego powodu wykona on w swym życiu jedynie skończoną liczbę lotów. Nie ma zatem problemu z wyliczeniem wartości średniej i wariancji, obie muszą okazać się skończone. Jeśli jednak zamiast albatrosów zaczniemy obserwować trwałe obiekty z mikroświata, w którym obowiązują prawa mechaniki kwantowej, to będziemy mieli problem. Istnieją takie wielkości fizyczne, dla których wartości średnie albo wariancje zwyczajnie nie istnieją. Spotykamy je nie tylko wśród zjawisk kwantowych, ale i klasycznych (np. turbulencje, przemiany fazowe), a także w innych, zgoła nieoczekiwanych miejscach: w internecie czy na giełdzie.



Rozwiązanie zadania F 585.

Dla uproszczenia rozważmy naczynie cylindryczne o przekroju S , przedzielone ciekłą ścianką, po której jednej stronie mamy $p = 0$, a po drugiej ciśnienie atmosferyczne $p = p_0$. Praca wykonana przez ciśnienie atmosferyczne to $W = p_0 S \Delta x$, a energia kinetyczna wdzierającej się masy powietrza wynosi

$$E_k = \frac{mv^2}{2} = \frac{\rho_0 S \Delta x v^2}{2}.$$

Z zasady zachowania energii $W = E_k$ otrzymujemy

$$v = \sqrt{\frac{2p_0}{\rho_0}} \approx 400 \frac{\text{m}}{\text{s}}.$$

