Czy losowa liczba jest duża czy mała?

Rafał SZTENCEL *

Oto problem z poprzedniego odcinka.

Wylosowano niezależnie i zgodnie z tym samym, nieznanym nam rozkładem o ciągłej dystrybuancie (albo wręcz ciągłym, jeśli komuś wygodniej), dwie liczby rzeczywiste. Pokazano nam jedną z nich. Jeśli prawidłowo odgadniemy, czy jest ona większa od drugiej, wygramy milion dolarów. Nietrudno zobaczyć, że możemy zapewnić sobie 50% szans na wygraną. A czy da się więcej?

Przy tak dużej nagrodzie warto walczyć o każdy dodatkowy procent szans.

Zadanie jest trudne. Piszący te słowa nie wie, czy poradziłby sobie z nim, jako że poznał je wraz z rozwiązaniem. Sondaż przeprowadzony wśród znajomych wykazał, że tylko nieliczni potrafili podać pomysł, prowadzący – po wskazówkach i dyskusji – do celu.

Jeśli chcemy rozstrzygnąć, czy zademonstrowana liczba jest duża, czy mała, potrzebujemy punktu odniesienia. Możemy, na przykład, porównać ją z zerem (właśnie taki pomysł miał jeden z ankietowanych).

Niech zatem X_1, X_2 będą wylosowanymi liczbami, pokazano nam X_1. Jeśli $X_1 < 0$, to uznajemy, że X_1 jest mniejsza od X_2; w przeciwnym razie uznajemy X_1 za większą od X_2. Ponieważ (wspólny) rozkład X_1 i X_2 ma ciągłą dystrybuantę, możemy ignorować zdarzenia o zerowym prawdopodobieństwie: $\{X_1 = 0\}$, $\{X_1 = X_2\}$ itp. Jest teraz sześć możliwych uporządkowań liczb X_1, X_2 i 0, które przedstawimy w tabeli. Plus w ostatniej kolumnie oznacza wygraną, minus przegraną.

1	$X_1 < X_2 < 0$	+
2	$0 < X_1 < X_2$	−
3	$X_2 < 0 < X_1$	+
4	$0 < X_2 < X_1$	+
5	$X_2 < X_1 < 0$	−
6	$X_1 < 0 < X_2$	+

Uporządkowania 1 i 5, 2 i 4, oraz 3 i 6 mają równe prawdopodobieństwa, skąd wynika, że szansa wygranej minus szansa przegranej jest równa

$$P(0 \in (\min(X_1, X_2), \max(X_1, X_2))).$$

Dla pewnych dystrybuant F powyższe prawdopodobieństwo będzie dodatnie, jeśli jednak zero znajdzie się poza zbiorem wartości X_1 i X_2, nie zyskamy nic.

Jak udoskonalić ten sposób? Należy zamiast zera użyć zmiennej losowej Z o ściśle rosnącej dystrybuancie. Wtedy

$$P(Z \in (\min(X_1, X_2), \max(X_1, X_2))) > 0.$$

Wykonanie takiego prywatnego losowania daje szansę wygranej większą niż $\frac{1}{2}$. Jeśli Z ma ten sam rozkład, co X_1 i X_2, szansa wygranej wzrasta do $\frac{2}{3}$, bowiem wtedy wszystkie uporządkowania w tabeli mają jednakowe prawdopodobieństwa.

Zadanie powyższe aż się prosi o rozmaite modyfikacje i uogólnienia. Można, na przykład, zapytać, co się zmieni, gdy rozkład X_1 (i X_2) jest znany. Zastanówmy się, który z nielosowych punktów odniesienia jest najlepszy, czyli dla jakiego a

$$(*)\quad P(a \in (\min(X_1, X_2), \max(X_1, X_2))) = P(X_1 < a < X_2) + P(X_2 < a < X_1)$$

jest największe. Za pomocą F (wspólnej dystrybuanty X_1 i X_2) łatwo wyliczyć prawdopodobieństwa dwóch zdarzeń po prawej stronie $(*)$:

$$\begin{aligned} P(X_1 < a < X_2) = P(X_2 < a < X_1) = P(X_2 < a, X_1 > a) = \\ = P(X_2 < a) \cdot P(X_1 > a) = F(a) \cdot [1 - F(a)], \end{aligned}$$

a to wyrażenie przyjmuje największą wartość, gdy $F(a) = \frac{1}{2}$. Istnienie takiego a wynika z ciągłości F. Liczba a (niekoniecznie wyznaczona jednoznacznie) jest *medianą* badanego rozkładu. Szansa wygranej wzrasta teraz do $\frac{1}{8} + \frac{1}{4} + \frac{1}{8} + \frac{1}{4} = \frac{3}{4}$.

Mediana jest jednym z ważnych parametrów rozkładu prawdopodobieństwa (prawie tak ważnym, jak wartość oczekiwana, która na ogół niewiele się od mediany różni). Medianą zmiennej losowej X nazywamy każdą liczbę a, dla której

$$P(X \leqslant a) \geqslant \frac{1}{2}, \quad P(X \geqslant a) \geqslant \frac{1}{2}.$$

Jak widzieliśmy w $(*)$, mediana (tak, jak inne ważne parametry rozkładów) ma pewną własność ekstremalną.

*Instytut Matematyki Uniwersytetu Warszawskiego