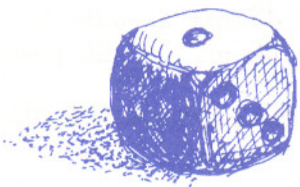
Probabilistyka w życiu codziennym

Bolesław
KOPOCIŃSKI

Przypadek pojawia się na każdym poziomie ogólności naszych rozważań; odgrywa rolę w fizycznym opisie świata, biologicznym opisie życia, zachowaniu mas w życiu społecznym itd. Jednostka postrzega przypadek jako zagrożenie lub wykorzystuje go jako element strategii w starciu z konkurentem. Probabilistykę spotykamy więc w codzienności. Oto przykłady.

Użyteczność wygranych na loterii. Loterią są m.in. ubezpieczenia majątkowe. Ubezpieczyciel szacuje prawdopodobieństwo np. pożaru, pobiera od klienta składkę równą np. 5% wartości obiektu i, w razie straty, wypłaca odszkodowanie. Klient ocenia wartość proponowanej loterii. Przy wspomnianej stawce może oczekiwać, że we wsi złożonej z 20 domów corocznie jeden dom spłonie lub (w bardziej efektownym wariancie) może liczyć na kompletną zagładę wsi średnio co 20 lat. Obserwacje rzeczywistości pozwalają ocenić, czy gra w ubezpieczenia jest sprawiedliwa.



Dla zabawy, aby przeżyć namiastkę gry, zaproponujmy znajomym wpłatę po złotówce do puli; niech pewien mechanizm losowy z równymi szansami przyzna całą pulę jednemu z nas. Większość osób, jakkolwiek nie widzi w tej grze większego sensu, przystaje na udział w grze. Zauważmy, że mało osób zagra, gdy stawkę podniesiemy do 100 zł.

Dlaczego ludzie biorą udział w grach niesprawiedliwych, a nie biorą udziału w sprawiedliwych? Odpowiedzi na to pytanie udziela teoria użyteczności. Każda wygrana i każda przegrana ma bowiem dla grających odpowiednią użyteczność.

Formalnie biorąc, do oceny gry losowej potrzebne jest pojęcie oczekiwanej wygranej. Niech $x_1, x_2, \ldots, x_n$ będą wypłatami w grze, a $p_1, p_2, \ldots, p_n$ – prawdopodobieństwami tych wypłat. Oczekiwaną wygraną definiuje się wzorem $\mathrm{E}(X) = \sum\limits_{i=1}^{n} x_i p_i$. Przypisując wygranej x użyteczność $u(x)$, bierzemy pod uwagę nową grę, w której z prawdopodobieństwem p_i zdarza się wypłata $u(x_i)$. Oczekiwaną użytecznością jest więc $\mathrm{E}(u(X)) = \sum\limits_{i=1}^{n} u(x_i)p_i$.

Więcej informacji na temat funkcji użyteczności znajdzie Czytelnik np. w książce: C.H. Coombs, R.M. Dawes, A. Tversky, *Wprowadzenie do psychologii matematycznej*, PWE, Warszawa 1977.

O kształt funkcji użyteczności toczą spory psychologowie i matematycy. Tu dla przykładu weźmiemy funkcję sklejoną z dwóch gałęzi parabol, określoną wzorami $u(x) = \sqrt{x/10}$ dla $x \geq 0$ i $u(x) = -(x/10)^2$ dla $x < 0$, która bagatelizuje pewne straty (np. te z przedziału $(-10, 0)$), a przecenia pewne zyski (np. te z przedziału $(0, 10)$). Jeśli $x_1 = -5$, $x_2 = 5$, $p_1 = p_2 = 0{,}5$, to $\mathrm{E}(X) = 0$, czyli gra w sensie oczekiwanej wygranej jest sprawiedliwa, emocjonalnie obojętna. Natomiast $\mathrm{E}(u(X)) = 0{,}5(-0{,}25 + \sqrt{0{,}5}) = 0{,}23$, czyli w sensie oczekiwanej użyteczności gra jest atrakcyjna, i to dla obu graczy!

Inna znana loteria to gra w ruletkę; w końcu XIX wieku w Monte Carlo przyjezdnych witał tłum tzw. Profesorów – ludzi, którzy sami zbankrutowali i usiłowali innych uczyć swych niezawodnych systemów gry.

Podczas długiej gry grający cząstka po cząstce traci kapitał (patrz np. *Delta* 6/1998); zmieniają się zarówno kwoty, którymi dysponuje, jak i użyteczność stawek. Jeśli się już gra, warto stawiać wszystko w jednej grze, która daje wygraną o pożądanej użyteczności.

Wracając do ubezpieczeń, przypuśćmy, że dom wartości 30 ($\times 10^4$ zł) jest zagrożony pożarem z prawdopodobieństwem 0,05. Strata X przyjmuje wartości $x_1 = -30$, $x_2 = 0$, z prawdopodobieństwami odpowiednio $p_1 = 0{,}05$, $p_2 = 0{,}95$. Zatem $\mathrm{E}(X) = -1{,}5$, a więc składka 1,5 byłaby sprawiedliwa. Oczekiwana użyteczność rezygnacji z ubezpieczania domu jest równa $\mathrm{E}(u(X)) = -(30/10)^2 \cdot 0{,}05 = -0{,}45$. Jeśli nawet ubezpieczymy się z niesprawiedliwą składką 2, to i tak oczekiwana użyteczność $\mathrm{E}(u(X)) = -0{,}04$ jest większa, niż gdy rezygnujemy z ubezpieczenia.

Metoda reprezentacyjna. Kiedy nadchodzą wybory, wiele osób chce zawczasu wiedzieć, jakie będą wyniki. Informacji dostarczają im biura badania opinii publicznej, ogłaszając np.: „kandydata A popiera 40% wyborców, z dokładnością 3%; reprezentatywna próba składała się z 1000 ankietowanych”.

Teoria matematyczna leżąca u podstaw takich stwierdzeń opiera się na rozkładzie dwumianowym i jego aproksymacji rozkładem normalnym. Zakłada się, że wyborcy mają zdeterminowane poglądy: w części p popierają A i w części $1 - p$ są mu przeciwni. Spośród wszystkich wyborców wybiera się losowo osoby ankietowane, które ujawniają swe poglądy. Frakcja popierających A w próbie jest estymatorem p poparcia w całym społeczeństwie. Głównym problemem

Rozwiązanie zadania F 514.
Potencjały punktów B i B' są równe i odcinek BB' nie wpływa na opór figury. Jeżeli przewodnik BB' usuniemy, to otrzymaną figurę można rozpatrywać jako równoległe połączenie dwóch jednakowych obwodów elektrycznych: $ABCDEF$ i $AB'C'D'E'F$. Wyznaczamy opór gałęzi $ABCDEF$. Opór pręta jest równy $r = \rho\frac{l}{S}$, a więc opór każdej części pręta wynosi $\frac{1}{3}r$. Opór trójkąta BCD wynosi $\frac{2}{9}r$, stąd opór całej gałęzi $ABCDEF$ ma wartość

$$R_1 = 2\cdot\frac{2}{9}r + \frac{1}{3}r = \frac{7}{9}r\,.$$

Opór całej figury jest równy:

$$R = \frac{1}{2}R_1 = \frac{7}{18}r = \frac{7}{18}\rho\frac{l}{S}\,.$$

badań reprezentacyjnych (obok sposobu losowania próby) jest ustalenie wielkości próby wystarczającej do oceny preferencji wyborców z wymaganą dokładnością.

Niech S_n będzie liczbą popierających A w próbie o liczebności n; ułamek $S_n/n = p'$ wyraża, jaka część próby popiera A. Podana informacja o błędzie prognozy oznacza, iż $|p'-p| \le 0{,}03$. Należy podkreślić, że żadna liczebność próby nie gwarantuje spełnienia tej nierówności; trzeba bowiem dopuścić myśl, iż losowa próba może się składać z samych zwolenników A nawet wtedy, gdy stanowią oni drobną część wyborców. Decydujemy się zatem na pewien *poziom ufności* α, np. $\alpha = 0{,}95$ – tzn. chcemy dobrać n tak, by nierówność $|p'-p| \le 0{,}03$ zachodziła z prawdopodobieństwem równym co najmniej 0,95.

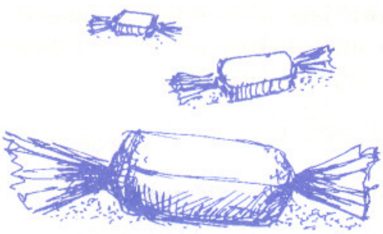
Używane metody losowania próby gwarantują zwykle, że S_n ma rozkład dwumianowy z parametrami n i p. Do rozwiązania naszego zadania wystarczy twierdzenie de Moivre'a–Laplace'a, które mówi, że przybliżeniem rozkładu dwumianowego jest rozkład normalny, a ściślej

$$P(|p'-p| \le 0{,}03) = P(|S_n/n - p| \le 0{,}03) = P(-0{,}03n \le S_n - np \le 0{,}03n) =$$
$$= P(-z \le Z \le z) \cong \phi(z) - \phi(-z),$$

gdzie $Z = (S_n - np)/\sqrt{npq}$, $z = 0{,}03n/\sqrt{npq}$, ϕ jest dystrybuantą standardowego rozkładu normalnego.

W odpowiednich tablicach można sprawdzić, że $\phi(z) - \phi(-z) = 0{,}95$ dla $z = 1{,}96$, czyli dla $n = 4268{,}4pq$. Ponieważ pq nie przekracza $\frac{1}{4}$, więc $n > 1067$ zapewnia wymagany poziom ufności. Pamiętajmy jednak, iż jest jedna szansa na 20, że błąd $|p'-p|$ przekroczy 0,03.

Celowe zaburzenia w badaniach reprezentacyjnych. Największym zagrożeniem metody reprezentacyjnej jest unikanie lub fałszowanie odpowiedzi. Zdarza się też często, że ankietowani – np. ze względu na wstyd lub inne uczucia – nie są skłonni ujawnić poglądów, ale chętnie się nimi dzielą, gdy mogą w jakiś sposób ukryć własne zdanie. Dla uwiarygodnienia ankiet dodaje się więc czasem jeszcze jeden element losowy. Oto przykład.



Przypuśćmy, że w ankiecie możliwa jest jedna z dwóch odpowiedzi: 0 lub 1. Ankietą objęto n osób; ich poglądy tworzą deterministyczny ciąg $x_1, x_2, \ldots, x_n$ zer i jedynek. Celem ankiety jest znalezienie frakcji $p = \frac{1}{n}\sum_{i=1}^{n} x_i$ osób wyznających pogląd 1.

Ankietowani odpowiadają zgodnie z następującą instrukcją: (1) rzuć monetą; (2) rzuć monetą po raz drugi; jeśli wypadnie orzeł, odnotuj w ankiecie własny pogląd, a jeśli reszka – wynik poprzedniego rzutu (0 oznacza reszkę, a 1 orła).

Zauważmy, że ankietowany może odpierać wszelkie uwagi, twierdząc, iż jego odpowiedź to wynik rzutu monetą. Nadto jednakowe odpowiedzi nie dowodzą jednomyślności w grupie.

Wprowadźmy zmienne losowe charakteryzujące i-tą osobę: δ_i – wynik pierwszego rzutu monetą (orzeł: $\delta_i = 1$; reszka: $\delta_i = 0$); Δ_i – wynik drugiego rzutu (orzeł: $\Delta_i = 1$; reszka: $\Delta_i = 0$). Wówczas odpowiedź i-tej osoby wyraża wzór $X_i = \Delta_i x_i + (1-\Delta_i)\delta_i$, a $S_n = \sum_{i=1}^{n} X_i$ to liczba odpowiedzi 1.

Wariancja zmiennej losowej X, określana wzorem

$$\mathrm{Var}(X) = \mathrm{E}((X - \mathrm{E}X)^2)\,,$$

jest miarą rozproszenia tej zmiennej.

Zmienne losowe $\delta_1, \delta_2, \ldots, \delta_n$, $\Delta_1, \Delta_2, \ldots, \Delta_n$ są niezależne i przyjmują z prawdopodobieństwem 1/2 wartości 0 i 1. Łatwo więc sprawdzić, że $\mathrm{E}(X_i) = \frac{1}{2}x_i + \frac{1}{4}$, $\mathrm{Var}(X_i) = \frac{3}{16}$. Proponowany estymator p^* frakcji p i jego wariancja (potrzebna do oceny błędu estymacji) są następujące:

$$p^* = \frac{2}{n}S_n - \frac{1}{2}, \qquad \mathrm{Var}(p^*) = \frac{3}{4n}.$$

Nasz estymator jest nieobciążony, tzn. jego wartość oczekiwana jest równa nieznanemu parametrowi: $\mathrm{E}(p^*) = p$. Jego wariancja maleje liniowo ze wzrostem n (tak samo jak w metodzie reprezentacyjnej bez zaburzania ankiet). Opisana metoda jest na tyle efektywna, że można ją stosować w praktyce, a grube oszacowania p nie wymagają wielu obserwacji.

Przy innej okazji napiszemy w *Delcie* o trafności losowych prognoz pogody.