C.4 Rozkłady: geometryczny i dwumianowy

Rzut monetą jest przykładem ***próby Bernoulliego***, która jest zdefiniowana jako doświadczenie mogące zakończyć się jednym z dwóch możliwych wyników: ***sukcesem***, który występuje z prawdopodobieństwem p, lub ***porażką***, która występuje z prawdopodobieństwem $q = 1 - p$. Kiedy mówimy o wielu ***próbach Bernoulliego***, mamy na myśli, że są one wzajemnie niezależne, i jeśli nie powiemy wyraźnie, że jest inaczej, to w każdej z nich jest takie samo prawdopodobieństwo sukcesu p. Z próbami Bernoulliego są związane dwa ważne rozkłady: rozkład geometryczny i dwumianowy.

Rozkład geometryczny

Przypuśćmy, że mamy ciąg prób Bernoulliego, każda z prawdopodobieństwem sukcesu p i prawdopodobieństwem porażki $q = 1 - p$. Ile nastąpi prób, zanim odniesiemy sukces? Niech zmienna losowa X będzie liczbą prób potrzebnych do osiągnięcia sukcesu. Wówczas X przyjmuje wartości z zakresu $\{1, 2, \ldots\}$ i dla $k \geq 1$ zachodzi równość

$$\Pr\{X = k\} = q^{k-1} p, \tag{C.35}$$

ponieważ mamy $k - 1$ porażek, zanim odniesiemy sukces. Rozkład prawdopodobieństwa opisany równaniem (C.35) jest nazywany ***rozkładem geometrycznym***. Rysunek C.1 ilustruje taki rozkład.

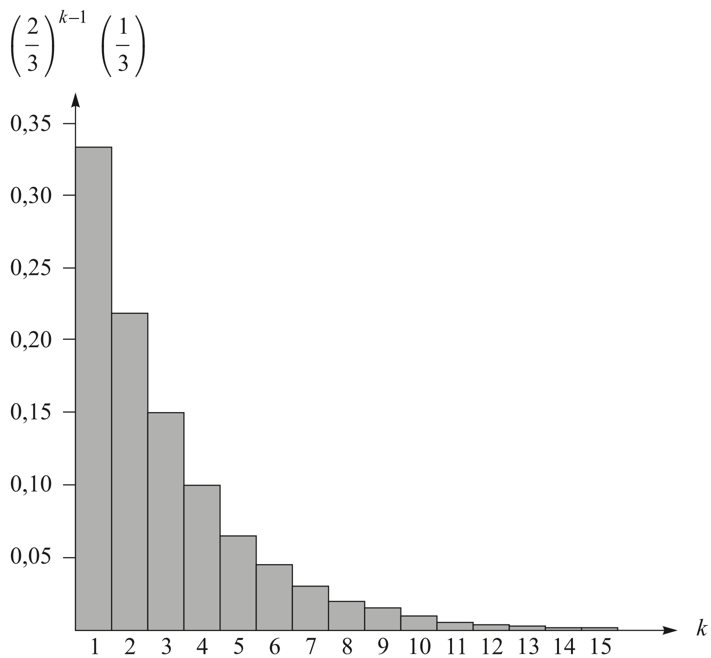
Przy założeniu, że $q < 1$, wartość oczekiwaną rozkładu geometrycznego można obliczyć następująco:

$$\begin{aligned} \mathrm{E}[X] &= \sum_{k=1}^{\infty} k q^{k-1} p \\ &= \frac{p}{q} \sum_{k=0}^{\infty} k q^k \\ &= \frac{p}{q} \cdot \frac{q}{(1-q)^2} \quad \text{(ze wzoru (A.11) na str. 1072)} \\ &= \frac{p}{q} \cdot \frac{q}{p^2} \\ &= 1/p. \end{aligned} \tag{C.36}$$

Potrzeba więc średnio $1/p$ prób, aby odnieść sukces, co jest zgodne z intuicją. Wariancja, którą można obliczyć w podobny sposób, korzystając z wyniku zadania C.4-4, wynosi

$$\mathrm{Var}[X] = q/p^2. \tag{C.37}$$

Przypuśćmy na przykład, że powtarzamy rzuty dwiema kostkami aż do otrzymania liczby oczek siedem lub jedenaście. Na 36 możliwych wyników 6 daje nam siódemkę i 2 – jedenastkę. Stąd prawdopodobieństwo sukcesu wynosi $p = 8/36 = 2/9$ i musimy średnio rzucić $1/p = 9/2 = 4{,}5$ razy, aby otrzymać w wyniku siedem lub jedenaście.



Rysunek C.1 Rozkład geometryczny z prawdopodobieństwem sukcesu $p = 1/3$ i prawdopodobieństwem porażki $q = 1 - p$. Wartość oczekiwana rozkładu wynosi $1/p = 3$

Rozkład dwumianowy

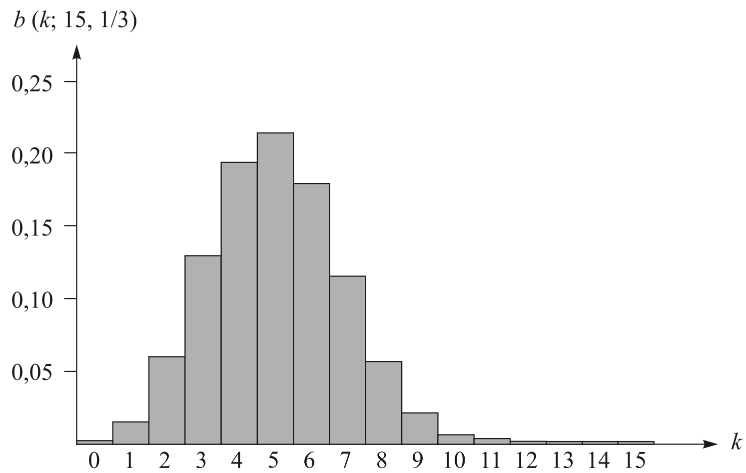
Ile sukcesów nastąpi podczas n prób Bernoulliego, jeżeli sukces następuje z prawdopodobieństwem p, a porażka z prawdopodobieństwem $q = 1 - p$? Zdefiniujmy zmienną losową X jako liczbę sukcesów, które wystąpiły w n próbach. Wówczas X przyjmuje wartości ze zbioru $\{0, 1, \ldots, n\}$ i dla $k = 0, \ldots, n$ zachodzi równość

$$\Pr\{X = k\} = \binom{n}{k} p^k q^{n-k}, \tag{C.38}$$

ponieważ istnieje $\binom{n}{k}$ sposobów na wybranie spośród n prób tych k, które kończą się sukcesem, a prawdopodobieństwo otrzymania właśnie takiego układu k sukcesów wynosi $p^k q^{n-k}$. Rozkład prawdopodobieństwa określony równaniem (C.38) jest nazywany ***rozkładem dwumianowym***. Dla wygody zdefiniujemy rodzinę rozkładów dwumianowych, używając zapisu

$$b(k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k}. \tag{C.39}$$

Rysunek C.2 ilustruje rozkład dwumianowy. Nazwa „dwumianowy” wynika z tego, że wzór (C.38) jest k-tym składnikiem rozwinięcia $(p + q)^n$. Zatem skoro $p + q = 1$, to ze wzoru (C.4) na str. 1109 dostajemy



Rysunek C.2 Rozkład dwumianowy $b(k; 15, 1/3)$ odpowiadający $n = 15$ próbom Bernoulliego, każda z prawdopodobieństwem sukcesu $p = 1/3$. Wartość oczekiwana rozkładu wynosi $np = 5$

$$\sum_{k=0}^{n} b(k; n, p) = 1 \tag{C.40}$$

zgodnie z wymogiem aksjomatu 2 prawdopodobieństwa.

Wartość oczekiwaną zmiennej losowej o rozkładzie dwumianowym możemy obliczyć, korzystając ze wzorów (C.9) i (C.40). Niech X będzie zmienną losową o rozkładzie $b(k; n, p)$ i niech $q = 1 - p$. Z definicji wartości oczekiwanej mamy

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{k=0}^{n} k \, \Pr\{X = k\} \\
&= \sum_{k=0}^{n} k \, b(k; n, p) \\
&= \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k} && \text{(z równania (C.9) na str. 1110)} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{(n-1)-k} \\
&= np \sum_{k=0}^{n-1} b(k; n-1, p) \\
&= np. && \text{(z równania (C.40)).}
\end{aligned} \tag{C.41}$$

Korzystając z liniowości wartości oczekiwanej, możemy otrzymać ten sam rezultat, wykonując znacznie mniej skomplikowanych obliczeń. Niech X_i będzie zmienną losową oznaczającą liczbę sukcesów w i-tej próbie. Wówczas $\mathrm{E}[X_i] = p \cdot 1 + q \cdot 0 = p$, a oczekiwana liczba sukcesów dla n prób wynosi

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} \mathrm{E}[X_i] \qquad \text{(z równania (C.24) na str. 1120)} \\
&= \sum_{i=1}^{n} p \\
&= np. && \text{(C.42)}
\end{aligned}$$

To samo podejście można zastosować do obliczenia wariancji naszego rozkładu. Korzystając z równania (C.31), otrzymujemy $\mathrm{Var}[X_i] = \mathrm{E}[X_i^2] - \mathrm{E}^2[X_i]$. Ponieważ X_i przyjmuje wyłącznie wartości 0 i 1, mamy $\mathrm{E}[X_i^2] = \mathrm{E}[X_i] = p$; zatem

$$\mathrm{Var}[X_i] = p - p^2 = p(1-p) = pq. \tag{C.43}$$

Aby obliczyć wariancję zmiennej losowej X, skorzystamy z niezależności n prób; stąd posługując się równaniem (C.33), mamy

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} \mathrm{Var}[X_i] \\
&= \sum_{i=1}^{n} pq \\
&= npq. && \text{(C.44)}
\end{aligned}$$

Jak można zobaczyć na rys. C.2, rozkład dwumianowy $b(k;n,p)$ rośnie w miarę jak k przebiega od 0 do np, a następnie maleje. Możemy udowodnić, że rozkład ten zawsze zachowuje się w taki sposób, patrząc na stosunek kolejnych prawdopodobieństw:

$$\begin{aligned}
\frac{b(k;n,p)}{b(k-1;n,p)} &= \frac{\binom{n}{k} p^k q^{n-k}}{\binom{n}{k-1} p^{k-1} q^{n-k+1}} \\
&= \frac{n!(k-1)!(n-k+1)!\,p}{k!(n-k)!n!q} \\
&= \frac{(n-k+1)p}{kq} && \text{(C.45)}
\end{aligned}$$

$$
\begin{aligned}
&= 1 + \frac{(n-k+1)p - kq}{kq}. \\
&= 1 + \frac{(n-k+1)p - k(1-p)}{kq}. \\
&= 1 + \frac{(n+1)p - k}{kq}.
\end{aligned}
$$

Stosunek ten jest większy niż 1, gdy $(n+1)p-k$ jest dodatnie. Zatem $b(k;n,p) > b(k-1;n,p)$ dla $k < (n+1)p$ (rozkład rośnie) i $b(k;n,p) < b(k-1;n,p)$ dla $k > (n+1)p$ (rozkład maleje). Jeżeli $(n+1)p$ jest całkowite, to dla $k = (n+1)p$ stosunek $b(k;n,p)/b(k-1;n,p)$ wynosi 1, czyli $b(k;n,p) = b(k-1;n,p)$. Rozkład ma wtedy dwa maksima: dla $k = (n+1)p$ i $k-1 = (n+1)p - 1 = np - q$. W przeciwnym wypadku osiąga on maksimum dla dokładnie jednego całkowitego k, które leży w przedziale $np - q < k < (n+1)p$.

Następujący lemat podaje górne ograniczenie rozkładu dwumianowego.

Lemat C.1
Niech $n \geq 0$, niech $0 < p < 1$, niech $q = 1 - p$ i niech $0 \leq k \leq n$. Wówczas

$$
b(k;n,p) \leq \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}.
$$

Dowód Mamy

$$
\begin{aligned}
b(k;n,p) &= \binom{n}{k} p^k q^{n-k} \\
&\leq \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k} p^k q^{n-k} \quad \text{(ze wzoru (C.7) na str. 1109)} \\
&= \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}.
\end{aligned}
$$

■

Zadania

C.4-1
Zweryfikuj aksjomat 2 z aksjomatów prawdopodobieństwa dla rozkładu geometrycznego.

C.4-2
Ile razy musimy średnio rzucić sześcioma monetami, zanim otrzymamy 3 orły i 3 reszki?

C.4-3
Pokaż, że wariancja rozkładu geometrycznego wynosi q/p^2. (*Wskazówka:* Wykorzystaj zadanie A.1-6 na str. 1074).

C.4-4
Pokaż, że $b(k;n,p) = b(n-k;n,q)$, gdzie $q = 1 - p$.

C.4-5

Pokaż, że wartość maksymalna rozkładu dwumianowego $b(k; n, p)$ wynosi w przybliżeniu $1/\sqrt{2\pi npq}$, gdzie $q = 1 - p$.

★ ***C.4-6***

Pokaż, że prawdopodobieństwo nieuzyskania żadnego sukcesu w n próbach Bernoulliego, każda z prawdopodobieństwem $p = 1/n$, wynosi w przybliżeniu $1/e$. Pokaż, że prawdopodobieństwo uzyskania dokładnie jednego sukcesu jest także w przybliżeniu równe $1/e$.

★ ***C.4-7***

Profesor Rosencrantz rzuca monetą n razy i to samo robi profesor Guildenstern. Pokaż, że prawdopodobieństwo, iż wyrzucą taką samą liczbę orłów, wynosi $\binom{2n}{n}/4^n$. (*Wskazówka:* Przyjmij, że dla profesora Rosencrantza sukcesem jest wyrzucenie orła, a dla profesora Guildensterna – wyrzucenie reszki). Opierając się na swoim dowodzie, wykaż, że

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}.$$

★ ***C.4-8***

Pokaż, że dla $0 \le k \le n$ zachodzi nierówność

$$b(k; n, 1/2) \le 2^{n H(k/n) - n},$$

gdzie $H(x)$ jest funkcją entropii zdefiniowaną równaniem (C.8) na str. 1110.

★ ***C.4-9***

Rozważmy n prób Bernoulliego, gdzie dla $i = 1, 2, \ldots, n$ prawdopodobieństwo sukcesu w i-tej próbie wynosi p_i, i niech X będzie zmienną losową oznaczającą łączną liczbę sukcesów. Niech $p \ge p_i$ dla każdego $i = 1, 2, \ldots, n$. Udowodnij, że dla $1 \le k \le n$ zachodzi nierówność

$$\Pr\{X < k\} \ge \sum_{i=0}^{k-1} b(i; n, p).$$

★ ***C.4-10***

Niech X będzie zmienną losową określającą łączną liczbę sukcesów w ciągu A n prób Bernoulliego, gdzie prawdopodobieństwo sukcesu w i-tej próbie wynosi p_i, i niech X' będzie zmienną losową określającą łączną liczbę sukcesów w ciągu A' n prób Bernoulliego, gdzie prawdopodobieństwo sukcesu w i-tej próbie wynosi $p'_i \ge p_i$. Udowodnij, że dla $0 \le k \le n$ zachodzi nierówność

$$\Pr\{X' \ge k\} \ge \Pr\{X \ge k\}.$$

(*Wskazówka:* Pokaż, jak uzyskać próby Bernoulliego w ciągu A' przez doświadczenie związane z próbami w ciągu A, i wykorzystaj wynik zadania C.3-7).

⋆ C.5 Krańce rozkładu dwumianowego

Prawdopodobieństwo osiągnięcia co najmniej lub co najwyżej k sukcesów w n próbach Bernoulliego, każda z prawdopodobieństwem sukcesu p, jest często bardziej interesujące niż prawdopodobieństwo osiągnięcia dokładnie k sukcesów. W tym podrozdziale badamy ***krańce*** rozkładu dwumianowego: dwa obszary rozkładu $b(k; n, p)$, które są dalekie od średniej np. Udowodnimy kilka ważnych oszacowań na sumy składników z obu krańców.

Na początku wyprowadzimy ograniczenie dotyczące prawego krańca rozkładu $b(k; n, p)$. Ograniczenia na lewym krańcu można ustalić przez zamianę ról sukcesu i porażki.

Twierdzenie C.2
Rozważmy ciąg n prób Bernoulliego, gdzie sukces występuje z prawdopodobieństwem p. Niech X będzie zmienną losową oznaczającą łączną liczbę sukcesów. Wówczas dla $0 \le k \le n$ prawdopodobieństwo uzyskania co najmniej k sukcesów wynosi

$$\begin{aligned}\Pr\{X \ge k\} &= \sum_{i=k}^{n} b(i; n, p) \\ &\le \binom{n}{k} p^k .\end{aligned}$$

Dowód Dla $S \subseteq \{1, 2, \ldots, n\}$ niech A_S oznacza zdarzenie, że i-ta próba jest sukcesem dla każdego $i \in S$. Ponieważ $\Pr\{A_S\} = p^k$, gdzie $|S| = k$, więc mamy

$$\begin{aligned}\Pr\{X \ge k\} &= \Pr\{\text{istnieje } S \subseteq \{1, 2, \ldots, n\} : \ |S| = k \text{ i } A_S\} \\ &= \Pr\left\{ \bigcup_{S \subseteq \{1,2,\ldots,n\} : |S| = k} A_S \right\} \\ &\le \sum_{S \subseteq \{1,2,\ldots,n\} : |S| = k} \Pr\{A_S\} \qquad \text{(z nierówności (C.21) na str. 1117)} \\ &= \binom{n}{k} p^k . \qquad \blacksquare\end{aligned}$$

Następujący wniosek jest odpowiednikiem powyższego twierdzenia dla lewego krańca rozkładu dwumianowego. W ogólności, adaptację twierdzeń zachodzących dla jednego z krańców na przypadek drugiego krańca będziemy pozostawiali Czytelnikowi.

Wniosek C.3
Rozważmy ciąg n prób Bernoulliego, gdzie sukces występuje z prawdopodobieństwem p. Jeśli X jest zmienną losową oznaczającą łączną liczbę sukcesów, to dla $0 \le k \le n$ prawdopodobieństwo uzyskania co najwyżej k sukcesów wynosi

$$\begin{aligned}
\Pr\{X \le k\} &= \sum_{i=0}^{k} b(i; n, p) \\
&\le \binom{n}{n-k}(1-p)^{n-k} \\
&= \binom{n}{k}(1-p)^{n-k} . \qquad \blacksquare
\end{aligned}$$

Nasze kolejne ograniczenie dotyczy lewego krańca rozkładu dwumianowego. Wniosek z niego mówi, że daleko od wartości średniej suma wartości na lewym krańcu zmniejsza się wykładniczo.

Twierdzenie C.4
Rozważmy ciąg n prób Bernoulliego, gdzie sukces występuje z prawdopodobieństwem p, a porażka z prawdopodobieństwem $q = 1 - p$. Niech X będzie zmienną losową oznaczającą łączną liczbę sukcesów. Wówczas dla $0 < k < np$ prawdopodobieństwo uzyskania mniej niż k sukcesów wynosi

$$\begin{aligned}
\Pr\{X < k\} &= \sum_{i=0}^{k-1} b(i; n, p) \\
&< \frac{kq}{np-k}\, b(k; n, p) .
\end{aligned}$$

Dowód Szacujemy szereg $\sum_{i=0}^{k-1} b(i; n, p)$ przez szereg geometryczny metodą z dodatku A.2, str. 1078. Dla $i = 1, 2, \ldots, k$ ze wzoru (C.45) mamy

$$\begin{aligned}
\frac{b(i-1; n, p)}{b(i; n, p)} &= \frac{iq}{(n-i+1)p} \\
&< \frac{iq}{(n-i)p} \\
&\le \frac{kq}{(n-k)p} .
\end{aligned}$$

Jeśli przyjmiemy

$$\begin{aligned}
x &= \frac{kq}{(n-k)p} \\
&< \frac{kq}{(n-np)p} \\
&= \frac{kq}{nqp} \\
&= \frac{k}{np} \\
&< 1 ,
\end{aligned}$$

to stąd wynika, że

$$b(i-1;n,p) < x\,b(i;n,p)$$

dla $0 < i \le k$. Powtarzając to rozumowanie $k-i$ razy, otrzymujemy

$$b(i;n,p) < x^{k-i}\,b(k;n,p)$$

dla $0 \le i < k$, a stąd

$$\begin{aligned}
\sum_{i=0}^{k-1} b(i;n,p) &< \sum_{i=0}^{k-1} x^{k-i} b(k;n,p) \\
&< b(k;n,p) \sum_{i=1}^{\infty} x^i \\
&= \frac{x}{1-x}\, b(k;n,p) \\
&= \frac{kq/((n-k)p)}{((n-k)p-kq)/((n-k)p)}\, b(k;n,p) \\
&= \frac{kq}{np-kp-kq}\, b(k;n,p) \\
&= \frac{kq}{np-k}\, b(k;n,p).
\end{aligned}$$

■

Wniosek C.5
Rozważmy ciąg n prób Bernoulliego, gdzie sukces występuje z prawdopodobieństwem p, a porażka z prawdopodobieństwem $q = 1 - p$. Dla $0 < k < np/2$ prawdopodobieństwo uzyskania mniej niż k sukcesów jest mniejsze niż połowa prawdopodobieństwa uzyskania mniej niż $k+1$ sukcesów.

Dowód Ponieważ $k \le np/2$, mamy

$$\begin{aligned}
\frac{kq}{np-k} &\le \frac{(np/2)q}{np-(np/2)} \\
&= \frac{(np/2)q}{np/2} \\
&\le 1,
\end{aligned} \tag{C.46}$$

bo $q \le 1$. Jeśli X jest zmienną losową oznaczającą liczbę sukcesów, to z twierdzenia C.4 i nierówności (C.46) wynika, że prawdopodobieństwo uzyskania mniej niż k sukcesów wynosi

$$\Pr\{X < k\} = \sum_{i=0}^{k-1} b(i;n,p) < b(k;n,p).$$

Mamy zatem

$$\begin{aligned}\frac{\Pr\{X < k\}}{\Pr\{X < k+1\}} &= \frac{\sum_{i=0}^{k-1} b(i;n,p)}{\sum_{i=0}^{k} b(i;n,p)} \\ &= \frac{\sum_{i=0}^{k-1} b(i;n,p)}{\sum_{i=0}^{k-1} b(i;n,p) + b(k;n,p)} \\ &< 1/2,\end{aligned}$$

ponieważ $\sum_{i=0}^{k-1} b(i;n,p) < b(k;n,p)$. ■

Oszacowania prawego krańca można wyznaczyć w podobny sposób. Ich dowody są pozostawione jako zadanie C.5-2.

Wniosek C.6
Rozważmy ciąg n prób Bernoulliego, gdzie sukces występuje z prawdopodobieństwem p. Niech X będzie zmienną losową oznaczającą łączną liczbę sukcesów. Wówczas dla $np < k < n$ prawdopodobieństwo uzyskania więcej niż k sukcesów wynosi

$$\begin{aligned}\Pr\{X > k\} &= \sum_{i=k+1}^{n} b(i;n,p) \\ &< \frac{(n-k)p}{k-np}\, b(k;n,p).\end{aligned}$$ ■

Wniosek C.7
Rozważmy ciąg n prób Bernoulliego, gdzie sukces występuje z prawdopodobieństwem p, a porażka z prawdopodobieństwem $q = 1 - p$. Dla $(np+n)/2 < k < n$ prawdopodobieństwo uzyskania więcej niż k sukcesów jest mniejsze niż połowa prawdopodobieństwa uzyskania więcej niż $k-1$ sukcesów. ■

W następnym twierdzeniu rozważanych jest n prób Bernoulliego, każda z prawdopodobieństwem p_i uzyskania sukcesu, dla $i = 1, 2, \ldots, n$. Jak wynika z zamieszczonego dalej wniosku, możemy użyć tego twierdzenia do wyznaczenia oszacowania prawego krańca rozkładu dwumianowego, przyjmując $p_i = p$ dla wszystkich prób.

Twierdzenie C.8
Rozważmy ciąg n prób Bernoulliego, gdzie w i-tej próbie, dla $i = 1, 2, \ldots, n$, sukces występuje z prawdopodobieństwem p_i, a porażka z prawdopodobieństwem $q_i = 1 - p_i$. Niech X będzie zmienną losową oznaczającą łączną liczbę sukcesów i niech $\mu = \mathrm{E}[X]$. Wówczas dla $r > \mu$ zachodzi zależność

$$\Pr\{X - \mu \ge r\} \le \left(\frac{\mu e}{r}\right)^r.$$

Dowód Ponieważ dla dowolnego $\alpha > 0$ funkcja $e^{\alpha x}$ jest ściśle rosnąca względem x, mamy

$$\Pr\{X - \mu \geq r\} = \Pr\left\{e^{\alpha(X-\mu)} \geq e^{\alpha r}\right\}, \tag{C.47}$$

gdzie α będzie ustalone później. Korzystając z nierówności Markowa (C.34), otrzymujemy

$$\Pr\left\{e^{\alpha(X-\mu)} \geq e^{\alpha r}\right\} \leq \mathrm{E}\left[e^{\alpha(X-\mu)}\right] e^{-\alpha r}. \tag{C.48}$$

Główny ciężar dowodu polega na oszacowaniu $\mathrm{E}\left[e^{\alpha(X-\mu)}\right]$ i podstawieniu odpowiedniej wartości α w nierówności (C.48). Najpierw wyznaczymy $\mathrm{E}\left[e^{\alpha(X-\mu)}\right]$. Stosując metodę wskaźnikowych zmiennych losowych z podrozdz. 5.2, przyjmijmy $X_i = \mathrm{I}\,\{i$-ta próba Bernoulliego kończy się sukcesem$\}$ dla $i = 1, 2, \ldots, n$, tzn. że X_i jest zmienną losową, która przyjmuje wartość 1, jeśli wynikiem i-tej próby Bernoulliego jest sukces, a 0 – jeśli porażka. Wtedy

$$X = \sum_{i=1}^{n} X_i,$$

a z liniowości wartości oczekiwanej mamy

$$\mu = \mathrm{E}[X] = \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} p_i,$$

skąd wynika, że

$$X - \mu = \sum_{i=1}^{n} (X_i - p_i).$$

Podstawiając prawą stronę tej równości za $X - \mu$, otrzymujemy

$$\begin{aligned}
\mathrm{E}\left[e^{\alpha(X-\mu)}\right] &= \mathrm{E}\left[e^{\alpha \sum_{i=1}^{n}(X_i - p_i)}\right] \\
&= \mathrm{E}\left[\prod_{i=1}^{n} e^{\alpha(X_i - p_i)}\right] \\
&= \prod_{i=1}^{n} \mathrm{E}\left[e^{\alpha(X_i - p_i)}\right],
\end{aligned}$$

co wynika z (C.27), ponieważ z wzajemnej niezależności zmiennych losowych X_i wynika wzajemna niezależność zmiennych losowych $e^{\alpha(X_i - p_i)}$ (patrz zad. C.3-5). Z definicji wartości oczekiwanej mamy

$$\begin{aligned}
\mathrm{E}\left[e^{\alpha(X_i - p_i)}\right] &= e^{\alpha(1-p_i)} p_i + e^{\alpha(0-p_i)} q_i \\
&= p_i e^{\alpha q_i} + q_i e^{-\alpha p_i} \\
&\leq p_i e^{\alpha} + 1 \\
&\leq \exp(p_i e^{\alpha}),
\end{aligned} \tag{C.49}$$

gdzie $\exp(x)$ oznacza funkcję wykładniczą: $\exp(x) = e^x$. (Nierówność (C.49) wynika z nierówności $\alpha > 0$, $q_i \le 1$, $e^{\alpha q_i} \le e^\alpha$ i $e^{-\alpha p_i} \le 1$, a ostatni wiersz wynika z nierówności (3.12)). Stąd otrzymujemy

$$\begin{aligned} \mathrm{E}\left[e^{\alpha(X-\mu)}\right] &= \prod_{i=1}^{n} \mathrm{E}\left[e^{\alpha(X_i-p_i)}\right] \\ &\le \prod_{i=1}^{n} \exp(p_i e^\alpha) \\ &= \exp\left(\sum_{i=1}^{n} p_i e^\alpha\right) \\ &= \exp(\mu e^\alpha), \end{aligned} \tag{C.50}$$

bo $\mu = \sum_{i=1}^{n} p_i$. Z równania (C.47) oraz nierówności (C.48) i (C.50) wynika zatem, że

$$\Pr\{X - \mu \ge r\} \le \exp(\mu e^\alpha - \alpha r). \tag{C.51}$$

Przyjmując $\alpha = \ln(r/\mu)$ (patrz zad. C.5-7), dostajemy

$$\begin{aligned} \Pr\{X - \mu \ge r\} &\le \exp(\mu e^{\ln(r/\mu)} - r\ln(r/\mu)) \\ &= \exp(r - r\ln(r/\mu)) \\ &= \frac{e^r}{(r/\mu)^r} \\ &= \left(\frac{\mu e}{r}\right)^r. \end{aligned}$$

■

Twierdzenie (C.8) zastosowane do prób Bernoulliego, gdzie każda z prób ma to samo prawdopodobieństwo zaistnienia sukcesu, daje nam następujący wniosek dotyczący ograniczenia prawego krańca rozkładu dwumianowego.

Wniosek C.9

Rozważmy ciąg n prób Bernoulliego, gdzie w każdej próbie sukces występuje z prawdopodobieństwem p, a porażka występuje z prawdopodobieństwem $q = 1 - p$. Niech X będzie zmienną losową oznaczającą łączną liczbę sukcesów. Wtedy dla $r > np$ zachodzi zależność

$$\begin{aligned} \Pr\{X - np \ge r\} &= \sum_{k=\lceil np+r \rceil}^{n} b(k; n, p) \\ &\le \left(\frac{npe}{r}\right)^r. \end{aligned}$$

Dowód Ze wzoru (C.41) mamy $\mu = \mathrm{E}[X] = np$. ■

Zadania

★ ***C.5-1***

Co jest bardziej prawdopodobne: uzyskanie dokładnie n orłów w $2n$ rzutach monetą symetryczną, czy uzyskanie n orłów w n rzutach monetą symetryczną?

★ ***C.5-2***

Udowodnij wnioski C.6 i C.7.

★ ***C.5-3***

Pokaż, że

$$\sum_{i=0}^{k-1} \binom{n}{i} a^i < (a+1)^n \frac{k}{na - k(a+1)} b(k; n, a/(a+1))$$

dla każdego $a > 0$ i każdego k takiego, że $0 < k < na/(a+1)$.

★ ***C.5-4***

Udowodnij, że jeśli $0 < k < np$, gdzie $0 < p < 1$ i $q = 1 - p$, to

$$\sum_{i=0}^{k-1} p^i q^{n-i} < \frac{kq}{np-k} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}.$$

★ ***C.5-5***

Pokaż, korzystając z twierdzenia C.8, że dla $r > n - \mu$ zachodzi nierówność

$$\Pr\{\mu - X \geq r\} \leq \left(\frac{(n-\mu)e}{r}\right)^r.$$

Podobnie, pokaż, korzystając z wniosku C.9, że dla $r > n - np$ zachodzi nierówność

$$\Pr\{np - X \geq r\} \leq \left(\frac{nqe}{r}\right)^r.$$

★ ***C.5-6***

Rozważmy ciąg n prób Bernoulliego, gdzie w i-tej próbie, dla $i = 1, 2, \ldots, n$, sukces występuje z prawdopodobieństwem p_i, a porażka z prawdopodobieństwem $q_i = 1 - p_i$. Niech X będzie zmienną losową oznaczającą łączną liczbę sukcesów i niech $\mu = \mathrm{E}[X]$. Pokaż, że dla $r \geq 0$

$$\Pr\{X - \mu \geq r\} \leq e^{-r^2/2n}.$$

(*Wskazówka:* Udowodnij, że $p_i e^{\alpha q_i} + q_i e^{-\alpha p_i} \leq e^{-\alpha^2/2}$. Następnie poprowadź dowód jak dowód twierdzenia C.8, używając tej nierówności w miejsce nierówności (C.49)).

★ ***C.5-7***

Pokaż, że prawa strona nierówności (C.51) jest najmniejsza dla $\alpha = \ln(r/\mu)$.

Problemy

C-1 Problem Monty'ego Halla

Wyobraź sobie, że jesteś uczestnikiem teleturnieju *Let's Make a Deal* z lat 60-tych, którego gospodarzem jest Monty Hall. Cenna nagroda jest ukryta za jednymi z trzech drzwi, a bezwartościowe nagrody za pozostałymi dwoma drzwiami. Wygrasz cenną nagrodę – samochód lub inną drogą rzecz – jeśli wybierzesz właściwe drzwi. Po wybraniu jednych drzwi, ale jeszcze przed ich otwarciem, Monty, który wie, gdzie jest samochód, poleca swojej asystentce Carol Merrill, aby otworzyła jedne z pozostałych drzwi, odsłaniając kozę (nie jest to cenna nagroda). Gospodarz pyta, czy chcesz pozostać przy obecnym wyborze, czy też przerzucić się na drugie zamknięte drzwi. Co powinieneś zrobić, aby zmaksymalizowac swoje szanse na wygranie samochodu, a nie drugiej kozy?

Odpowiedź na to pytanie – pozostać przy pierwotnym wyborze, czy zmienić decyzję? – była przedmiotem wielu dyskusji, po części dlatego, że opis problemu nie jest całkowicie jednoznaczny. Zbadamy konsekwencje rozmaitych subtelnych założeń.

(a) Załóżmy, że twój pierwszy wybór jest losowy, czyli że prawdopodobieństwo wyboru właściwych drzwi wynosi $1/3$. Wiesz przy tym, że Monty zawsze daje każdemu uczestnikowi możliwość zmiany decyzji (a więc da ją też i tobie). Udowodnij, że lepiej jest zmienić decyzję niż pozostać przy pierwotnym wyborze. Jakie jest prawdopodobieństwo wygrania samochodu?

To właśnie ta odpowiedź jest najczęściej podawana jako rozwiązanie, mimo że w oryginalnym sformułowaniu problemu nie zakłada się, że Monty *zawsze* oferuje uczestnikowi możliwość zmiany decyzji. Jak wyjaśnimy dalej, twoja optymalna strategia może być inna, jeśli to niejawne założenie nie jest spełnione. W rzeczywistości w prawdziwym teleturnieju po wyborze drzwi przez uczestnika Monty czasami po prostu polecał Carol otwarcie wskazanych drzwi.

Stwórzmy model interakcji między tobą a Montym w postaci doświadczenia probabilistycznego, w którym obie strony stosują zrandomizowane strategie. Dokładniej, po wybraniu przez ciebie drzwi Monty oferuje ci możliwość zmiany z prawdopodobieństwem p_{right}, jeśli wybrałeś właściwe drzwi i z prawdopodobieństwem p_{wrong}, jeśli twój wybór był zły. Jeśli dostałeś możliwość zmiany decyzji, to zmieniasz ją z prawdopodobieństwem p_{switch}. Jeśli na przykład Monty zawsze daje ci możliwość zmiany decyzji, to jego strategię określają wartości $p_{\text{right}} = p_{\text{wrong}} = 1$. Jeśli ty zawsze zmieniasz decyzję, to twoją strategię określa $p_{\text{switch}} = 1$.

Grę można teraz potraktować jako eksperyment składający się z pięciu kroków:

1. Wybierasz drzwi losowo, trafiając samochód (właściwy wybór) z prawdopodobieństwem $1/3$, albo kozę (niewłaściwy wybór) z prawdopodobieństwem $2/3$.
2. Carol otwiera jedne z dwóch pozostałych drzwi, odsłaniając kozę.
3. Monty oferuje ci możliwość zmiany decyzji z prawdopodobieństwem p_{right}, jeśli twój wybór był właściwy i z prawdopodobieństwem p_{wrong}, jeśli twój wybór był niewłaściwy.
4. Jeśli Monty zaoferował ci możliwość zmiany decyzji w kroku 3, to zmieniasz ją z prawdopodobieństwem p_{switch}.
5. Carol otwiera ostatecznie wybrane przez ciebie drzwi, odsłaniając albo samochód (wygrałeś), albo kozę (przegrałeś).

Przeanalizujemy teraz tę grę i wyjaśnimy, jak wartości p_{right}, p_{wrong} i p_{switch} wpływają na prawdopodobieństwo wygranej.

(b) Jakich jest sześć możliwych wyników w przestrzeni zdarzeń dla tej gry? Które wyniki odpowiadają wygraniu przez ciebie samochodu? Jak wyrażają się prawdopodobieństwa poszczególnych wyników w zależności od wartości p_{right}, p_{wrong} i p_{switch}? Przedstaw swoje odpowiedzi w formie tabeli.

(c) Skorzystaj z wyników ze swojej tabeli (lub użyj innego sposobu), żeby udowodnić, że prawdopodobieństwo wygrania samochodu wynosi

$$\frac{1}{3}(2p_{\text{wrong}}p_{\text{switch}} - p_{\text{right}}p_{\text{switch}} + 1).$$

Załóżmy, że Monty zna prawdopodobieństwo p_{switch}, z jakim zmieniasz decyzję, a jego celem jest minimalizacja twojej szansy na wygraną.

(d) Jaka jest najlepsza strategia Monty'ego, czyli najlepszy wybór wartości p_{right} i p_{wrong}, w przypadku, kiedy $p_{\text{switch}} > 0$ (to znaczy, zmieniasz decyzję z dodatnim prawdopodobieństwem)?

(e) Uzasadnij, że jeśli $p_{\text{switch}} = 0$ (nigdy nie zmieniasz decyzji), to każda z możliwych strategii Monty'ego jest dla niego optymalna.

Załóżmy teraz, że strategia Monty'ego jest ustalona, czyli wartości p_{right} i p_{wrong} są z góry określone.

(f) W przypadku, kiedy znasz wartości p_{right} i p_{wrong}, jaka jest twoja najlepsza strategia wyboru wartości prawdopodobieństwa p_{switch} w zależności od p_{right} i p_{wrong}?

(g) W przypadku, kiedy nie znasz wartości p_{right} i p_{wrong}, jaki wybór p_{switch} maksymalizuje najmniejsze prawdopodobieństwo wygranej dla wszystkich możliwych konfiguracji p_{right} i p_{wrong}?

Wróćmy do pierwotnego sformułowania problemu, w którym Monty dał ci możliwość zmiany decyzji, ale nie znasz jego zamiarów ani strategii.

(h) Uzasadnij, że proawdopodobieństwo warunkowe wygrania samochodu, jeśli wiadomo, że Monty daje ci możliwość zmiany decyzji, wynosi

$$\frac{p_{\text{right}} - p_{\text{right}}p_{\text{switch}} + 2p_{\text{wrong}}p_{\text{switch}}}{p_{\text{right}} + 2p_{\text{wrong}}}. \tag{C.52}$$

Wyjaśnij, dlaczego $p_{\text{right}} + 2p_{\text{wrong}} \neq 0$.

(i) Jaka jest wartość wyrażenia (C.52) dla $p_{\text{switch}} = 1/2$? Pokaż, że wybór $p_{\text{switch}} < 1/2$ lub $p_{\text{switch}} > 1/2$ pozwala Monty'emu dobrać wartości p_{right} i p_{wrong}, które dają mniejszą wartość wyrażenia (C.52) niż przy wyborze $p_{\text{switch}} = 1/2$.

(j) Załóżmy, że nie znasz strategii Monty'ego. Wyjaśnij, dlaczego zmiana decyzji z prawdopodobieństwem $1/2$ jest dobrą strategią postępowania w przypadku problemu w jego pierwotnym sformułowaniu. Podsumuj, czego się nauczyłeś, analizując ten problem.

C-2 Kule i urny

W tym problemie prześledzimy efekt różnych założeń na liczbę sposobów umieszczenia n kul w b różnych urnach.

(a) Przypuśćmy, że jest n różnych kul i że ich kolejność w urnie nie ma znaczenia. Pokaż, że liczba sposobów umieszczenia kul w urnach wynosi b^n.

(b) Przypuśćmy, że kule są różne i że kule w każdej z urn są uporządkowane. Udowodnij, że liczba sposobów umieszczenia kul w urnach wynosi $(b+n-1)!/(b-1)!$. (*Wskazówka:* Rozważ liczbę możliwości ustawienia n różnych kul i $b-1$ nierozróżnialnych patyków w jednym rzędzie).

(c) Przypuśćmy, że kule są identyczne, a więc ich kolejność w urnie nie ma znaczenia. Pokaż, że liczba sposobów umieszczenia kul w urnach wynosi $\binom{b+n-1}{n}$. (*Wskazówka:* Ile ustawień z punktu (b) powtarza się, jeśli przyjmiemy, że kule są identyczne?).

(d) Przypuśćmy, że kule są identyczne i że żadna z urn nie może zawierać więcej niż jedną kulę, czyli $n \leq b$. Pokaż, że liczba sposobów umieszczenia kul wynosi $\binom{b}{n}$.

(e) Przypuśćmy, że kule są identyczne i że żadna z urn nie może pozostać pusta. Przy założeniu $n \geq b$ pokaż, że liczba sposobów umieszczenia kul wynosi $\binom{n-1}{b-1}$.

Uwagi do dodatku

Pierwsze ogólne metody rozwiązywania problemów prawdopodobieństwa były rozważane w sławnej korespondencji pomiędzy B. Pascalem a P. de Fermatem, która rozpoczęła się w 1654 r., oraz w książce C. Huygensa z 1657 r. Początki ścisłej teorii prawdopodobieństwa datuje się od prac J. Bernoulliego z 1713 r. i A. De Moivre'a z 1730 r. Kolejne udoskonalenia teorii wprowadzili P.S. de Laplace, S.-D. Poisson oraz C.F. Gauss.

Sumy zmiennych losowych były pierwotnie przedmiotem badań P.L. Czebyszewa i A.A. Markowa. Teoria prawdopodobieństwa została zaksjomatyzowana przez A.N. Kołmogorowa w 1933 r. Oszacowania krańców rozkładu zostały podane przez Czernoffa [91] i Hoeffdinga [222]. Podstawowe zagadnienia dotyczące losowych struktur kombinatorycznych opisał P. Erdős.

Książki Knutha [259] i Liu [302] są dobrymi podręcznikami elementarnej kombinatoryki i zagadnień zliczania. Standardowe podręczniki Billingsleya [56], Chunga [93], Drake'a [125], Fellera [139] i Rozanova [390] dają wszechstronne wprowadzenie do probabilistyki.