Elementy rachunku prawdopodobienstwa

Antoni Kościelski

Tekst przedtawionych elementów nie został starannie przeczytany przed ogłoszeniem i może zawierać wiele niezauważonych błędów.

Spis treści

1 Przestrzenie probabilistyczne

1.1 Zdarzenia elementarne

Zbiór zdarzeń elementarnych to dowolny niepusty zbiór. Elementy tego zbioru nazywamy zdarzeniami elementarnymi. Można myśleć, że jest to zbiór wszystkich możliwych wyników pewnego doświadczenia losowego.

1.2 Rozkład prawdopodobieństwa dla skończonych zbiorów zdarzeń elementarnych

Przypuśćmy, że Ω jest skończonym zbiorem zdarzeń elementarnych. Rozkładem prawdopodobieństwa nazywamy dowolną funkcję $p : \Omega \to [0,1]$ taką, że

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

Wartość $p(\omega)$ jest nazywana prawdopodobieństwem wystąpienia zdarzenia (elementarnego) ω i jest interpretowana jako szansa wystąpienia tego zdarzenia lub częstość występowania tego zdarzenia.

1.3 Zdarzenia

W doświadczeniu losowym (a więc na poziomie intuicyjnym), zdarzeniem nazywamy zajście jednego z wielu zdarzeń elementarnych. Na przykład zdarzenie polegające na tym, że przy rzucie kostką wypadnie parzysta liczba oczek składa się z trzech zdarzeń elementarnych polegających na wyrzuceniu 2, 4 lub 6 oczek. Zdarzenie może więc być rozumiane jako alternatywa zdarzeń elemenarnych i może być utożsamiane ze zbiorem zdarzeń elementarnych.

Przyjmujemy, że zdarzenie to zbiór zdarzeń elementarnych. Jeżeli zbiór wszystkich zdarzeń elementarnych jest skończony, to dowolny zbiór zdarzeń elementarnych jest zdarzeniem. Dla nieskończonych zbiorów zdarzeń elementarnych definicja będzie bardziej restrykcyjna.

Taka definicja powoduje pewne komplikacje. Nie widać istotnych powodów, aby rozróżniać zdarzenie elementarne ω i zdarzenie $\{\omega\}$. Będziemy zwykle posługiwać się zdarzeniami, a zdarzenie elementarne będziemy utożsamiać ze zdarzeniem jednoelementowym.

1.4 Ciało zbiorów

Niepustą rodzinę (zbiór) $\mathcal{C}$ (Ω) nazywamy ciałem zbiorów, jeżeli $\mathcal{C}$ jest zamknięta ze względu na sumę mnogościową i dopełnienie (do zbioru Ω).

Jest oczywiste, że ciało zbiorów jest zamknięte ze względu na wszystkie skończone działania mnogościowe (iloczyn kartezjański nie jest uważany za działanie mnogościowe), należy do niego zbiór Ω i zbiór pusty. Zdarzenia tworzą ciało zbiorów.

Przykładem ciała jest zbiór $\mathcal{P}(\Omega)$ wszystkich podzbiorów Ω. Ciałem jest też zbiór

$$\{X \subseteq N : X \text{ jest skończony lub ma skończone dopełnienie}\}.$$

W rachunku prawdopodobieństwa często stosuje się zamiast mnogościowej inną, charakterystyczną terminologię. Zbiór wszsytkich zdarzeń elementarnych Ω nazywa się zdarzeniem pewnym, zbiór pusty – zdarzeniem niemożliwym, dopełnienie zdarzenia nazywa się zdarzeniem przeciwnym. Elementy zdarzenia A (a więc zdarzenia elementarne) nazywa się zdarzeniami sprzyjającymi zajściu zdarzenia A. Mówi się też o sumie zdarzeń, czyli o sumie mnogościowej, oraz o iloczynie zdarzeń, czyli o przekroju.

1.5 Prawdopodobieństwo zdarzenia

Prawdopodobieństwo zdarzenia A ma odpowiadać intuicyjnemu pojęciu szansy zajścia zdarzenia A i jest rozumiane jako suma szans zajścia zdarzeń elementarnych składających się na A. Zwykle uważa się, że przy rzucie kostką szansa wypadnięcia parzystej liczby oczek jest sumą szans wypadnięcia 2, 4 i 6 oczek.

Jeżeli mamy dany rozkład prawdopodobieństwa p, to prawdopodobieństwo zdarzenia A definiuje się jako liczbę $P(A)$ daną wzorem

$$P(A) = \sum_{\omega \in A} p(\omega).$$

Nietrudno zauważyć, że dla każdej pary rozłącznych zbiorów $A, B \in \mathcal{P}(\Omega)$, zachodzi równość

$$P(A \cup B) = P(A) + P(B).$$

Mamy też $P(\Omega) = 1$.

1.6 Skończone przestrzenie probabilistyczne

Przestrzenią probabilistyczną (albo losową) nazywamy układ złożony ze zbioru zdarzeń elementarnych Ω, zbioru (wszystkich) zdarzeń i prawdopodobieństwa. Na razie rozważamy przypadek, w którym zbiór Ω jest skończony. W tym przypadku, zbiorem wszystkich możliwych zdarzeń jest zwykle zbiór $\mathcal{P}(\Omega)$ wszystkich podzbiorów Ω, a prawdopodobieństwo jest funkcją $P : \mathcal{P}(\Omega) \to [0, 1]$ (określoną na zbiorze wszystkich zdarzeń) i spełniającą spełniające warunki:

1. $P(\Omega) = 1$ oraz
2. $P(A \cup B) = P(A) + P(B)$ dla każdej pary rozłącznych zdarzeń $A, B \in \mathcal{P}(\Omega)$.

Ta druga własność prawdopodobieństwa nazywa się addytywnością. Wartość $P(A)$ nazywamy prawdopodobieństwem (zajścia) zdarzenia A.

Widać już pewne trudności językowe. Będziemy zwykle mówić zdarzenie lub prawdopodobieństwo zdarzenia, mimo że pojęcia te zależą od przestrzeni probabilistycznej. Nie grozi to żadnymi konsekwencjami przynajmniej tak długo, dopóki analizujemy jedną, określoną przestrzeń probabilistyczną i bez trudu domyślamy się, jaką przestrzeń rozważamy w tym momencie.

1.7 Najprostsze własności prawdopodobieństwa

Z definicji przestrzeni probabilistycznej wynika, że prawdopodobieństwo ma następujące własności:

1. Jeżeli $A \subseteq B$, to $P(A) \leqslant P(B)$ (monotoniczność),
2. $P(A^c) = 1 - P(A)$ (A^c oznacza zdarzenie przeciwne do A, czyli dopełnienie A),
3. $P(\emptyset) = 0$ oraz $0 \leqslant P(A) \leqslant 1$ dla dowolnego zdarzenia A,
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ dla każdej pary zdarzeń A i B,
5. $P(A \cup B) \leqslant P(A) + P(B)$ dla każdej pary zdarzeń A i B (podaddytywność).

Z przyjętej definicji wynika także, że dla dowolnego zdarzenia A zachodzi wzór

$$P(A) = \sum_{\omega \in A} P(\{\omega\}).$$

W przypadku skończonych zbiorów zdarzeń elementarnych mamy wzajemnie jednoznaczną odpowiedniość między rozkładami prawdopodobieństwa i przestrzeniami probabilistycznymi. Mając przestrzeń probabilistyczną łatwo zdefiniować rozkład prawdopodobieństwa p odpowiadający tej przestrzeni. Możemy to zrobić przyjmując, że

$$p(\omega) = P(\{\omega\}).$$

Z drugiej strony, mając rozkład prawdopodobieśtwa p na skończonym zbiorze Ω, na zbiorze $\mathcal{P}(\Omega)$ możemy zdefiniować prawdopodobieństwo przyjmując, że

$$p(A) = \sum_{\omega \in A} p(\omega).$$

Sprawdzenie, że zachodzą wymagane własności prawdopodobieństwa, bądź rozkładu, a także że opisana odpowiedniość jest wzajemnie jednoznaczna, pozostawiam jako zadanie.

1.8 Sumy przeliczalnych zbiorów liczb

Chcemy posługiwać się symbolem

$$\sum_{\omega\in A} p(\omega)$$

także w przypadku, gdy A jest zbiorem nieskończonym i przeliczalnym. Są dwie metody definiowania tego symbolu w tym przypadku. Możemy ustawić elementy A w różnowartościowy ciąg $\omega_0, \omega_1, \omega_2, \ldots$ i przyjąć, że

$$\sum_{\omega\in A} p(\omega) = \sum_{i=0}^{\infty} p(\omega_i).$$

Wadą tej definicji jest, że może zależeć od wyboru ustawienia elementów zbioru A (patrz dodatek o sumowaniu wyrazów ciągu nieskończonego). Można też przyjąć, że

$$\sum_{\omega\in A} p(\omega) = \sup\{\sum_{\omega\in X} p(\omega) : X \subseteq A \wedge X \text{ jest skończony}\}.$$

Ta definicja z kolei nie odpowiada intuicjom, jeżeli funkcja p przyjmuje wartości ujemne. Zauważmy jednak, że zachodzi następujący lemat.

Lemat 1.1 *Jeżeli funkcja p przyjmuje na zbiorze A wartości nieujemne i $\omega_0, \omega_1, \omega_2, \ldots$ jest różnowartościowym ciągiem, takim, że $A = \{\omega_0, \omega_1, \omega_2, \ldots\}$, to szereg*

$$\sum_{i=0}^{\infty} p(\omega_i)$$

jest zbieżny wtedy i tylko wtedy, gdy istnieje

$$\sup\{\sum_{\omega\in X} p(\omega) : X \subseteq A \wedge X \textit{ jest skończony}\},$$

a jeżeli podany szereg jest zbieżny, to zachodzi równość

$$\sum_{i=0}^{\infty} p(\omega_i) = \sup\{\sum_{\omega\in X} p(\omega) : X \subseteq A \wedge X \textit{ jest skończony}\}.$$

Dowód. Suma szeregu o wyrazach dodatnich, czyli granica niemalejącego ciągu liczb jest równa supremum wyrazów tego ciągu. Wobec tego porównujemy suprema dwóch zbiorów:

$$A = \{\sum_{i=0}^{n} p(\omega_i) : n \in N\} \text{ oraz } B = \{\sum_{\omega\in X} p(\omega) : X \subseteq A \wedge X \text{ jest skończony}\}.$$

Oczywiście, $A \subseteq B$. Ponadto, każdy element zbioru B jest mniejszy od pewnego elementu B bądź jest jemu równy. Stąd wynika, że zbiory ograniczeń górnych zbiorów A i B są identyczne. Jeżeli zbiory ograniczeń są puste, to oba suprema nie istnieją. W przeciwnym razie oba suprema są identyczne. $\square$

Wniosek 1.2 *Przestawiając wyrazy zbieżnego szeregu o wyrazach nieujemnych otrzymujemy szereg o tej samej sumie.* $\square$

Zauważmy jeszcze, że na wykładzie analizy dowodzi się nieco ogólniejsze twierdzenie o przestawianiu wyrazów szeregów bezwzględnie zbieżnych, zaś sumę dowolnego zbioru przeliczalnego można zdefiniować wzorując się na teorii całki w następujący sposób:

$$\sum_{i=0}^{\infty} p(\omega_i) = \sup\{\sum_{\omega\in X} p(\omega) : X \subseteq A \text{ i jest skończony}\} + \inf\{\sum_{\omega\in X} p(\omega) : X \subseteq A \text{ i jest skończony}\}.$$

Licząc w ten sposób sumę najpierw sumujemy wyrazy dodatnie (suma ta jest równa supremum zbioru skończonych sum częściowych), następnie sumujemy wyrazy ujemne (otrzymujemy infimum tego zbioru), a w końcu obie sumy dodajemy. Pamiętajmy także, że $\sum_{\omega\in\emptyset} p(\omega) = 0$.

1.9 Rozkład prawdopodobieństwa dla zbiorów przeliczalnych

Kłopoty z defimicją rozkładu prawdopodobieństwa (i samego prawdopodobieństwa) w przypadku nieskończonych zbiorów zdarzeń elementarnych biorą się stąd, że powinniśmy sumować nieskończone zbiory liczb. Jest to łatwe jeszcze dla zbiorów przeliczalnych.

Przypuśćmy, że Ω jest przeliczalnym zbiorem zdarzeń elementarnych. Rozkładem prawdopodobieństwa nazywamy dowolną funkcję $p : \Omega \to [0, 1]$ taką, że

$$\sup\{\sum_{\omega\in A} p(\omega) : A \subseteq \Omega \wedge A \text{ jest skończony}\} = 1.$$

Wartość $p(\omega)$ jest nazywana prawdopodobieństwem wystąpienia zdarzenia (elementarnego) ω i jest interpretowana jako szansa wystąpienia tego zdarzenia lub częstość występowania tego zdarzenia.

Przytoczona definicja stwierdza dwa fakty. Jeżeli sumujemy prawdopodobieństwa skończenie wielu zdarzeń elementarnych, to zawsze otrzymamy liczbę $\leqslant 1$, a ponadto, sumując prawdopodobieństwa skończenie wiele (odpowiednio dobranych) zdarzeń elementarnych możemy otrzymać wartości dowolnie bliskie 1.

Jeżeli zbiór Ω jest nieskończony i ponumerujemy jego elementy w sposób różnowartościowy, np. przyjmiemy, że $\Omega = \{\omega_0, \omega_1, \omega_2, \ldots\}$, gdzie $\omega_k \neq \omega_l$ dla dowolnych różnych liczb naturalnych k i l, to warunek z definicji rozkładu można zastąpić równoważną równością

$$\sum_{k=0}^{\infty} p(\omega_k) = 1.$$

W ten sposób otrzymujemy równoważną definicję rozkładu prawdopodobieństwa.

1.10 Przeliczalne przestrzenie probabilistyczne

Przypuśćmy, że Ω jest przeliczalnym (nieskończonym) zbiorem zdarzeń elementarnych. W tym przypadku przestrzenią probabilistyczną nazywamy układ złożony ze zbioru zdarzeń elementarnych Ω, zbioru zdarzeń, czyli rodziny $\mathcal{P}(\Omega)$ wszystkich podzbiorów Ω, oraz prawdopodobieństwa $P : \mathcal{P}(\Omega) \to [0, 1]$, czyli funkcji spełniającej warunek

1. $P(\Omega) = 1$ oraz

2. dla dowolnego ciągu $A_0, A_1, A_2, \ldots$ parami rozłącznych zdarzeń (podzbiorów Ω) równości

$$P(\bigcup_{i=0}^{\infty} A_i) = \sum_{i=0}^{\infty} P(A_i).$$

Wartość $P(A)$ nazywamy prawdopodobieństwem zdarzenia A.

Zauważmy, że z równości podanej w definicji przestrzeni probabilistycznej wynika, że

$$P(A \cup B) = P(A) + P(B)$$

dla dowolnych rozłącznych zbiorów $A, B \subseteq \Omega$. Aby to wykazać, rozważamy ciąg $A, B, \emptyset, \emptyset, \emptyset, \ldots$. Prawdziwe są też wszystkie własności prawdopodobieństwa podane w rozdziale 1.7 oraz wzór wyrażający prawdopodobieństwo zdarzenia przez prawdopodobieństwo zdarzeń elementarnych podobny do wzoru znanego z przestrzeni skończonych. W przypadku przestrzeni przeliczalnych i nieskończonych, dla skończonego zbioru A wzór ten jest dokładnie taki sam, a dla nieskończonego zbioru $A = \{\omega_0, \omega_1, \omega_2, \ldots\}$ (i różnowartościowej numeracji jego elementów) – przyjmuje postać

$$P(A) = \sup\{P(X) : X \subseteq A \wedge X \text{ jest skończony}\} = \sum_{i=0}^{\infty} P(\omega_i).$$

Dla przeliczalnych zbiorów zdarzeń elementarnych – podobnie jak dla skończonych – mamy wzajemnie jednoznaczną odpowiedniość między rozkładami prawdopodobieństwa i przestrzeniami probabilistycznymi.

1.11 Addytywne i σ-addytywne funkcje zbiorów

Niech f będzie rzeczywistą funkcją określoną na pewnym ciele zbiorów $\mathcal{C}$. Funkcja f jest addytywna, jeżeli

$$f(A \cup B) = f(A) + f(B)$$

dla każdej pary rozłącznych zbiorów $A, B \in \mathcal{C}$.

Funkcja f jest σ-addytywna, jeżeli

$$f(\bigcup_{i=0}^{\infty} A_i) = \sum_{i=0}^{\infty} f(A_i)$$

dla każdego ciągu $A_0, A_1, A_2, \ldots$ parami rozłącznych zbiorów z $\mathcal{C}$ takich, że $\bigcup_{i=0}^{\infty} A_i \in \mathcal{C}$.

Nietrudno zauważyć, że każda σ-addytywna funkcja jest też addytywna.

Lemat 1.3 *Niech f będzie addytywną funkcją określoną na ciele podzbiorów Ω. Następujące warunki są równoważna:*

1. *f jest σ-addytywna,*

2. *Dla dowolnego wstępującego ciągu zbiorów $A_0, A_1, A_2, \ldots$ takiego, że $\bigcup_{i=0}^{\infty} A_i \in \mathcal{C}$ zachodzi równość*

$$f(\bigcup_{i=0}^{\infty} A_i) = \lim_{i} f(A_i).$$

3. *Dla dowolnego zstępującego ciągu zbiorów $A_0, A_1, A_2, \ldots$ takiego, że $\bigcap_{i=0}^{\infty} A_i \in \mathcal{C}$ zachodzi równość*
$$f(\bigcap_{i=0}^{\infty} A_i) = \lim_i f(A_i).$$

4. *Dla dowolnego zstępującego ciągu zbiorów $A_0, A_1, A_2, \ldots$ takiego, że $\bigcap_{i=0}^{\infty} A_i = \emptyset$ zachodzi równość*
$$\lim_i f(A_i) = 0.$$

Dowód. Aby dowieść, że warunek 1) implikuje warunek 2), korzystamy ze wzoru

$$\bigcup_{i=0}^{\infty} A_i = A_0 \cup \bigcup_{i=0}^{\infty} (A_{i+1} \setminus A_i)$$

słusznego dla dowolnego wstępującego ciągu zbiorów $A_0, A_1, A_2, \ldots$ i umożliwiającego urozłącznienie jego wyrazów. W dowodzie warunku 3) przechodzimy przez dopełnienia korzystając z warunku 2) i wzoru $f(A^c) = f(\Omega) - f(A)$. Warunek 4) wynika w oczywisty sposób z warunku 3). Pokażemy tylko, jak z warunku 4) wyprowadzić warunek 1).

Przypuśćmy, że $B_0, B_1, B_2, \ldots$ jest ciągiem parami rozłącznych elementów $\mathcal{C}$ takim, że $\bigcap_{i=0}^{\infty} B_i \in \mathcal{C}$. Przyjmijmy, że

$$A_n = (\bigcup_{i=0}^{\infty} B_i) \setminus (\bigcup_{i=0}^{n-1} B_i) = \bigcup_{i=n}^{\infty} B_i.$$

Nietrudno zauważyć, że
$$\bigcap_{n=0}^{\infty} A_n = \emptyset.$$

Wobec tego,

$$0 = \lim_n f(A_n) = \lim_n (f(\bigcup_{i=0}^{\infty} B_i) - \sum_{i=0}^{n-1} f(B_i)) = f(\bigcup_{i=0}^{\infty} B_i) - \lim_n \sum_{i=0}^{n-1} f(B_i) = f(\bigcup_{i=0}^{\infty} B_i) - \sum_{i=0}^{\infty} f(B_i). \ \Box$$

1.12 σ-ciała zbiorów

σ-ciałem nazywamy ciało zbiorów, które dodatkowo jest zamknięte ze względu na przeliczalne sumy mnogościowe. Oczywiście, każde ciało zbiorów jest zawarte w pewnym σ-ciele i wśród tych σ-ciał zawsze jest najmniejsze.

Lemat 1.4 *Niech $\mathcal{C}$ będzie ciałem zbiorów. Następujące warunki są równoważne:*

- *$\mathcal{C}$ jest σ-ciałem zbiorów,*
- *$\mathcal{C}$ jest zamknięte ze względu na sumy wstępujących ciągów zbiorów,*
- *$\mathcal{C}$ jest zamknięte ze względu na przekroje zstępujących ciągów zbiorów.* $\Box$

Każdy zbiór postaci $\mathcal{P}(\Omega)$ jest zarówno ciałem, jak i σ-ciałem zbiorów. Przykład ciała zbiorów, które nie jest σ-ciałem jest opisany w rozdziale 1.4. Każde skończone ciało zbiorów jest σ-ciałem, ponieważ każda przeliczalna suma zbiorów jest też sumą skończoną.

1.13 Dowolne przestrzenie probabilistyczne

Przestrzenią probabilistyczną nazywamy układ złożony

1. ze zbioru zdarzeń elementarnych Ω,
2. z σ-ciała $\mathcal{Z}$ zawartego w $\mathcal{P}(\Omega)$, czyli zbioru zdarzeń,
3. i prawdopodobieństwa $P : \mathcal{Z} \to [0,1]$ będącego σ-addytywną funkcją taką, że $P(\Omega) = 1$.

W przeciwieństwie do przeliczalnych przestrzeni probabilistycznych, w przypadku ogólnym ograniczamy się do pewnego σ-ciała zawartego w $\mathcal{P}(\Omega)$, ponieważ w niektórych ważnych przypadkach nie można określić prawdopodobieństwa dla wszystkich podzbiorów Ω. W przypadku przestrzeni przeliczalnych też można rozważać zbiory zdarzeń różne od $\mathcal{P}(\Omega)$, ale rozważanie takich przestrzeni nie wnosi nic interesującego.

2 Definiowanie przestrzeni probabilistycznych

2.1 Pierwszy przykład

Zdefiniujemy przestrzeń probabilistyczną, w której zbiorem zdarzeń elementarnych Ω będzie odcinek

$$[0,1) = \{x \in R : 0 \leqslant x < 1\} = \Omega.$$

Definiowana przestrzeń odpowiada doświadczeniu losowemu, w którym losujemy liczbę rzeczywistą z przedziału $[0,1)$. Prawdopodobieństwo wylosowania liczby z jakiegoś przedziału będzie proporcjonalne do długości przedziału.

Niech $\mathcal{C}$ oznacza zbiór

$$\mathcal{C} = \{\bigcup_{i=0}^{n} [a_i, b_i) \subseteq [0,1) : 0 \leqslant a_1 < b_1 < a_2 < b_2 < \ldots < a_n < b_n \leqslant 1\}.$$

Fakt 2.1 Rodzina $\mathcal{C}$ jest ciałem zbiorów.

Dowód. Nietrudno zauważyć, że jeżeli $[a,b) \in \mathcal{C}$ i $0 < a < b < 1$, to

$$[a,b)^c = [0,a) \cup [b,1).$$

W ogólnym przypadku

$$(\bigcup_{i=1}^{n} [a_i, b_i))^c = [0, a_1) \cup \bigcup_{i=1}^{n-1} [b_i, a_{i+1}) \cup [b_n, 1)$$

(czasem pierwszy i/lub ostatni składnik jest pusty i można go pominąć).

Łatwo też dowieść zamkniętość $\mathcal{C}$ ze względu na przekrój. Jeżeli $1 \leqslant a < c < b < d \leqslant 1$, to

$$[a,b) \cap [c,d) = [c,b).$$

Nietrudno zauważyć, że przekrój dwóch sum parami rozłącznych odcinków jest – na mocy prawa rozdzielności – sumę przekrojów odcinków i przekroje te są parami rozłączne. $\square$

Można też wykazać, że przedstawienie zbioru należącego do $\mathcal{C}$ w postaci podanej w definicji $\mathcal{C}$ jest jednoznaczne, a więc jeżeli

$$\bigcup_{i=0}^{n}[a_i, b_i) = \bigcup_{i=0}^{m}[c_i, d_i)$$

dla pewnych ciągów spełniających nierówności

$$0 \leqslant a_1 < b_1 < a_2 < b_2 < \ldots < a_n < b_n \leqslant 1 \text{ oraz } 0 \leqslant c_1 < d_1 < c_2 < d_2 < \ldots < c_m < d_m \leqslant 1,$$

to $n = m$ oraz $a_i = c_1$ i $b_i = d_i$ dla wszystkich $i = 1, \ldots, n$.

Dzięki jednoznaczności przedstawienia, zbiorowi $A = \bigcup_{i=1}^{n}[a_i, b_i)$ należącemu do $\mathcal{C}$ możemy przyporządkować prawdopodobieństwo

$$P(A) = \sum_{i=1}^{n}(b_i - a_i).$$

Jednoznaczność przedstawienia gwarantuje poprawność definicji P.

Fakt 2.2 Jeżeli odcinek

$$[a, b) = \bigcup_{i=1}^{n}[a_i, b_i)$$

dla parami rozłącznych odcinków $[a_i, b_i)$, $i = 1, \ldots, n$, to

$$b - a = \sum_{i=1}^{n}(b_i - a_i).$$

Dowód. Jeżeli sumowane odcinki ponumerujemy tak, aby $a_i < a_{i+1}$ dla wszystkich i, to z założonej równości i rozłączności sumowanych odcinków wynika, że $b_i = a_{i+1}$ dla każdego $i = 1, \ldots, n-1$. Wiedząc to łatwo wyliczyć interesującą nas sumę. □

Udowodniony właśnie wzór pozwala wykazać

Fakt 2.3 Funkcja P jest addytywna.

Dowód. Przypuśćmy, że $A, B \in \mathcal{C}$ są zbiorami rozłącznymi. Oczywiście, $A \cup B$ należy do $\mathcal{C}$ i ma wymagane w definicji $\mathcal{C}$ przedstawienie w postaci sumy odcinków. Suma przedstawień A i B jest innym, drobniejszym przedstawieniem $A \cup B$ w postaci sumy odcinków. Grupując odpowiednio odcinki z tego przedstawienia otrzymujemy odcinki z przedstwienia z definicji. Dzięki temu, posługując się powyższym wzorem pokazujemy, że $P(A \cup B) = P(A) + P(B)$. □

Fakt 2.4 Funkcja P jest także σ-addytywna.

Dowód. Aby się o tym przekonać, wystarczy wziąć wstępujący ciąg $A_0, A_1, A_2, \ldots$ elementów $\mathcal{C}$ o sumie równej $[0, 1)$ i wykazać, że $\lim_i P(A_i) = 1$ (patrz lemat 1.3).

Weźmy więc taki ciąg. Najpierw powiększymy trochę jego wyrazy. Przypuśćmy, że

$$A_n = \bigcup_{i=1}^{m}[a_{n,i}, b_{n,i}).$$

Niech $\varepsilon > 0$ będzie ustaloną liczbą. Zamiast A_n będziemy rozważać zbiory

$$A'_n = \bigcup_{i=1}^{m} (a_i - \frac{\varepsilon}{m2^{n+2}}, b_i)$$

oraz

$$A''_n = \bigcup_{i=1}^{m} [a_i - \frac{\varepsilon}{m2^{n+2}}, b_i)$$

Oczywiście,

$$[0,1) \subseteq \bigcup_{i=1}^{\infty} A'_i \subseteq \bigcup_{i=1}^{\infty} A''_i,$$

zbiór A'_n powstał z A_n przez dodanie pewnych odcinków i sam jest sumą odcinków otwartych, oraz

$$P(([0,1) \cap A''_n) \setminus A_n) \leqslant \frac{\varepsilon}{2^{n+2}}.$$

Niech a będzie równe kresowi górnemu (supremum) zbioru

$$\mathcal{X} = \{x \leqslant 1 : \exists k \in N \; [0,x] \subseteq \bigcup_{i=0}^{k} A'_i\}.$$

Oczywiście, $0 \leqslant a \leqslant 1$. Pokażemy, że $a = 1$ sprowadzając do sprzeczności założenie, że $a < 1$. Jeżeli $a \in [0,1)$, to $a \in A_n \subseteq A'_n$ dla pewnego n. Zbiór A'_n jest sumą odcinków otwartych, więc

$$a \in (a_{n,i} - \frac{\varepsilon}{m2^{n+2}}, b_{n,i})$$

dla pewnego i. Dowolnie blisko $\sup \mathcal{X}$ są elementy zbioru $\mathcal{X}$. Wobec tego, istnieją liczby $a' \in \mathcal{X}$ i $a'' < 1$, oraz liczba naturalna k takie, że

$$a' < a < a'', \quad [a', a''] \subseteq A'_n \text{ oraz } [0, a'] \subseteq \bigcup_{i=0}^{k} A'_i.$$

Mamy więc także zawieranie

$$[0, a''] \subseteq \bigcup_{i=0}^{\max(n,k)} A'_i.$$

Z definicji a oraz tego zawierania wynika, że $a'' \leqslant a$, a to przeczy wyborowi a''. Tak więc $a = 1$.

Z tego, że $a = 1$ wynika, że dla pewnego K zachodzi zawieranie

$$[0, 1 - \frac{\varepsilon}{2}) \subseteq \bigcup_{i=0}^{K} A'_i \subseteq \bigcup_{i=0}^{K} ([0,1) \cap A''_i).$$

Stąd z kolei i z zawierań $A_i \subseteq A_K$ dla $i \leqslant K$ otrzymujemy, że

$$[0,1) \setminus A_K \subseteq [1 - \frac{\varepsilon}{2}, 1) \cup (\bigcup_{i=0}^{K} ([0,1) \cap A''_i)) \setminus A_K \subseteq [1 - \frac{\varepsilon}{2}, 1) \cup \bigcup_{i=0}^{K} (([0,1) \cap A''_i) \setminus A_i).$$

Wobec tego,

$$P([0,1) \setminus A_K) \leqslant \frac{\varepsilon}{2} + \sum_{i=0}^{K} \frac{\varepsilon}{2^{i+2}} \leqslant \varepsilon.$$

Tym bardziej,

$$\lim_i P(A_i) \geqslant P(A_K) \geqslant 1 - \varepsilon.$$

Ponieważ liczba ε może być dowolnie mała, więc $\lim_i P(A_i) \geqslant 1$. Odwrotna nierówność jest oczywista. □

Ciało $\mathcal{C}$ nie jest jednak σ-ciałem. Oczywiście, dla dowolnej liczby a takiej, że $0 \leqslant a < 1$, zbiór $\{a\}$ nie należy do $\mathcal{C}$, ale

$$\{a\} = \bigcap_{i=0}^{\infty} [a, a + \frac{1}{i+1}).$$

Z udowodnionych wyżej własności funkcji P wynika jednak, że rozszerza się ona jednoznacznie na najmniejsze σ-ciało podzbiorów $[0,1)$ zawierające $\mathcal{C}$. Oznaczmy to σ-ciało symbolem $\mathcal{B}$. Zwykle nazywa się ono σ-ciałem zbiorów borelowskich. Jest to dostatecznie obszerna rodzina zbiorów i wystarczająca z punktu widzenia rachunku prawdopodobieństwa. Możemy więc uważać, że zdefiniowaliśmy przestrzeń probabilistyczna Ω z σ-ciałem $\mathcal{B}$ i prawdopodobieństwem P. Dowód twierdzenia o możliwości rozszerzenia jest zawarty w jednym z dodatków, patrz twierdzenie 7.1.

Przedstawiana konstrukcja jest często wykorzystywana w matematyce, często też pojawiają się konstrukcje analogiczne. Jeżeli zamiast odcinka weźmiemy kwadrat $[0,1)^2$, to w podobny sposób zdefiniujemy funkcję, która zbiorom przyporządkowuje ich powierzchnię. Podobnie możemy definiować objętość brył. Zamiast odcinka lub kwadratu bierze się też prostą R lub płaszczyznę R^2. Wtedy trzeba np. rozważać „odcinki” nieskończone, takie jak $(-\infty, 0)$ oraz $[1, \infty)$ i trzeba coś zrobić z warunkiem $P(\Omega) = 1$. Można odrzucić go lub inaczej zdefiniować prawdopodobieństwo, tak aby ten warunek był jednak spełniony.

2.2 Drugi przykład

Będziemy budować przestrzeń probabilistyczną pozwalającą analizować następujący program (który można interpretować jako program losujący liczbę naturalną, a właściwie przedstawienie dwójkowe liczby naturalnej):

1. **podstaw do zmiennej** *wynik* **słowo puste**
2. **i powtarzaj następujące czynności**
   (a) **wylosuj znak 0, 1 lub 2,**
   (b) **jeżeli wylosowany znak jest różny od 2, to dopisz go do zmiennej** *wynik***,**
3. **tak długo, aż zostanie wylosowany znak 2,**
4. **zwróć** *wynik***.**

Przyjmujemy, że zbiór zdarzeń elementarnych Ω jest równy zbiorowi wszystkich ciągów o wyrazach 0 i 1, skończonych i nieskończonych. Taki ciąg opisuje przebieg losowania podczas wykonania algorytmu. Ciąg pusty oznacza, że za pierwszym razem został wylosowany znak 2. Ciąg 01 odpowiada sytuacji, gdy algorytm losował trzykrotnie, kolejno 0, 1 i 2. Ciągi nieskończone odpowiadają takim przebiegom losowania, w których nigdy nie została wylosowana cyfra 2.

Niech teraz α oznacza skończony ciąg zerojedynkowy. Symbol $[\alpha]$ będzie oznaczać zbiór

$$\{\omega \in \Omega : \alpha \text{ jest odcinkiem początkowym } \omega\},$$

czyli zbiór tych przebiegów losowań, które rozpoczęły się wylosowaniem kolejnych wyrazów ciągu α. Jeżeli na chwilę przyjmiemy, że losujemy przedstawienie dwójkowe liczby rzeczywistej z przedziału $[0,1)$ i liczba a ma przedstawienie $0,\alpha(1)\alpha(2)\ldots\alpha(n)$ (n to długość α), to zbiór $[\alpha]$ możemy utożsamiać z odcinkiem $[a, a+\frac{1}{2^n}]$. Fakt ten wskazuje na podobieństwo formalne tego i poprzedniego przykładu przestrzeni probabilistycznej.

Oczywiście,

$$[\alpha] = \{\alpha\} \cup [\alpha 0] \cup [\alpha 1], \tag{1}$$

gdzie $\alpha 0$ oznacza ciąg α z dopisanym znakiem 0 na końcu. Przyjmujemy, że interesuje nas prawdopodobieństwo P, dla którego zbiory $\{\alpha\}$ i $[\alpha]$ są zdarzeniami i takie, że

$$P(\{\alpha\}) = p \cdot P([\alpha])$$

oraz

$$P([\alpha 0]) = P([\alpha 1]) = \frac{1-p}{2} P([\alpha])$$

dla pewnej, ustalonej liczby $p \in (0,1)$. Równości te powinny wynikać ze sposobu losowania podczas jednokrotnego wykonania pętli algorytmu. Przyjmujemy, że 2 jest losowane z częstością p, a cyfry 0 i 1 są losowane z tą samą częstością, które musi być więc równa $\frac{1-p}{2}$. Interesuje nas, czy jest przestrzeń probabilistyczna spełniająca podane warunki.

Zauważmy jeszcze, że z równości (1), z założonych własności P i faktu, że $P([\emptyset]) = P(\Omega) = 1$, stosując zasadę indukcji otrzymujemy, że

$$P([\alpha]) = \frac{(1-p)^n}{2^n} \quad \text{oraz} \quad P(\{\alpha\}) = p \cdot \frac{(1-p)^n}{2^n},$$

gdzie n oznacza długość α.

Jeżeli przyjęliśmy, że dla skończonych ciągów α zbiory $[\alpha]$ i $\{\alpha\}$ są zdarzeniami, to zdarzeniami powinny być także zbiory należące do

$$\mathcal{C} = \{\bigcup_{i=1}^{n} [\alpha_i] \cup \bigcup_{i=1}^{m} \{\beta_i\} \subseteq \Omega : n, m \in N \wedge \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m \text{ są skończone}\}.$$

Zauważmy, że $\emptyset \in \mathcal{C}$, ponieważ jest to suma zera zbiorów postaci $[\alpha_i]$ i zera zbiorów postaci $\{\beta_i\}$.

Fakt 2.5 $\mathcal{C}$ jest ciałem zbiorów.

Dowód. Stosunkowo łatwo pokazuje się, że rodzina $\mathcal{C}$ jest zamknięta ze względu na przekrój i dopełnienie.

Jeżeli liczymy przekrój elementów $\mathcal{C}$ i korzystamy z prawa rozdzielności, to przekrój ten okazuje się sumą przekrojów zbiorów postaci $[\alpha]$ i/lub $\{\beta\}$. Wystarczy zauważyć, że przekroje takich zbiorów też mają taką postać lub są zbiorami pustymi. Jeżeli jeden z czynników ma postać $\{\beta\}$, to iloczyn jest pusty lub równy $\{\beta\}$. Jeżeli α jest odcinkiem początkowym β, to $[\alpha] \cap [\beta] = [\beta]$. Jeżeli ani α nie jest odcinkiem początkowym β, ani β nie jest odcinkiem początkowym α, to $[\alpha] \cap [\beta] = \emptyset$.

Dzięki prawu de Morgana zamkniętość $\mathcal{C}$ ze względu na dopełnienie wynika z zamkniętości na przekrój. Wystarczy zauważyć, że dopełnienia zbiorów postaci $[\alpha]$ i $\{\beta\}$ należą do $\mathcal{C}$. Na przykład

$$[01110]^c = [1] \cup [00] \cup [010] \cup [0110] \cup [01111] \cup \{\varepsilon, 0, 01, 011, 0111\}$$

oraz

$$\{01110\}^c = [011100] \cup [011101] \cup [01110]^c.$$

Uogólnienie podanych wzorów pozostawiam czytelnikowi. □

Łatwo określić P dla argumentów $A \in \mathcal{C}$. Przypuśćmy, że

$$A = \bigcup_{i=1}^{n} [\alpha_i] \cup \bigcup_{i=1}^{m} \{\beta_i\}.$$

Jeżeli składniki w podanym wzorze są parami rozłączne, to oczywiście przyjmujemy, że

$$P(A) = \sum_{i=1}^{n} P([\alpha_i]) + \sum_{i=1}^{m} P(\{\beta_i\}).$$

Jeżeli nie są rozłączne, to łatwo można znaleźć inne przedstawienie A z rozłącznymi składnikami. Wystarczy usunąć składniki $[\alpha_i]$ dla i takich, że α_i jest dla pewnego $j \neq i$ wydłużeniem α_j i sprawdzić, czy słowa postaci β_i nie należą do kilku składników (jeżeli należą, to można je pominąć). Tak więc potrafimy określić wartość $P(A)$ dla dowolnego $A \in \mathcal{C}$. Nie jest jednak jasne, czy potrafimy to zrobić w sposób jednoznaczny.

Fakt 2.6 Jeżeli

$$[\alpha] = \bigcup_{i=1}^{n} [\alpha_i] \cup \bigcup_{i=1}^{m} \{\beta_i\}$$

i składniki w podanym wzorze są parami rozłączne, to

$$P([\alpha]) = \sum_{i=1}^{n} P([\alpha_i]) + \sum_{i=1}^{m} P(\{\beta_i\}).$$

Dowód. Fakt ten dowodzimy przez indukcję ze względu na n. Zauważmy, że zbiory postaci $[\alpha]$ mają kilka oczywistych własności. Każdy taki zbiór jest nieskończony (mocy continuum), a nawet nieskończona jest każda różnica $[\alpha] \setminus [\beta]$ dla każdego ciągu β będącego istotnym wydłużeniem α. Ponadto, jeżeli $\beta \in [\alpha]$, to $[\beta] \subseteq [\alpha]$.

Z przytoczonych własności wynika, że $n > 0$, a gdy $n = 1$, to $m = 0$ i $\alpha = \alpha_1$.

Załóżmy, że $n > 1$. Najpierw znajdujemy najdłuższy ciąg spośród $\alpha_1, \ldots, \alpha_n$. Przypuśćmy, że jest to ciąg $\gamma 0$. Wtedy wśród $\alpha_1, \ldots, \alpha_n$ znajduje się także ciąg $\gamma 1$, a wśród ciągów $\beta_1, \ldots \beta_m$ – ciąg γ. Jeżeli we wzorze na $[\alpha]$ składniki $[\gamma 0]$, $[\gamma 1]$ i $\{\gamma\}$ zastąpimy przez $[\gamma]$, to otrzymamy nowe przedstawienie zbioru $[\alpha]$. Dla tego przedstawienia możemy skorzystać z założenia indukcyjnego. W ten sposób otrzymujemy wzór na $P(A)$ zawierający składnik $P([\gamma])$. Teraz wystarczy podstawić za $P([\gamma])$ równe mu wyrażenie $P([\gamma 0]) + P([\gamma 1]) + P(\{\gamma\})$. □

Fakt 2.7 Definicja P na ciele $\mathcal{C}$ jest poprawna.

Dowód. Jeżeli weźmiemy dwa przedstawienia zbioru $A \in \mathcal{C}$ w postaci sum parami rozłącznych składników i wyliczymy ich przekrój korzystając z prawa rozdzielności, to znajdziemy kolejne, drobniejsze przedstawienia A. Odpowiednio grupując i sumując składniki tego przedstawienia otrzymujemy składniki wyjściowych przedstawień. Dalej wystarczy skorzystać z poprzedniego faktu. □

Addytywność funkcji P na ciele $\mathcal{C}$ jest oczywista. Pokażemy jeszcze, że funkcja P jest σ-addytywna, ale najpierw wykażemy potrzebny lemat techniczny.

Lemat 2.8 *Załóżmy, że*

$$A = \bigcup_{i=1}^{n_1}[\alpha_{1,i}] \cup \bigcup_{i=1}^{m_1}\{\beta_{1,i}\} \supseteq \bigcup_{i=1}^{n_2}[\alpha_{2,i}] \cup \bigcup_{i=1}^{m_2}\{\beta_{2,i}\}$$

oraz zbioru A nie można inaczej przedstawić w podany sposób zmniejszając $n_1 + m_1$. Wtedy dla każdego $i = 1, \ldots, n_2$ istnieje j takie, że $1 \leqslant j \leqslant n_1$ oraz $\alpha_{1,j}$ jest odcinkiem początkowym $\alpha_{2,i}$.

Dowód. Najpierw zauważmy, że warunek $[\alpha'] \subseteq [\alpha]$ oznacza, że α jest odcinkiem początkowym α'. Aby się o tym przekonać, wystarczy α' wydłużyć samymi 0, a następnie samymi 1, oba te ciągi należą do $[\alpha]$ i z tego powodu zawierają α. Ponadto, jeżeli między zbiorami $[\alpha']$ i $[\alpha]$ nie zachodzi żadne zawieranie, to zbiory te są rozłączne.

Weźmy teraz jeden z ciągów postaci $\alpha_{2,i}$. Są możliwe dwa przypadki. Albo zbiór $[\alpha_{2,i}]$ jest zawarty w zbiorze postaci $[\alpha_{1,j}]$ dla pewnego j i teza wynika z powyższej uwagi. Albo takie zawieranie nie zachodzi dla żadnego j. Wtedy, po przemnożeniu założonego zawierania przez $[\alpha_{2,i}]$ otrzymujemy wzór postaci

$$[\alpha_{2,i}] = \bigcup_k [\alpha_{1,j_k}] \cup \bigcup_l \{\beta_{1,i_l}\}.$$

Wzór ten pozwala uprościć przedstawienie zbioru A i przeczy założeniu o postaci przedstawienia. □

Fakt 2.9 Funkcja $P : \mathcal{C} \quad R$ jest σ-addytywna.

Dowód. Z lematu 1.3 wynika, że wystarczy pokazać dla dowolnego zstępującego ciągu $A_0, A_1, A_2, \ldots$ elementów $\mathcal{C}$ o pustym przekroju, że

$$\lim_{k\to\infty} P(A_k) = 0.$$

Udowodnimy własność silniejszą, a mianowicie pokażemy, że od pewnego miejsca ciąg $A_0, A_1, A_2, \ldots$ zawiera wyłącznie zbiory puste. Przypuśćmy, że

$$A_k = \bigcup_{i=1}^{n_k}[\alpha_{k,i}] \cup \bigcup_{i=1}^{m_k}\{\beta_{k,i}\} \tag{2}$$

i podane przedstawienia mają możliwie mało składników.

Najpierw załóżmy, że ze zbioru $I = \{\alpha_{k,i} : k \in N \wedge i \leqslant n_k\}$ można wybrać ciąg $\gamma_0, \gamma_1, \gamma_2, \ldots$ coraz dłuższych elementów (takich, że γ_i jest właściwym odcinkiem początkowym γ_{i+1}). Niech

γ będzie ciągiem nieskończonym zawierającym jako odcinki początkowe wszystkie ciągi postaci γ_i. Pokażemy, że $\gamma \in \bigcap_{k=0}^{\infty} A_k$.

Niech k będzie dowolną liczbą naturalną. W przedstawieniach (2) zbiorów $A_0, \ldots, A_k$ występuje skończenie wiele ciągów postaci $\alpha_{j,i}$. Przyjmijmy, że γ_l nie występuje w tych przedstawieniach. Wobec tego, dla pewnej liczby $t > k$ mamy

$$\gamma \in [\gamma_l] \subseteq A_t \subseteq A_k.$$

Oznacza to, że $\gamma \in \bigcap_k A_k$ i przeczy założeniu o pustości tego przekroju. Wobec tego, nie można utworzyć ciągu $\gamma_0, \gamma_1, \gamma_2, \ldots$ o podanych własnościach.

Znana jest jednak konstrukcja takiego ciągu $\gamma_0, \gamma_1, \gamma_2, \ldots$ przy założeniu, że zbiór I zawiera dowolnie długie ciągi. Można to zrobić w następujący sposób: Rozważamy ciągi $\alpha_{0,1}, \ldots, \alpha_{0,n_0}$. Z lematu 2.8 wynika, że wszystkie inne ciągi ze zbioru I są wydłużeniami wymienionych. Dokładniej, zbiór I zawiera dowolnie długie ciągi i rozbija się na zbiory wydłużeń poszczególnych, wymienionych ciągów. Tak więc dla pewnego $i \leqslant n_0$ ciąg $\alpha_{0,i}$ ma w zbiorze I dowolnie długie wydłużenia. Bierzemy ten ciąg i powtarzamy to samo rozumowanie dla elementów ciągu $\alpha_{1,1}, \ldots, \alpha_{1,n_1}$ należących do $[\alpha_{0,i}]$ i zbioru $I_1 = I \cap [\alpha_{0,i}]$. W ten sposób możemy skonstruować ciąg coraz dłuższych elementów I. Jak wykazaliśmy, takiego ciągu nie można utworzyć. Wobec tego, zbiór I nie zawiera dowolnie długich ciągów, długości jego elementów są ograniczone i ma skończenie wiele elementów.

Skończoność zbioru I oznacza, że dla wszystkich, dostatecznie dużych k w przedstawieniu (2) nie ma składników postaci α, czyli $n_k = 0$. Dla takich k zbiory A_k są skończone, są coraz mniejsze i ich przekrój jest pusty. Mogą zmniejszać się skończenie wiele razy. Ostatecznie, dla dostatecznie dużych k także $m_k = 0$ i zbiory A_k są puste. $\square$

Twierdzenie o rozszerzaniu stwierdza, że funkcję P można rozszerzyć do prawdopodobieństwa określonego na najmniejszym σ-ciele podzbiorów Ω zawierającym $\mathcal{C}$.

Dalej zajmiemy się zbiorem

$$X = \{\omega \in \Omega : \omega \text{ jest nieskończony}\}.$$

Po pierwsze zauważmy, że dopełnienie X^c jest zbiorem przeliczalnym. Co więcej, jeżeli $\omega \notin X$, to $\{\omega\} \in \mathcal{C}$. Wobec tego, zbiór

$$X^c = \bigcup_{\beta \notin X} \{\beta\}$$

jest przeliczalną sumą elementów $\mathcal{C}$ i należy do σ-ciała zawierającego $\mathcal{C}$. To samo można więc powiedzieć o jego dopełnieniu, czyli o zbiorze X.

Dodatkowo, zbiór X nie należy do $\mathcal{C}$. Elementy $\mathcal{C}$ są albo skończone, albo mają moc continuum. Wobec tego, dopełnienie X^c nie należy do $\mathcal{C}$, a więc to samo dotyczy X. Obliczymy jeszcze $P(X)$ i $P(X^c)$. Zauważmy, że

$$P(X^c) = \sum_{\beta \notin X} P(\{\beta\}) = \sum_{n=0}^{\infty} \sum_{|\,\beta\,|=n} p \cdot \frac{(1-p)^n}{2^n} = p \sum_{n=0}^{\infty} (1-p)^n = p \cdot \frac{1}{1-(1-p)} = 1.$$

Stąd wynika, że $P(X) = 1 - P(X^c) = 0$. Tak więc rozważany algorytm kończy pracę z prawdopodobieństwem 1, a zapętla się z prawdopodobieństwem 0.

2.3 Dystrybuanta

Przypuśćmy, że rozważamy przestrzeń probabilistyczną $(R, \mathcal{B}, P)$, w której zbiorem zdarzeń elementarnych jest zbiór liczb rzeczywistych (a więc losujemy liczby rzeczywiste), a $\mathcal{B}$ jest σ-ciałem zbiorem borelowskich ($\mathcal{B}$ jest generowane np. przez półproste postaci $(-\infty, x)$). Przyjmijmy, że funkcja $f : R \to R$ jest zdefiniowana wzorem

$$f(x) = P((-\infty, x)).$$

Funkcję f nazywamy dystrybuantą (rozkładu) prawdopodobieństwa P (lub miary probabilistycznej P).

Lemat 2.10 *Dystrybuanta f prawdopodobieństwa P ma następujące własności:*

1. $\lim_{x\to-\infty} f(x) = 0$,
2. $\lim_{x\to\infty} f(x) = 1$,
3. $\lim_{x\to a-} f(x) = f(a)$,
4. *f jest niemalejąca, czyli $f(x) \leqslant f(y)$ jeżeli tylko $x \leqslant y$.*

Dowód. Monotoniczność dystrybuanty jest oczywistą konsekwencją monotoniczności prawdopodobieństwa. Trzy granice z tezy obliczamy w taki sam sposób. Najpierw zauważamy, że wystarczy rozważać monotoniczne ciągi argumentów (malejące w pierwszym przypadku i rosnące w pozostałych). Następnie korzystamy z ciągłości prawdopodobieństwa. Na przykład, jeżeli $x_0, x_1, x_2 \dots$ jest niemalejącym ciągiem zbieżnym do a, to

$$\lim_i f(x_i) = \lim_i P((-\infty, x_i)) = P(\bigcup_{i=0}^{\infty}(-\infty, x_i)) = P((-\infty, a)) = f(a). \quad \Box$$

2.4 Dystrybuanty a prawdopodobieństwa

Przypuśćmy, że mamy funkcję $f : R \to [0, 1]$ o własnościach podanych w lemacie 2.10. Wtedy możemy zdefiniować prawdopodobieństwo określone na σ-ciele zbiorów borelowskich przyjmując, że

$$P([a, b)) = f(b) - f(a)$$

dla wszystkich $a, b \in R$ takich, że $a < b$. Podobnie, jak w rozdziale 2.1, pokazujemy, że istnieje dokładnie jedno prawdopodobieństwo określone na σ-ciele zbiorów borelowskich i spełniające podane równości.

Zauważmy, że prawdopodobieństwo P spełnia także warunki

$$P((-\infty, a)) = P(\bigcup_{n=1}^{\infty}[a-n, a)) = \lim_{n\to\infty} P([a-n, a)) = \lim_{n\to\infty}(f(a) - f(a-n)) = f(a)$$

dla wszystkich $a \in R$. Oznacza to, że funkcja f jest dystrybuantą prawdopodobieństwa P.

Pokazaliśmy więc, że istnieje wzajemnie jednoznaczna odpowiedniość między prawdopodobieństwami określonymi na σ-ciele zbiorów borelowskich i funkcjami spełniającymi tezę lematu 2.10.

2.5 Gęstość prawdopodobieństwa

Przypuśćmy, że $g : R \to [0, \infty)$ jest funkcją taką, że

$$\int_{-\infty}^{\infty} g(x)dx = 1.$$

Wtedy funkcja $f : R \to R$ zdefiniowana wzorem

$$f(a) = \int_{-\infty}^{a} g(x)dx$$

spełnia warunki podane w lemacie 2.10 i jest dystrybuantą pewnego prawdopodobieństwa P. Nietrudno zauważyć, że

$$P([a,b)) = \int_{a}^{b} g(x)dx$$

dla wszystkich $a, b \in R$ takich, że $a < b$. Funkcję g nazywamy gęstością prawdopodobieństwa P. Pojęcie gęstości jest uogólnieniem pojęcia rozkładu prawdopodobieństwa wprowadzonego, gdy rozważaliśmy skończone i przeliczalne przestrzenie probabilistyczne.

Dowodzi się, że jeżeli dla pewnego prawdopodobieństwa P istnieje gęstość, to dystrybuanta P jest ciągła. Łatwo podać przykłady nieciągłych funkcji spełniających warunki z lematu 2.10. Wobec tego, nie każde prawdopodobieństwo określone na σ-ciele zbiorów borelowskich ma gęstość.

Ważnym przykładem gęstości jest funkcja $g : R \to R$ zdefiniowana wzorem

$$g(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}.$$

Jest to gęstość tzw. rozkładu normalnego.

3 Zmienne losowe

Przypuśćmy, że doświadczenie losowe łączy się z pomiarem jakiejś wielkości. Na przykład, chcemy bardzo dokładnie zważyć pewien przedmiot. Wobec tego, kładziemy go na wadze i ważymy popełniając wiele błędów zależnych od ustawienia wagi, jej sprawności technicznej, dokładności podziałki, ostrości wzroku itp. Pojedyńczy pomiar jest więc obarczonym pewnym błędem losowym, a wykonując ważenie otrzymujemy pewną liczbę, która jest sumą wagi przedmiotu i popełnionego błędu. W takiej sytuacji z doświadczeniem losowym (a więc pojedyńczym pomiarem wagi) jest związana tzw. zmienna losowa (czyli wynik pomiaru). Będziemy formalizować i opisywać taką sytuację i podobne.

3.1 Definicja zmiennej losowej

Przypuśćmy, że mamy przestrzeń probabilistyczną ze zbiorem zdarzeń elementarnych Ω, zbiorem zdarzeń $\mathcal{Z}$ i prawdopodobieństwem P. Zmiennymi losowymi będziemy nazywać funkcje $X : \Omega \to R$. Dopuszcza się również, że wartościami zmiennej losowej są symbole ∞ i $-\infty$. Zmienna losowa X powinna spełniać pewne warunki. Wydaje się naturalne, że chcemy na przykład mówić o prawdopodobieństwie zdarzenia polegającego na tym, że wartość zmiennej losowej jest z określonego przedziału, a więc

$$\{\omega \in \Omega : a \leqslant X(\omega) < b\} \in \mathcal{Z}.$$

Przyjmujemy, że funkcja $f : \Omega \to R$ jest zmienną losową, jeżeli spełnia podany warunek dla wszystkich $a, b \in R$, $a < b$.

Zauważmy od razu, że jeżeli zbiorem zdarzeń jest $\mathcal{P}(\Omega)$, to każda funkcja jest zmienną losową. Z własności przeciwobrazów bez trudu otrzymujemy, że

Lemat 3.1 *Jeżeli $f : \Omega \to R$ jest funkcją losową, to*

$$\vec{X}^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{Z}$$

dla każdego zbioru borelowskiego $A \subseteq R$. □

Wobec tego, jeżeli X jest zmienną losową, to zdarzeniami są także zbiory takie, jak $\{\omega \in \Omega : a \leqslant X(\omega) \leqslant b\}$, $\{\omega \in \Omega : a < X(\omega) < b\}$, $\{\omega \in \Omega : X(\omega) < b\}$, $\{\omega \in \Omega : X(\omega) = b\}$ itp.

Lemat 3.2
1. *Funkcje stałe są zmiennymi losowymi.*
2. *Jeżeli $f : R \to R$ jest funkcją ciągłą i X jest zmienną losową, to funkcja Y definiowana wzorem $Y(\omega) = f(X(\omega))$ jest zmienną losową.*
3. *Jeżeli X i Y są zmiennymi losowymi, to $X + Y$ oraz XY są zmiennymi losowymi.*

Dowód. Zajmiemy się tylko jedną częścią tezy. Aby pokazać, że suma zmiennych losowych jest zmienną losową, wystarczy zauważyć, że warunek

$$a < X(\omega) + Y(\omega) < b$$

jest równoważny warunkowi $\exists\, r_1, r_2, s_1, s_2 \in Q$

$$(r_1 < r_2) \wedge (s_1 < s_2) \wedge (a < r_1 + s_1 < r_2 + s_2 < b) \wedge (r_1 < X(\omega) < r_2) \wedge (s_1 < Y(\omega) < s_2). \quad \Box$$

Dla przestrzeni opisanej w rozdziale 2.2 przykładem zmiennej losowej może być funkcja f, która skończonemu ciągowi α przyporządkowuje długość tego ciągu $\mid \alpha \mid$ powiększoną o 1. Ciągom nieskończonym funkcja f przyporządkowuje ∞. Funkcję tę można interpretować jako liczbą losowań wykonanych po uruchomieniu rozważanego algorytmu. Można zauważyć, że przeciwobrazy ograniczonych podzbiorów R wyznaczone przez tę funkcję są zbiorami skończonymi. Co więcej, elementy γ tych przeciwobrazów są takie, że zbiory $\{\gamma\}$ są zdarzeniami (wynika to bezpośrednio z definicji zbioru zdarzeń). Tak więc interesujące nas przeciwobrazy są skończonymi sumami zdarzeń i, jako takie, są zdarzeniami. Prawdę mówiąc w rozważanej przestrzeni wszystkie zbiory jednoelementowe są zdarzeniami.

3.2 Rozkład zmiennej losowej

Jeżeli analizujemy tylko wyniki pomiaru, a nie interesują nas czynniki wpływające na pomiar, to zamiast rozważać trudną do określenia i skomplikowaną przestrzeń, na której jest określona zmienna losowa, możemy zajmować się przestrzenią wyników pomiarów, zwykle liczb rzeczywistych (przyjęliśmy to w definicji zmiennej losowej), z prawdopodobieństwem zwanym rozkładem zmiennej losowej (słowo rozkład ma tu inne znaczenie niż dotychczas).

Przypuśćmy, że $X : \Omega \to R$ jest zmienną losową określoną w przestrzeni probabilistycznej ze zbiorem zdarzeń $\mathcal{Z}$ i prawdopodobieństwem P. Rozkładem zmiennej X nazywamy prawdopodobieństwo P_X zdefiniowane wzorem

$$P_X(A) = P(X^{-1}(A)) = P(\{\omega \in \Omega : X(\omega) \in A\})$$

i określone na σ-ciele podzbiorów borelowskich zbioru R. Fakt, że P_X jest prawdopodobieństwem jest łatwą konsekwencją definicji i własności przeciwobrazu.

Zwykle łatwiej jest opisać lub przewidzieć i zbadać rozkład zmiennej losowej niż prawdopodobieństwo określone w abstrakcyjnej przestrzeni, na której ta zmienna jest określona. Jeżeli uda się to, to otrzymujemy także jakieś szczególne informacje o tej abstrakcyjnej przestrzeni.

Nietrudno zauważyć, że rozkład prawdopodobieństwa P_f zmiennej f zdefiniowanej w rozdziale 3.1 spełnia następujące równości

$$P_f(\{n+1\}) = p(1-p)^n, \quad P_f(N_+) = 1 \quad \text{oraz} \quad p_f(R \setminus N_+) = 0,$$

gdzie N_+ oznacza zbiór dodatnich liczbn naturalnych.

3.3 Wartość oczekiwana zmiennej losowej

3.3.1 Przypadek przestrzeni przeliczalnych i prostych zmiennych

Jeżeli przestrzeń zdarzeń elementarnych Ω jest przeliczalna (i nieskończona), a X jest zmienną losową określoną na Ω, to wartość oczekiwana zmiennej X jest definiowana wzorem

$$E(X) = \sum_{i=0}^{\infty} X(\omega_i)P(\{\omega_i\}).$$

Analogicznie definiujemy wartość oczekiwaną zmiennej X dla skończonych przestrzeni zdarzeń elementarnych.

Lemat 3.3 *Wartość oczekiwana ma następujące własności:*

1. $E(1) = 1$ *(tutaj pierwsza jedynka oznacza funkcję stale równą 1, druga – liczbę 1),*
2. $E(aX + bY) = aE(X) + bE(Y)$*,*
3. *jeżeli* $x_1, x_2, \ldots, x_n$ *są wszystkimi wartościami przyjmowanymi przez zmienną losową* X*, to*
$$E(X) = \sum_{i=1}^{n} x_i P(\{\omega \in \Omega : X(\omega) = x_i\}), \tag{3}$$
4. *jeżeli* $X(\omega) \leqslant Y(\omega)$ *dla wszystkich* $\omega \in \Omega$*, to* $E(X) \leqslant E(Y)$*.*

Zmienną losową, która przyjmuje tylko skończenie wiele wartości nazywamy (funkcją) prostą. Wartość oczekiwaną prostych zmiennych losowych zawsze definiujemy zgodnie ze wzorem z lematu 3.3, punkt 3.

3.3.2 Przypadek nieujemnych zmiennych losowych

Przypuśćmy, że X jest nieujemną zmienną losową. Dla liczby naturalnej n definiujemy funkcję $X_n : \Omega \to R$ przyjmując

$$X_n(\omega) = \begin{cases} n & \text{jeżeli } X(\omega) \geqslant n \\ \dfrac{k}{2^n} & \text{jeżeli } \dfrac{k}{2^n} \leqslant X(\omega) < \dfrac{k+1}{2^n} \text{ dla pewnego naturalnego } k < n2^n. \end{cases} \tag{4}$$

Nietrudno zauważyć, że X_n jest prostą zmienną losową mniejszą od X $(X_n(\omega) \leqslant X(\omega))$ i przybliżającą w pewnym sensie zmienną X, a mianowicie jeżeli $X(\omega) \leqslant n$, to

$$0 \leqslant X(\omega) - X_n(\omega) \leqslant \frac{1}{2^n}.$$

Stąd wynika, że dla wszystkich $\omega \in \Omega$

$$\lim_n X_n(\omega) = X(\omega),$$

a ponadto wartości $X_n(\omega)$ tworzą ciąg niemalejący. Przyjmujemy, że

$$E(X) = \lim_n E(X_n).$$

Wartość oczekiwana zmiennej losowej X może nie być określona. Ciąg wartości oczekiwanych $E(X_n)$ jest niemalejący. Jeżeli nie jest zbieżny, to jest rozbieżny do nieskończoności. Wtedy przyjmuje się, że $E(X) = \infty$.

3.3.3 Przypadek ogólny

Jeżeli X jest zmienną losową przyjmującą dowolne wartości, to rozbijamy ją na część dodatnią i ujemną:

$$X = X_+ + X_-, \text{ gdzie } X_+(\omega) = \max(X(\omega), 0) \text{ oraz } X_-(\omega) = \min(X(\omega), 0).$$

Przyjmujemy, że

$$E(X) = E(X_+) - E(-X_-).$$

Własności wartości oczekiwanej zmiennej losowej podane w lemacie 3.3 pozostają prawdziwe we wszystkich przypadkach.

3.3.4 Wartość oczekiwana a rozkład zmiennej

Będziemy rozważać przestrzeń probabilistyczną ze zbiorem zdarzeń elementarnych Ω, z σ-ciałem $\mathcal{Z}$, prawdopodobieństwem P i zmienną losową $X : \Omega \to R$. Wartość oczekiwaną zmiennej $Y : \Omega \to R$ będziemy oznaczać $E_1(Y)$ dla odróżnienia od wartości oczekiwanych w innej przestrzeni. Zmienna X pozwala rozważać przestrzeń probabilistyczną złożoną z R, σ-ciała zbiorów borelowskich i rozkładu P_X zmiennej X. W tej przestrzeni wartość oczekiwaną zmiennej $Y : R \to R$ będziemy oznaczać symbolem $E_2(Y)$.

Niech I oznacza funkcję identycznościową określoną na zbiorze R. Funkcję I możemy też uważać za zmienną losową (w przestrzeni z prawdopodobieństwem P_X). Nietrudno zauważyć, że rozkładem zmiennej I jest również P_X.

Lemat 3.4 *Zachodzi wzór $E_1(X) = E_2(I)$.*

Dowód. Aby dowieść podany wzór wystarczy cierpliwie porównać definicje wartości oczekiwanych w obu przypadkach. □

Wniosek 3.5 *Zmienne losowe o tym samym rozkładzie mają tę samą wartość oczekiwaną.* □

Poprzedni lemat można znacznie wzmocnić. Można między innymi dowieść, że

Lemat 3.6 *Jeżeli $F : R \to R$ jest funckją ciągłą, to $E_1(F \circ X) = E_2(F)$.* □

Na ogół nie stosuje się indeksów do rozróżnienia różnych wartości oczekiwanych. Zwykle można domyślić się jak wartość oczekiwana jest rozumiana, a jeżeli jest to trudne, to korzysta się z symboliki stosowane w ogólnej teorii całki. Wtedy podany wzór mógłby mieć postać

$$E_1(F \circ X) = \int_\Omega F(X(\omega))dP(\omega) = \int_R F(x)dP_X(x) = E_2(F).$$

Wyliczmy jeszcze wartość oczekiwaną zmiennej losowej f zdefiniowanej w rozdziale 3.1. Wartość oczekiwaną f powinniśmy liczyć zgodnie ze wzorem (4). Nietrudno zauważyć, że dla funkcji f o wartościach naturalnych ten wzór definiuje funkcję $f_n(\omega) = \min\{f(\omega), n\}$. Tak więc $E(f) = \lim_{n\to\infty} E(\min(f, n))$. Wartość $E(\min(f, n))$ liczymy zgodnie ze wzorem (3). Zauważmy, że

$$P(\{\omega : f(\omega) = n\}) = P(\{\omega : |\omega| = n-1\}) = \sum_{|\,\omega\,|=n-1} P(\{\omega\}) = 2^{n-1}p\frac{(1-p)^{n-1}}{2^{n-1}} = p(1-p)^{n-1}$$

oraz

$$P(\{\omega : f(\omega) \geqslant n\}) = \sum_{i=n}^{\infty} P(\{\omega : f(\omega) = i\}) = \sum_{i=n}^{\infty} p(1-p)^{i-1} = p(1-p)^{n-1}\sum_{i=0}^{\infty}(1-p)^i = (1-p)^{n-1}.$$

Stąd

$$E(\min(f, n)) = \sum_{i=1}^{n-1} ip(1-p)^{i-1} + n(1-p)^{n-1}$$

oraz

$$E(f) = \lim_{n\to\infty} E(\min(f, n)) = \sum_{i=1}^{\infty} ip(1-p)^{i-1} + 0 = p\frac{1}{(1-(1-p))^2} = \frac{1}{p}.$$

Tak więc dla $p = \frac{1}{3}$, należy spodziewać się, że rozważany algorytm wykona zawartą w nim pętlę 3 razy.

3.4 Nierówność Czebyszewa

Lemat 3.7 (Nierówność Czebyszewa (1)) *Jeżeli X jest nieujemną zmienną losową i $t > 0$, to*

$$P(\{\omega \in \Omega : X(\omega) \geqslant t\}) \leqslant \frac{E(X)}{t}.$$

Dowód. Niech $A = \{\omega \in \Omega : X(\omega) \geqslant t\}$. Zdefiniujmy zmienną losową Y przyjmując, że

$$Y(\omega) = \begin{cases} t & \text{jeżeli } \omega \in A \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Oczywiście, $Y(\omega) \leqslant X(\omega)$ oraz

$$t \cdot P(A) = E(Y) \leqslant E(X). \ \square$$

3.5 Wariancja zmiennej losowej

Wariancją zmiennej losowej X nazywamy wartość oczekiwaną zmiennej losowej Y takiej, że $Y(\omega) = (X - E(X))^2$. Wariancją zmiennej losowej X będziemy oznaczać symbolem $Var(X)$. Mamy więc $Var(X) = E((X - E(X))^2)$.

Z rezultatów zamieszczonych w rozdziale 3.3.4 wynika, że wariancja zmiennej (podobnie jak wartość oczekiwana) jest wyznaczona przez rozkład zmiennej. Zmienne o tym samym rozkładzie mają taką samą wariancję.

Lemat 3.8 *1.* $Var(X) = E(X^2) - (E(X))^2$,

2. $Var(aX + b) = a^2 Var(X)$.

3. Jeżeli $a \leqslant X(\omega) \leqslant b$, *to* $Var(X) \leqslant \frac{(b-a)^2}{4}$.

Lemat 3.9 (Nierówność Czebyszewa (2)) *Jeżeli X jest zmienną losową i $t > 0$, to*

$$P(\{\omega \in \Omega : |\, X(\omega) - E(X)\, | \geqslant t\}) \leqslant \frac{Var(X)}{t^2}.$$

Dowód. Jest to konsekwencja nierówności Czebyszewa z lematu 3.7. Wystarczy zauważyć, że

$$\{\omega \in \Omega : |\, X(\omega) - E(X)\, | \geqslant t\} = \{\omega \in \Omega : (X(\omega) - E(X))^2 \geqslant t^2\},$$

a wartość oczekiwana zmiennej $Y(\omega) = (X(\omega) - E(X))^2$ jest niczym innym jak wariancją zmiennej X. $\Box$

4 Prawdopodobieństwo warunkowe, zdarzenia niezależne i iloczyn kartezjański

4.1 Prawdopodobieństwo warunkowe

Jeżeli obserwujemy, jak dużo jest ludzi wysokich (np. o wzroście przekraczającym 180 cm), to po przeprowadzeniu stosownych doświadczeń losowych będziemy mogli też wiedzieć, jak wiele jest wysokich kobiet i jak dużo jest wysokich mężczyzn. Tego typu problemy prowadzą do pojęcia prawdopodobieństwa warunkowego.

Przypuśćmy, że mamy przestrzeń probabilistyczną (a więc mamy Ω, $\mathcal{Z}$ i P) oraz zdarzenie $A \in \mathcal{Z}$ takie, że $P(A) > 0$. Wtedy możemy skonstruować inną przestrzeń probabilistyczną. Zbiorem zdarzeń elementarnych tej przestrzeni będzie znowu zbiór Ω, zbiorem zdarzeń – rodzina $\mathcal{Z}$. Prawdopodobieństwem w tej przestrzeni jest funkcja zdefiniowana wzorem

$$P(X|A) = \frac{P(X \cap A)}{P(A)}.$$

Łatwo przekonać się, że zdefiniowana w ten sposób przestrzeń rzeczywiście ma własności wymagane od przestrzeni probabilistycznej.

Zauważmy, że $P(A|A) = 1$ oraz $P(\Omega \setminus A|A) = 0$. Wartość $P(X|A)$ nazywa się prawdopodobieństwem warunkowym zdarzenia X pod warunkiem zajścia zdarzenia A. Wartość tą zwykle interpretuje się jako prawdopodobieństwo zajścia zdarzenia X w sytuacji, gdy w jakiś sposób wymuszamy zajście zdarzenia A, np. świadomie pomijamy wyniki takich doświadczeń losowych, w których nie zaszło zdarzenie A.

Prawdopodobieństwo warunkowe $P(\cdot|A)$ można też rozważać w przestrzeni probabilistycznej złożonej ze zbioru zdarzeń elementarnych A i zbioru zdarzeń $\mathcal{Z}_A = \{X \subseteq \Omega : X \subseteq A\}$.

4.2 Zdarzenia niezależne

Zdarzenia A i B są niezależne, jeżeli zajście zdarzenia B nie zależy od tego, czy w jakiś sposób wymusiliśmy zajście zdarzenia A, czy też nie zrobiliśmy tego. Tak więc formalną definicję zdarzeń niezależnych można wyrazić wzorem

$$P(B) = P(B|A).$$

Obowiązująca definicja jest jednak inna. Zwykle przyjmuje się, że zdarzenia A i P są niezależne, jeżeli

$$P(A \cap B) = P(A) \cdot P(B).$$

Drobna różnica między tymi definicjami ma miejsce w sytuacjach, gdy $P(A) = 0$.

4.3 Iloczyn kartezjański przestrzeni probabilistycznych

Niech Ω_1 i Ω_2 będą zbiorami zdarzeń elementarnych dwóch przestrzeni probabilistycznych ze zbiorami zdarzeń $\mathcal{Z}_1$ i $\mathcal{Z}_2$ oraz prawdopodobieństwami P_1 i P_2 odpowiednio. Zdefiniujemy przestrzeń probabilistyczną, w której zbiorem zdarzeń elementarnych jest $\Omega_1 \times \Omega_2$.

Łatwo zdefiniować zbiór zdarzeń tej przestrzeni. Jest to najmniejsze σ-ciało $\mathcal{Z}$ podzbiorów $\Omega_1 \times \Omega_2$ zawierające

$$\{A \times B : A \in \mathcal{Z}_1 \wedge B \in \mathcal{Z}_2\}.$$

Trudniej zdefiniować prawdopodobieństwo. Chcemy jednak, aby to prawdopodobieństwo P spełniało równości

$$P(A \times B) = P_1(A) \cdot P_2(B) \tag{5}$$

dla wszystkich $A \in \mathcal{Z}_1$ i $B \in \mathcal{Z}_2$. Pokażemy w rozdziale 4.5, że warunek (5) jednoznacznie definiuje pewne prawdopodobieństwo na $\mathcal{Z}$.

4.4 Iloczyn kartezjański a niezależność

Przypuśćmy, że mamy przestrzeń probabilistyczną złożoną z Ω, $\mathcal{Z}$ i P opisującą pewne doświadczenie losowe i mamy dwa zdarzenia losowe A i B (zależne lub nie). Chcemy dwukrotnie powtórzyć doświadczenie losowe tak, aby wyniki pierwszego i drugiego doświadczenia były niezależne, a w szczególności, aby zajście zdarzenia A podczas pierwszego doświadczenia było niezależne od zajścia zdarzenia B podczas drugiego doświadczenia.

Zauważmy, że iloczyn kartezjański przestrzeni pozwala na sformalizowanie tego zagadnienia. Wynikiem takiego podwójnego doświadczenia będą pary złożone z wyników pierwszego i drugiega doświadczenia, a więc zbiorem zdarzeń elementarnych będzie iloczyn $\Omega \times \Omega$. Zajściu zdarzenia A podczas pierwszego doświadczenia odpowiada zbiór $A \times \Omega$, a zbiór $\Omega \times B$ oznacza zajście zdarzenia B podczas drugiego doświadczenia. Zauważmy, że w iloczynie (kwadracie) kartezjańskim rozważanej przestrzeni zachodzą równości

$$P((A \times \Omega) \cap (\Omega \times B)) = P(A \times B) = P(A) \cdot P(B) = P(A \times \Omega) \cdot P(\Omega \times B).$$

Oczywiście, konstrukcję iloczynu kartezjańskiego możemy rozszerzyć na przypadek dowolnej skończonej liczby przestrzeni (a nawet nieskończonej) i w ten sposób definiujemy przestrzenie probabilistyczne służące do opisu serii niezależnych doświadczeń losowych.

4.5 Poprawność definicji iloczynu kartezjańskiego

Niech $\mathcal{Z}_0$ będzie najmniejszym ciałem zbiorów zawierającym

$$\{A \times B : A \in \mathcal{Z}_1 \wedge B \in \mathcal{Z}_2\}.$$

Jest oczywiste, że $\mathcal{Z}$ jest najmniejszym σ-ciałem zawierającym $\mathcal{Z}_0$. Łatwo też można scharakteryzować $\mathcal{Z}_0$. Mamy bowiem

$$\mathcal{Z}_0 = \{\bigcup_{i=1}^{n} A_i \times B_i : n \in N \wedge \forall i, j \leqslant n\ (A_i \in \mathcal{Z}_1 \wedge B_i \in \mathcal{Z}_2 \wedge (i \neq i \Rightarrow A_i \cap A_j = \emptyset))\}.$$

Korzystając z tej charakteryzacji bez trudu pokazujemy, że istnieje dokładnie jedna funkcja addytywna P określona na $\mathcal{Z}_0$ i spełniająca warunek (5).

Pokażemy teraz, że funkcja P jest σ-addytywna na $\mathcal{Z}_0$. Razem z twierdzeniem 7.1 o rozszerzaniu pozwala to zdefiniować na $\mathcal{Z}$ prawdopodobieństwo o własności (5).

Lemat 4.1 *Funkcja P jest σ-addytywna na $\mathcal{Z}_0$.*

Dowód. Wystarczy wykazać, że jeżeli mamy wstępujący ciąg zbiorów

$$X_n = \bigcup_{k=0}^{m_n} A_{n,k} \times B_{n,k} \in \mathcal{Z}_0$$

taki, że $A_{n,k_1} \cap A_{n,k_2} = \emptyset$ dla dowolnego n i dowolnych różnych liczb $k_1, k_2 \leqslant m_n$ oraz zachodzi wzór

$$\bigcup_{n=0}^{\infty} \bigcup_{k=0}^{m_n} A_{n,k} \times B_{n,k} = \Omega_1 \times \Omega_2, \tag{6}$$

to

$$\lim_{n\to\infty} P(\bigcup_{k=0}^{m_n} A_{n,k} \times B_{n,k}) = \lim_{n\to\infty} \sum_{k=0}^{m_n} P_1(A_{n,k}) P_2(B_{n,k}) = 1.$$

Ustalmy na chwilę $\omega \in \Omega_1$ i przyjmijmy, że

$$B_n(\omega) = \begin{cases} B_{n,k} & \text{jeżeli } \omega \in A_{n,k} \\ \emptyset & \text{jeżeli } \omega \notin \bigcup_{k=0}^{m_n} A_{n,k}. \end{cases}$$

Nietrudno zauważyć, że ciąg $B_0(\omega), B_1(\omega), B_2(\omega), \ldots$ jest wstępujący. Pokażemy, że jego suma jest równa Ω_2. Dla dowolnego $\omega' \in \Omega_2$ mamy $(\omega, \omega') \in \Omega_1 \times \Omega_2$. Na podstawie równości

(6) istnieją liczby n i $k \leqslant m_n$ takie, że $(\omega, \omega') \in A_{n,k} \times B_{n,k}$. Stąd wynika, że $\omega' \in B_n(\omega)$. Pokazaliśmy więc, że

$$\bigcup_{n=0}^{\infty} B_n(\omega) = \Omega_2.$$

Ponieważ P_2 jest miarą, więc także

$$\lim_{n \to \infty} P_2(B_n(\omega)) = 1.$$

Weźmy dowolną liczbę $\varepsilon \in (0,1)$ i przyjmijmy, że

$$A_n = \{\omega \in \Omega_1 : P_2(B_n(\omega)) \geqslant 1 - \varepsilon\}.$$

Nietrudno zauważyć, że $A_0, A_1, A_2, \ldots$ jest wstępującym ciągiem zbiorów takim, że

$$\bigcup_{n=0}^{\infty} A_n = \Omega_1.$$

Wobec tego,

$$\lim_{n \to \infty} P_1(A_n) = 1.$$

Dalej będą nam potrzebne dwie równości dotyczące zbiorów A_n:

$$A_n = \bigcup_{k : A_n \cap A_{n,k} \neq \emptyset} A_{n,k} \quad \text{oraz} \quad B_n(\omega) = B_{n,k}$$

dla wszystkich $\omega \in A_n \cap A_{n,k}$.

Weźmy liczbę n, która spełnia nierówność $P_1(A_n) \geqslant 1 - \varepsilon$. Oszacujemy teraz

$$P(\bigcup_{k=0}^{m_n} A_{n,k} \times B_{n,k}) = \sum_{k=0}^{m_n} P_1(A_{n,k}) P_2(B_{n,k}).$$

Mamy

$$\sum_{k=0}^{m_n} P_1(A_{n,k})P_2(B_{n,k}) \geqslant \sum_{k : A_n \cap A_{n,k} \neq \emptyset} P_1(A_{n,k})P_2(B_{n,k}) \geqslant (1-\varepsilon) \sum_{k : A_n \cap A_{n,k} \neq \emptyset} P_1(A_{n,k}) =$$
$$= (1-\varepsilon) P_1(\bigcup_{k : A_n \cap A_{n,k} \neq \emptyset} A_{n,k}) = (1-\varepsilon)P_1(A_n) \geqslant (1-\varepsilon)^2 \geqslant 1 - 2\varepsilon.$$

Z dowiedzionego oszacowania wynika teza. □

4.6 Prawo wielkich liczb

Przypuśćmy, że mamy przestrzeń probabilistyczną ze zbiorem zdarzeń elementarnych Ω, zbiorem zdarzeń $\mathcal{Z}$ i prawdopodobieństwem P. Niech $S \in \mathcal{Z}$ będzie ustalonym zdarzeniem. Będziemy mówić o sukcesie, jeżeli został wylosowany element zbioru S. Przyjmujemy też, że ma to miejsce z prawdopodobieństwem p (czyli $P(S) = p$), a $q = 1 - p$.

Doświadczenie losowe opisane przez tę przestrzeń powtarzamy w sposób niezależny pewną liczbę razy. Można powiedzieć, że rozważamy iloczyny kartezjańskie Ω^n z prawdopodobieństwem P_n. Niech X_n oznacza liczbę sukcesów w n niezależnych doświadczeniach. Tak więc $X_n : \Omega^n \to R$ oraz $X_n(x_1, \ldots, x_n) = |\{i : x_i \in S\}|$.

Przyjmijmy, że $S^1 = S^c$ i $S^0 = S$. Posługując się tą umową możemy rozważać zbiory postaci

$$S(\varepsilon_1, \ldots, \varepsilon_n) = S^{\varepsilon_1} \times \ldots \times S^{\varepsilon_n}$$

dla dowolnych ciągów $\varepsilon_1, \ldots, \varepsilon_n$ złożonych z 0 i 1. Bez trudu, z definicji P_n wyprowadzamy wzór

$$P_n(S^{\varepsilon_1} \times \ldots \times S^{\varepsilon_n}) = p^k(1-p)^{n-k} = p^k q^{n-k},$$

gdzie k oznacza liczbę wyrazów ciągu $\varepsilon_1, \ldots, \varepsilon_n$ równych 0. Jest też oczywisty wzór

$$\{\omega \in \Omega_n : X_n(\omega) = k\} = \bigcup S^{\varepsilon_1} \times \ldots \times S^{\varepsilon_n},$$

w którym sumujemy wszystkie iloczyny podanej postaci wyznaczone przez ciągi zawierające k zer. Ponieważ sumowane zbiory są parami rozłączne, więc rozkład zmiennej X_n spełnia równości

$$P_n(\{\omega \in \Omega_n : X_n(\omega) = k\}) = \binom{n}{k} p^k q^{n-k}.$$

Znajomość rozkładu pozwala na wyliczenie wartość oczekiwanej i wariancji zmiennej X_n .

Lemat 4.2 *Wartość oczekiwana zmiennej X_n jest dana wzorem $E(X_n) = np$, a wariancja $Var(X_n) = np(1-p)$.*

Dowód. Zauważmy, że

$$E(X_n) = \sum_{k=0}^{n} k \cdot P_n(\{\omega : X_n(\omega) = k\}) = \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k q^{n-k} = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k} =$$
$$= np(p+q)^{n-1} = np.$$

Przed wyliczeniem wariancji obliczymy wartość oczekiwaną kwadratu X_n. Zauważmy, że

$$E(X_n^2) = \sum_{k=1}^{n} k^2 \cdot \binom{n}{k} p^k q^{n-k} = np \sum_{k=1}^{n} k \cdot \binom{n-1}{k-1} p^{k-1} q^{n-k} =$$
$$= np \sum_{k=0}^{n-1} (k+1) \cdot \binom{n-1}{k} p^k q^{n-1-k} = np \left(\sum_{k=0}^{n-1} k \cdot \binom{n-1}{k} p^k q^{n-1-k} + \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} \right) =$$
$$= np \left((n-1)p \sum_{k=1}^{n-1} \binom{n-2}{k-1} p^{k-1} q^{n-1-k} + (p+q)^{n-1} \right) = np\left((n-1)p(p+q)^{n-2} + 1\right) =$$
$$= np(np+q).$$

Teraz łatwo podać wzór na wariancję:

$$Var(X_n) = E(X_n^2) - E^2(X_n) = np(np+q) - (np)^2 = npq. \ \square$$

Twierdzenie 4.3 (Prawo wielkich liczb) *W opisane sytuacji*

$$\lim_{n\to\infty} P(\{\omega \in \Omega^n : | \frac{X_n(\omega)}{n} - p \,| \geqslant \varepsilon\}) = 0.$$

Dowód. Prawo wielkich liczb wynika z nierówności Czebyszewa 3.9 i lematu 4.2. Zauważmy, że

$$
\begin{aligned}
&P(\{\omega \in \Omega^n : \mid \frac{X_n(\omega)}{n} - p \mid \geqslant \varepsilon\}) = P(\{\omega \in \Omega^n : \mid X_n(\omega) - np \mid \geqslant n\varepsilon\}) = \\
&= P(\{\omega \in \Omega^n : \mid X_n(\omega) - E(X_n) \mid \geqslant n\varepsilon\}) \leqslant \frac{Var(X_n)}{n^2\varepsilon^2} = \frac{npq}{n^2\varepsilon^2}.
\end{aligned}
$$

Teza oczywiście wynika z udowodnionej nierówności. □

Prawo wielkich liczb daje nadzieję na obliczenie nieznanego prawdopodobieństwa $p = P(S)$. W sposób oczywisty wynika z niego, że także

$$\lim_{n\to\infty} P(\{\omega \in \Omega^n : \mid \frac{X_n(\omega)}{n} - p \mid < \varepsilon\}) = 1,$$

a więc poza nieprawdopodobnymi sytuacjami, praktycznie zawsze, jeżeli n-krotnie (dla dostatecznie dużej liczby n) sprawdzimy, czy wylosowany element należy do S, to $P(S)$ różni się mało od ilorazu liczby wylosowanych elementów S przez n.

5 Twierdzenie de Moivre'a - Laplace'a

5.1 Uzasadnienie

Przypuśćmy, że mamy przestrzeń probabilistyczną i zmienną losową X. Możemy myśleć, że zmienna X zwraca wynik losowego pomiaru jakieś wielkości. O wartości oczekiwanej $E(X)$ możemy myśleć jako o rzeczywistej wartości mierzonej mierzonej wielkości. Różnica $X - E(X)$ jest wtedy błędem pomiaru. Natomiast pierwiastek wariancji $\sqrt{Var(X)}$ w pewnym sensie odpowiada średniemu błędowi popełnianemu podczas pomiaru. Tak więc iloraz

$$\frac{X - E(X)}{\sqrt{Var(X)}}$$

odpowiada pojęciu błędu względnego i opisuje jak duży jest popełniony błąd w stosunku do błędu średniego. Będzie nas interesować w pewnym przypadku prawdopodobieństwo tego, że błąd względny przyjmuje wartość z danego przedziału, czyli

$$P(\{\omega \in \Omega : A \leqslant \frac{X(\omega) - E(X)}{\sqrt{Var(X)}} < B\}).$$

Mówiąc inaczej, interesuje nas rozkład błędu względnego.

Przypuśćmy, że mamy ciąg niezależnych zmiennych losowych $X_1, X_2, X_3, \ldots$. Przyjmijmy, że $S_n = \sum_{i=1}^n X_n$ oraz, że jest spełniony pewien warunek, często nazywany warunkiem Lapunowa. Dowodzi się, że

$$\lim_{n\to\infty} P(\{\omega \in \Omega : A \leqslant \frac{S_n(\omega) - E(S_n)}{\sqrt{Var(S_n)}} < B\}) = \frac{1}{\sqrt{2\pi}} \int_A^B e^{-\frac{z^2}{2}}\, dz.$$

Wprowadźmy oznaczenie dla całki z powyższego wzoru. Niech

$$N_{m,s}([A,B)) = \frac{1}{\sqrt{2\pi s}} \int_A^B e^{-\frac{(z-m)^2}{2s}}\, dz.$$

Wzór ten możemy uważać za definicję pewnego prawdopodobieństwa, określonego poprzez wskazanie gęstości. To prawdopodobieństwo nazywamy rozkładem normalnym z wartością oczekiwaną m i wariancją s , i rzeczywiście ma te własności.

Przytoczone twierdzenie można wzmocnić zastępując tezę następującą:

$$\lim_{n\to\infty} P(\{\omega \in \Omega : \frac{S_n(\omega) - E(S_n)}{\sqrt{Var(S_n)}} \in Z\}) = N_{0,1}(Z)$$

dla dowolnego zbioru borelowskiego Z. Co więcej, granica ta jest jednostajna ze względu na Z, a więc jeżeli przyjmiemy pewną dokładność, to dla dostatecznie dużych n prawdopodobieństwa

$$P(\{\omega \in \Omega : \frac{S_n(\omega) - E(S_n)}{\sqrt{Var(S_n)}} \in Z\}) \text{ oraz } N_{0,1}(Z)$$

są w przybliżeniu równe z zadaną dokładnością dla wszystkich zbiorów borelowskich Z (patrz twierdzenie 8.7).

Korzystając z powyższego twierdzenia można też w przybliżeniu określić rozkład zmiennych S_n. Można wykazać, że

$$P(\{\omega \in \Omega : S_n(\omega) \in Z\}) \approx N_{E(S_n),Var(S_n)}(Z)$$

w takim sensie, jak wyżej (patrz wniosek 8.8).

Wszystko to wskazuje na rolę rozkładu normalnego i zdaje się wyjaśniać, dlaczego obserwując zjawiska losowe często napotykamy na ten rozkład. Na przykład uważa się, że rozkład błędów pomiaru losowego jest rozkładem normalnym.

Dowody podanych twierdzeń są dość skomplikowane. Dalej naszkicujemy je dla pewnych, konkretnych zmiennych.

5.2 Sformułowanie twierdzenia

Przypuśćmy, że mamy daną przestrzeń probabilistyczną i zdarzenie Q. Sukcesem będziemy nazywać zajście zdarzenia Q. Przyjmujemy, że uzyskujemy sukces z prawdopodobieństwem p, z zdarzenie przeciwne (porażka) ma miejsce z prawdopodobieństwem $1 - p = q$. Będziemy wielokrotnie i w sposób niezależny powtarzać doświadczenie losowe opisane przez daną przestrzeń. Niech X_n oznacza zmienną losową, która n-krotnej serii doświadczeń losowych przyporządkowuje liczbę sukcesów. Zauważmy, że zmienna X_n jest sumą n zmiennych zero-jedynkowych przyjmujących wartość 1 wtedy i tylko wtedy, gdy w doświadczeniu losowym odnieśliśmy sukces.

Twierdzenie 5.1 (de Moivre'a-Laplace'a) *Niech A i B będą dowolnymi liczbami takimi, że $A < B$. Wtedy*

$$\lim_{n\to\infty} P_n(\{\omega \in \Omega_n : A \leqslant \frac{X_n(\omega) - E(X_n)}{\sqrt{Var(X_n)}} \leqslant B\}) = \frac{1}{\sqrt{2\pi}} \int_A^B e^{-\frac{z^2}{2}}\, dz.$$

W rozdziale 4.6 została wyliczona wartość oczekiwana i wariancja zmiennej X_n. Korzystając z wyników tych obliczeń wzorowi z twierdzenia de Moivre'a - Laplace'a można nadać postać

$$\lim_{n\to\infty} P_n(\{\omega \in \Omega_n : A \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} < B\}) = \frac{1}{\sqrt{2\pi}} \int_A^B e^{-\frac{z^2}{2}} dz.$$

Dowód twierdzenia został zamieszczony w jednym z dodatków. Dalej przedstawimy jedynie plan dowodu.

5.3 Plan dowodu

Przypuśćmy, że zmienna X przyjmuje tylko wartości naturalne z przedziału od 0 do n. Wtedy

$$\{\omega \in \Omega : A \leqslant \frac{X(\omega) - E(X)}{\sqrt{Var(X)}} \leqslant B\} = \bigcup_{k=0}^{n} \{\omega \in \Omega : A \leqslant \frac{k - E(X)}{\sqrt{Var(X)}} \leqslant B \wedge X(\omega) = k\}.$$

Oznaczmy symbolem W zbiór

$$\{k \in N : k \leqslant n \wedge A \leqslant \frac{k - E(X)}{\sqrt{Var(X)}} \leqslant B\}.$$

Nietrudno zauważyć, że

$$\{\omega \in \Omega : A \leqslant \frac{X(\omega) - E(X)}{\sqrt{Var(X)}} \leqslant B\} = \bigcup_{k \in W} \{\omega \in \Omega : X(\omega) = k\}.$$

oraz

$$P(\{\omega \in \Omega : A \leqslant \frac{X(\omega) - E(X)}{\sqrt{Var(X)}} \leqslant B\}) = \sum_{k \in W} P(\{\omega \in \Omega : X(\omega) = k\}).$$

Aby wyliczyć podaną sumę, skorzystamy z pojęcia całki Riemanna. Przyjmijmy, że

$$z_k = \frac{k - E(X)}{\sqrt{Var(X)}}.$$

Mamy więc ciąg liczb rzeczywistych $z_0, z_1, \ldots, z_n$, każde dwie kolejne z tych liczb są odległe od siebie o $\frac{1}{\sqrt{Var(X)}}$. Może się zdarzyć, że $z_0 \leqslant A \leqslant B \leqslant z_n$. Wtedy fragment ciągu $z_0, z_1, \ldots, z_n$ o indeksach $k \in W$ jest podziałem odcinka $[A, B]$. Przypuśćmy, że uda się nam znaleźć funkcję f taką, że

$$P(\{\omega \in \Omega : X(\omega) = k\}) \cdot \sqrt{Var(X)} = f(z_k).$$

Wtedy

$$\sum_{k \in W} P(\{\omega \in \Omega : X(\omega) = k\}) = \sum_{k \in W} f(z_k) \cdot \frac{1}{\sqrt{Var(X)}}$$

jest sumą przybliżającą całkę Riemanna z funkcji f i można twierdzić, że

$$P(\{\omega \in \Omega : A \leqslant \frac{X(\omega) - E(X)}{\sqrt{Var(X)}} \leqslant B\}) \approx \int_A^B f(z) dz.$$

Poszukiwanie funkcji f o podanych własnościach wymaga odwołania się do analizy matematycznej. Współczynniki Newtona wyraża się przez silnie, które z kolei wyraża się za pomocą funkcji wykładniczej korzystając z wzoru Stirlinga. Pojawiające się logarytmy przybliża się korzystając rozwinięcia Taylora. To wszystko razem daje efekt w postaci twierdzenia de Moivre'a - Laplace'a.

6 Dodatek: Uwaga o sumowaniu wyrazów ciągu nieskończonego

W twierdzeniu o przestawianiu wyrazów szeregu założenie o dodatniości wyrazów ciągu (lub słabsze założenie o bezwzględnej zbieżności) jest potrzebne. Dowodzi tego następujący przykład.

Z twierdzenia o szeregach naprzemiennych wynika zbieżność szeregu

$$\sum_{i=0}^{\infty}(-1)^i\frac{1}{i+1} = 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\dots$$

Sumę tego szeregu oznaczamy symbolem c. Znajdując rozwinięcie funkcji $\log(1+x)$ w szereg potęgowy pokazuje się, że $c = \log 2$. Zauważmy, że

$$c = (1-\frac{1}{2})+\frac{1}{3}-\frac{1}{4}+(\frac{1}{5}-\frac{1}{6})+\frac{1}{7}-\frac{1}{8}+(\frac{1}{9}-\frac{1}{10})+\frac{1}{11}-\frac{1}{12}\dots$$

oraz

$$\frac{c}{2} = \frac{1}{2} + 0 - \frac{1}{4} + \frac{1}{6} + 0 - \frac{1}{8} + \frac{1}{10} + 0 - \frac{1}{12} + \dots$$

Sumując dwa ostatnie ciągi otrzymujemy, że

$$\frac{3c}{2} = 1+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\frac{1}{9}+\frac{1}{11}-\frac{1}{6}\dots$$

Oznacza to, że sumując wyrazy ciągu

$$1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \frac{1}{5}, \dots$$

w kolejności, w jakiej są wymienione, otrzymujemy c, a sumując na przemian dwa kolejne dodatnie i jeden ujemny wyraz tego ciągu otrzymujemy $\frac{3}{2}c$.

7 Dodatek: Twierdzenie o rozszerzaniu

Twierdzenie 7.1 *Każda nieujemna σ-addytywna funkcja zbiorów określona na ciele $\mathcal{C}$ podzbiorów Ω daje się jednoznacznie rozszerzyć na najmniejsze σ-ciało $\mathcal{B}$ podzbiorów Ω zawierające $\mathcal{C}$.*

Dowód. Niech m będzie nieujemną σ-addytywną funkcją zbiorów o wartościach rzeczywistych, określoną na ciele $\mathcal{C}$ podzbiorów Ω. Zdefiniujmy funkcję $\mu : \mathcal{P}(\Omega) \to R$ przyjmując, że

$$\mu(A) = \inf\{\lim_{i\to\infty} m(A_i) : \forall i\ (A_i \in \mathcal{C} \quad A_i \subseteq A_{i+1}) \wedge A \subseteq \bigcup_{i=1}^{\infty} A_i\}.$$

Funkcja μ nazywa się miarą zewnętrzną Caratheodory'ego i ma następujące własności:

1. $0 \leqslant \mu(A) \leqslant m(\Omega)$ dla wszystkich $A \subseteq \Omega$.
2. μ jest monotoniczna, a więc jeżeli $A \subseteq B$, to $\mu(A) \leqslant \mu(B)$.
3. $\mu(A \cup B) \leqslant \mu(A) + \mu(B)$ dla dowolnych $A, B \subseteq \Omega$.
4. Dla dowolnego ciągu wstępującego $A_0, A_1, A_2, \ldots$ zachodzi nierówność
$$\lim_i \mu(A_i) \leqslant \mu(\bigcup_{i=0}^{\infty} A_i).$$
5. Dla dowolnego ciągu zstępującego $A_0, A_1, A_2, \ldots$ zachodzi nierówność
$$\lim_i \mu(A_i) \geqslant \mu(\bigcap_{i=0}^{\infty} A_i).$$

Wszystkie wymienione wyżej własności μ są oczywiste. Zachodzi też następujący fakt.

Fakt 7.2 Dla dowolnego ciągu wstępującego $A_0, A_1, A_2, \ldots$ zachodzi też równość
$$\lim_n \ \mu(A_n) = \mu(\bigcup_{n=0}^{\infty} A_n).$$

Dowód. Będziemy obliczać $\mu(\bigcup_{n=0}^{\infty} A_n)$. W tym celu weźmy więc $\varepsilon > 0$ i – dla każdego n – wstępujący ciąg $E_{n,0}, E_{n,1}, E_{n,2}, \ldots$ elementów $\mathcal{C}$ dobrze przybliżający $\mu(A_n)$, a dokładniej – spełniający warunki
$$A_n \subseteq \bigcup_{i=0}^{\infty} E_{n,i} \text{ oraz } \mu(A_n) \leqslant \lim_i \ m(E_{n,i}) \leqslant \mu(A_n) + \frac{\varepsilon}{2^{n+1}}.$$

Przyjmijmy, że
$$F_{n,i} = \bigcup_{k=0}^{n} E_{k,i}.$$
Oczywiście, $F_{n,0}, F_{n,1}, F_{n,2}, \ldots$ jest dla dowolnego $n \in N$ wstępującym ciągiem elementów $\mathcal{C}$ takim, że $E_{n,i} \subseteq F_{n,i}$. Będziemy szacować z góry granicę $\lim_i \ m(F_{n,i})$. Zauważmy, że
$$\lim_i \ m(E_{n,i} \cup E_{n-1,i}) = \lim_i \ m(E_{n,i}) + \lim_i \ m(E_{n-1,i}) - \lim_i \ m(E_{n,i} \cap E_{n-1,i}) \leqslant$$
$$\leqslant (\mu(A_n) + \frac{\varepsilon}{2^{n+1}}) + (\mu(A_{n-1}) + \frac{\varepsilon}{2^n}) - \mu(A_{n-1}) = \mu(A_n) + \frac{\varepsilon}{2^{n+1}} + \frac{\varepsilon}{2^n}.$$

Powtarzając podane rozumowanie wielokrotnie otrzymujemy, że
$$\lim_i \ m(F_{n,i}) \leqslant \mu(A_n) + \frac{\varepsilon}{2^{n+1}} + \frac{\varepsilon}{2^n} + \ldots + \frac{\varepsilon}{2} = \mu(A_n) + \frac{(2^{n+1}-1)\varepsilon}{2^{n+1}}.$$

Jeżeli teraz przejdziemy do granicy po n, to otrzymamy
$$\lim_n \ \lim_i \ m(F_{n,i}) \leqslant \lim_n \ \mu(A_n) + \varepsilon.$$

Ciąg $m(F_{n,i})$ jest niemalejący względem każdej zmiennej. Stąd wynika, że

$$\lim_n \lim_i m(F_{n,i}) = \lim_n m(F_{n,n}).$$

Zauważmy także, że

$$\bigcup_{n=0}^{\infty} A_n \subseteq \bigcup_{n=0}^{\infty}\bigcup_{i=0}^{\infty} E_{n,i} = \bigcup_{n=0}^{\infty} F_{n,n}.$$

Łącząc przytoczone wyżej fakty otrzymujemy, że

$$\mu(\bigcup_{n=0}^{\infty} A_n) \leqslant \lim_n m(F_{n,n}) \leqslant \lim_n \mu(A_n) + \varepsilon.$$

Jeżeli teraz skorzystamy z dowolności ε, to otrzymamy nierówność

$$\mu(\bigcup_{n=0}^{\infty} A_n) \leqslant \lim_n \mu(A_n).$$

Przeciwna nierówność wynika z własności 4). $\square$

Fakt 7.3 $\mu(A) = m(A)$ dla wszystkich $A \in \mathcal{C}$.

Dowód. Przypuśćmy, że $A \in \mathcal{C}$. Jest oczywiste, że $\mu(A) \leqslant m(A)$. Aby dowieść nierówność przeciwną weźmy wstępujący ciąg $A_0, A_1, A_2, \ldots$ elementów z $\mathcal{C}$ taki, że $A \subseteq \bigcup_{i=0}^{\infty} A_i$. Ciąg $A \cup A_0, A \cup A_1, A \cup A_2, \ldots$ elementów z $\mathcal{C}$ też jest wstępujący i dodatkowo spełnia równość

$$A = \bigcup_{i=0}^{\infty}(A \cup A_i).$$

Ponieważ funkcja m jest σ-addytywna, więc

$$m(A) = m(\bigcup_{i=0}^{\infty}(A \cup A_i)) = \lim_i m(A \cup A_i) \leqslant \lim_i m(A_i).$$

Tak więc $m(A)$ ogranicza z dołu zbiór liczb z definicji $\mu(A)$. Stąd wynika nierówność $m(A) \leqslant \mu(A)$. $\square$

Będziemy mówić, że zbiór E dobrze dzieli zbiór A, jeżeli

$$\mu(A) = \mu(A \cap E) + \mu(A \cap E^c).$$

Najpierw pokażemy, że

Fakt 7.4 Jeżeli $E \in \mathcal{C}$, to E dobrze dzieli dowolny zbiór.

Dowód. Niech A będzie dowolnym zbiorem. Weźmy dowolny wstępujący ciąg $A_0, A_1, A_2, \ldots$ elementów z $\mathcal{C}$ taki, że $A \subseteq \bigcup_{i=0}^{\infty} A_i$. Będziemy rozważać dwa ciągi:

$$A_0 \cap E, A_1 \cap E, A_2 \cap E, \ldots \text{ oraz } A_0 \cap E^c, A_1 \cap E^c, A_2 \cap E^c, \ldots$$

Wyrazy tych ciągów należą do $\mathcal{C}$, oba ciągi są wstępujące, a ich sumy zawierają odpowiednio $A \cap E$ oraz $A \cap E^c$. Zauważmy, że

$$\lim_i \; m(A_i) = \lim_i \; m(A_i \cap E) + \lim_i \; m(A_i \cap E^c) \geqslant \mu(A \cap E) + \mu(A \cap E^c).$$

Stąd wynika, że liczba $\mu(A \cap E) + \mu(A \cap E^c)$ ogranicza z dołu zbiór z definicji $\mu(A)$ i z tego powodu spełnia nierówność

$$\mu(A) \geqslant \mu(A \cap E) + \mu(A \cap E^c).$$

Przeciwna nierówność wynika z podaddytywności μ, czyli z własności 3). □

Niech $\tilde{\mathcal{C}}$ oznacza rodzinę tych podzbiorów Ω, które dobrze dzielą wszystkie zbiory zawarte w Ω. Z udowodnionej wyżej własności otrzymujemy, że

Fakt 7.5 $\mathcal{C} \quad \tilde{\mathcal{C}}$. □

Fakt 7.6 Rodzina $\tilde{\mathcal{C}}$ jest zamknięta ze względu na dopełnienie. □

Fakt 7.7 Rodzina $\tilde{\mathcal{C}}$ jest zamknięta ze względu na sumę, a więc jest ciałem zbiorów.

Dowód. Przypuśćmy, że $A \subseteq \Omega$ i $E, F \in \tilde{\mathcal{C}}$. Teza wynika z następujących konsekwencji dobrego dzielenia przez zbiory E i F:

$$\mu(A) = \mu(A \cap E) + \mu(A \cap E^c) = \mu(A \cap E) + \mu(A \cap E^c \cap F) + \mu(A \cap E^c \cap F^c)$$

oraz

$$\mu(A \cap (E \cup F)) = \mu(A \cap (E \cup F) \cap E) + \mu(A \cap (E \cup F) \cap E^c) = \mu(A \cap E) + \mu(A \cap F \cap E^c).$$

Stąd

$$\mu(A) = \mu(A \cap (E \cup F)) + \mu(A \cap E^c \cap F^c) = \mu(A \cap (E \cup F)) + \mu(A \cap (E \cup F)^c). \;\; \square$$

Fakt 7.8 Funkcja μ jest addytywna na $\tilde{\mathcal{C}}$, a więc dla parami rozłącznych zbiorów $A, B \in \tilde{\mathcal{C}}$ zachodzi równość $\mu(A \cup B) = \mu(A) + \mu(B)$.

Dowód. Tak jest, ponieważ A dobrze dzieli $A \cup B$. □

Fakt 7.9 Funkcja μ na zbiorze $\tilde{\mathcal{C}}$ jest σ-addytywna, a więc dla dowolnego ciągu $A_0, A_1, A_2, \ldots$ parami rozłącznych zbiorów należących do $\tilde{\mathcal{C}}$ zachodzi równość

$$\mu(\bigcup_{i=0}^{\infty} A_i) = \sum_{i=0}^{\infty} \mu(A_i).$$

Dowód. Jest to oczywisty wniosek z faktów 7.8 i 7.2 □

Fakt 7.10 Jeżeli $E_0, E_1, E_2, \ldots$ jest zstępującym ciągiem elementów $\tilde{\mathcal{C}}$, to

$$\lim_n \; \mu(E_n \setminus (\bigcap_{i=0}^{\infty} E_i)) = 0.$$

Dowód. Zachodzą następujące równości

$$\mu(E_n \setminus (\bigcap_{i=0}^{\infty} E_i)) = \mu(\bigcup_{i=n}^{\infty}(E_i \setminus E_{i+1}) = \sum_{i=n}^{\infty}(\mu(E_i) - \mu(E_{i+1})) =$$
$$= \lim_k \sum_{i=n}^{k}(\mu(E_i) - \mu(E_{i+1})) = \lim_k (\mu(E_n) - \mu(E_{k+1})) = \mu(E_n) - (\lim_k \mu(E_{k+1})).$$

Dalej wystarczy przejść do granicy po n. □

Fakt 7.11 Przypuśćmy, że $E_0, E_1, E_2, \ldots$ jest zstępującym ciągiem elementów $\tilde{\mathcal{C}}$, a A jest dowolnym podzbiorem Ω. Wtedy

$$\lim_i \mu(A \cap E_i) = \mu(A \cap \bigcap_{i=0}^{\infty} E_i).$$

Dowód. Oczywiście,

$$A \cap E_n \subseteq (A \cap \bigcap_{i=0}^{\infty} E_i) \cup (E_n \setminus \bigcap_{i=0}^{\infty} E_i).$$

Wobec tego,

$$\mu(A \cap E_n) \leqslant \mu(A \cap \bigcap_{i=0}^{\infty} E_i) + \mu(E_n \setminus \bigcap_{i=0}^{\infty} E_i).$$

Jeżeli teraz przejdziemy do granicy po n i skorzystamy z faktu 7.10, to otrzymamy

$$\lim_n \mu(A \cap E_n) \leqslant \mu(A \cap \bigcap_{i=0}^{\infty} E_i).$$

Odwrotna nierówność wynika z własności 5). □

Fakt 7.12 Rodzina $\tilde{\mathcal{C}}$ jest zamknięta na przeliczalne sumy wstępujących ciągów, a więc dla dowolnego wstępującego ciągu $E_0, E_1, E_2, \ldots$ elementów $\tilde{\mathcal{C}}$ mamy $\bigcup_{i=0}^{\infty} E_i \in \tilde{\mathcal{C}}$.

Dowód. Niech $E_0, E_1, E_2, \ldots$ będzie wstępującym ciągiem elementów $\tilde{\mathcal{C}}$ i niech A będzie dowolnym zbiorem. Pokażemy, że $\bigcup_{i=0}^{\infty} E_i$ dobrze dzieli zbiór A. Z faktu 7.2 wynika, że

$$\mu(A \cap \bigcup_{i=0}^{\infty} E_i) = \mu(\bigcup_{i=0}^{\infty}(A \cap E_i)) = \lim_i \mu(A \cap E_i).$$

Podobnie, na podstawie faktu 7.11,

$$\mu(A \cap (\bigcup_{i=0}^{\infty} E_i)^c) = \mu(A \cap \bigcap_{i=0}^{\infty} E_i^c) = \lim_i \mu(A \cap E_i^c).$$

Po zsumowaniu udowodnionych równości stronami i skorzystaniu z należenia zbiorów E_i do $\tilde{\mathcal{C}}$ otrzymujemy, że

$$\mu(A \cap \textstyle\bigcup_{i=0}^{\infty} E_i) + \mu(A \cap (\bigcup_{i=0}^{\infty} E_i)^c) =$$
$$= \lim_i \mu(A \cap E_i) + \lim_i \mu(A \cap E_i^c) = \lim_i (\mu(A \cap E_i) + \mu(A \cap E_i^c)) = \lim_i \mu(A) = \mu(A). \quad \square$$

Fakt 7.13 Rodzina $\tilde{\mathcal{C}}$ jest σ-ciałem zbiorów.

Dowód. Jest to oczywisty wniosek z faktu 7.12. $\square$

Z faktów 7.5, 7.3, 7.9 i 7.13 wynika, że funkcję m udało się nam rozszerzyć do σ-addytywnej funkcji μ określonej na pewnym σ-ciele zawierającym $\mathcal{C}$, a więc także na najmniejsze σ-ciało zawierające $\mathcal{C}$. Pozostała jeszcze do wykazania jednoznaczność rozszerzenia.

Fakt 7.14 Jeżeli μ_1 i μ_2 są σ-addytywnymi określonymi na σ-ciele $\mathcal{C}'$ podzbiorów Ω i $\mu_1(\Omega) = \mu_2(\Omega)$, to zbiór

$$\{A \in \mathcal{C}' : \mu_1(A) = \mu_2(A)\}$$

też jest σ-ciałem.

8 Dodatek: Dowód twierdzenia de Moivre'a - Laplace'a

8.1 Sformułowanie twierdzenia raz jeszcze

Przypuśćmy, że mamy daną przestrzeń probabilistyczną i zdarzenie Q. Sukcesem będziemy nazywać zajście zdarzenia Q. Przyjmujemy, że uzyskujemy sukces z prawdopodobieństwem p, z zdarzenie przeciwne (porażka) ma miejsce z prawdopodobieństwem $1 - p = q$. Będziemy wielokrotnie i w sposób niezależny powtarzać doświadczenie losowe opisane przez daną przestrzeń. Niech X_n oznacza zmienną losową, która n-krotnej serii doświadczeń losowych przyporządkowuje liczbę sukcesów.

Twierdzenie 8.1 (de Moivre'a-Laplace'a) *Niech A i B będą dowolnymi liczbami takimi, że $A < B$. Wtedy*

$$\lim_{n\to\infty} P_n(\{\omega \in \Omega_n : A \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} < B\}) = N_{0,1}([A, B)) = \frac{1}{\sqrt{2\pi}} \int_A^B e^{-\frac{z^2}{2}}\, dz.$$

Dowód twierdzenia zajmie kilka kolejnych rozdziałów.

8.2 Wzór Stirlinga i konsekwencje

Znany (z wykładu analizy) wzór Stirlinga stwierdza, że

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} = 1.$$

Weźmy dodatnią liczbę $\varepsilon < 1$. Z definicji granicy otrzymujemy, że jest taka liczba n_0, że dla dowolnej liczby $n > n_0$ zachodzą nierówności

$$1 - \frac{\varepsilon}{7} < \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} < 1 + \frac{\varepsilon}{7}$$

oraz

$$1 - \frac{\varepsilon}{7} < \frac{\sqrt{2\pi n} n^n e^{-n}}{n!} < 1 + \frac{\varepsilon}{7}.$$

Nierówności te pozwalają oszacować symbole Newtona. Przypuśćmy, że mamy liczby n i k takie, że $n, k, n-k > n_0$. Wtedy, po wymnożeniu odpowiednich nierówności stronami, otrzymujemy

$$1-\varepsilon < (1-\frac{\varepsilon}{7})^3 < \frac{\binom{n}{k}}{\frac{\sqrt{2\pi n}n^n e^{-n}}{\sqrt{2\pi k}k^k e^{-k}\sqrt{2\pi(n-k)}(n-k)^{n-k}e^{-n+k}}} < (1+\frac{\varepsilon}{7})^3 < 1+\varepsilon.$$

Uprościmy jeszcze mianownik tego skomplikowanego ułamka:

$$\frac{\sqrt{2\pi n}n^n e^{-n}}{\sqrt{2\pi k}k^k e^{-k}\sqrt{2\pi(n-k)}(n-k)^{n-k}e^{-n+k}} = \frac{1}{\sqrt{2\pi n}\sqrt{\frac{k(n-k)}{n^2}}}(\frac{k}{n})^{-k}(\frac{n-k}{n})^{-n+k}$$

W ten sposób dowiedliśmy, że

Lemat 8.2 *Dla każdego dodatniego $\varepsilon < 1$ istnieje liczba naturalna n_0 taka, że dla wszystkich liczb n i k spełniających warunki $n, k, n-k > n_0$ zachodzą nierówności*

$$1-\varepsilon < \frac{\binom{n}{k}}{\frac{1}{\sqrt{2\pi n}\sqrt{\frac{k(n-k)}{n^2}}}(\frac{k}{n})^{-k}(\frac{n-k}{n})^{-n+k}} < 1+\varepsilon.$$

8.3 Zbiór W i jego własności

Przyjmijmy, że

$$W = \{(n,k) \in N^2 : pn + A\sqrt{npq} \leqslant k \leqslant pn + B\sqrt{npq}\}.$$

Będziemy rozważać pary liczb naturalnych n i k należące do W. Zauważmy, że takie pary spełniają nierówność

$$qn - B\sqrt{npq} \leqslant n-k \leqslant qn - A\sqrt{npq}.$$

Wobec tego, dla wszystkich $(n,k) \in W$ mamy

$$1 + \frac{A}{\sqrt{npq}} \leqslant \frac{k}{pn} \leqslant 1 + \frac{B}{\sqrt{npq}} \text{ oraz } 1 - \frac{B}{\sqrt{npq}} \leqslant \frac{n-k}{qn} \leqslant 1 - \frac{A}{\sqrt{npq}}. \tag{7}$$

Stąd wynika, że dla dowolnej dodatniej liczby $\varepsilon < 1$ istnieje liczba naturalna n_0 taka, że dla wszystkich $n > n_0$ i dla k takich, że $(n,k) \in W$ zachodzą nierówności

$$1-\varepsilon \leqslant \frac{pn}{k} \leqslant 1+\varepsilon \text{ oraz } 1-\varepsilon \leqslant \frac{qn}{n-k} \leqslant 1+\varepsilon.$$

Jeżeli powyższe nierówności pomnożymy przez siebie i spierwiastkujemy stronami to otrzymamy, że

$$1-\varepsilon \leqslant \sqrt{\frac{pqn^2}{k(n-k)}} \leqslant 1+\varepsilon.$$

Lemat 8.3 *Dla każdego dodatniego $\varepsilon < 1$ istnieje liczba naturalna n_0 taka, że dla wszystkich liczb $n > n_0$ i k spełniających warunek $(n,k) \in W$ zachodzą nierówności*

$$1-\varepsilon < \frac{\binom{n}{k}}{\frac{1}{\sqrt{2\pi npq}}(\frac{k}{n})^{-k}(\frac{n-k}{n})^{-n+k}} < 1+\varepsilon.$$

Dowód. Aby dowieść ten lemat, wystarczy nierówność z lematu 8.2 pomnożyć przez wyżej podaną. Samo mnożenie wykonujemy podobnie, jak w dowodzie lematu 8.2. □

Udowodniony lemat wykorzystamy do szacowania wyrażenia

$$\binom{n}{k} p^k q^{n-k}.$$

Będziemy więc przekształcać mianownik z wzoru z tego lematu przemnożony przez odpowiednie potęgi p i q. Na razie nadajemy mu postać

$$\frac{1}{\sqrt{2\pi npq}}(\frac{k}{n})^{-k}(\frac{n-k}{n})^{-n+k}p^k q^{n-k} = \frac{1}{\sqrt{2\pi npq}}e^S,$$

gdzie

$$S = -k\log\frac{k}{n} - (n-k)\log\frac{n-k}{n} + k\log p + (n-k)\log q.$$

Dalej będziemy się posługiwać liczbą

$$z = \frac{k-np}{\sqrt{npq}} = \left(\frac{k}{np}-1\right)\sqrt{\frac{np}{q}}.$$

Na mocy nierówności (7), dla $(n,k) \in W$ liczba ta spełnia nierówności

$$A \leqslant z \leqslant B. \tag{8}$$

Zauważmy, że

$$\log\frac{k}{n} = \log\frac{np + z\sqrt{npq}}{n} = \log(p + z\sqrt{\frac{pq}{n}})$$

oraz

$$\log\frac{n-k}{n} = \log\frac{nq - z\sqrt{npq}}{n} = \log(q - z\sqrt{\frac{pq}{n}})$$

8.4 Wzór Taylora i dalszy przekształcenia

Zgodnie ze wzorem Taylora, jeżeli f jest funkcją trzykrotnie różniczkowalną w przedziale $[a,b]$, to

$$f(b) = f(a) + \frac{b-a}{1!}f'(a) + \frac{(b-a)^2}{2!}f''(a) + \frac{(b-a)^3}{3!}f'''(a+\theta(b-a))$$

dla pewnej liczby $\theta \in (0,1)$. Ten sam wzór jest słuszny dla funkcji trzykrotnie różniczkowalnych w przedziale $[b,a]$.

Z wzoru Taylora skorzystamy z dla funkcji $f(x) = \log(a + x)$. W tym przypadku wzór ten przyjmuje postać

$$\log(a+x) = \log a + \frac{x}{a} - \frac{x^2}{2a^2} + \frac{x^3}{3a^3}\frac{1}{(1+\theta\frac{x}{a})^3}. \tag{9}$$

Będziemy z niego korzystać na dwa sposoby: dla $a = p$ i $x = z\sqrt{\frac{pq}{n}}$ oraz dla $a = q$ i $x = -z\sqrt{\frac{pq}{n}}$. Zauważmy, że w interesujących nas przypadkach

$$\frac{x}{a} = \frac{z}{p}\sqrt{\frac{pq}{n}} = \frac{k}{np} - 1$$

lub

$$\frac{x}{a} = -\frac{z}{q}\sqrt{\frac{pq}{n}} = \frac{n-k}{nq} - 1.$$

Najpierw oszacujemy ostatnie składniki wzoru (9).

Weźmy znowu dodatnią liczbę ε. Z nierówności (7) wynika, że

$$\left|\frac{x}{a}\right| \leqslant \frac{\varepsilon}{2} \tag{10}$$

dla wszystkich liczb $n > N$ i liczb k takich, że $(n, k) \in W$. W szczególności dla liczb $n \geqslant 4\max(A^2, B^2)$ mamy, że

$$\left|\frac{x}{a}\right| \leqslant \frac{1}{2}. \tag{11}$$

Tym bardziej

$$\left|\theta\frac{x}{a}\right| \leqslant \frac{1}{2}$$

oraz

$$\frac{1}{2} \leqslant 1 + \theta\frac{x}{a} \leqslant \frac{3}{2}.$$

Jeżeli tę ostatnią nierówność podniesiemy stronami do trzeciej potęgi i odwrócimy, to otrzymamy

$$\frac{8}{27} \leqslant \frac{1}{(1+\theta\frac{x}{a})^3} \leqslant 8. \tag{12}$$

Środkowy ułamek w nierówności (12) jest więc dodatni. Podniesiemy teraz nierówność (10) do trzeciej potęgi i przemnożymy przez ten ułamek. Otrzymujemy

$$\left|\frac{x^3}{a^3}\frac{1}{(1+\theta\frac{x}{a})^3}\right| \leqslant \frac{\varepsilon}{8}\frac{1}{(1+\theta\frac{x}{a})^3} \leqslant \varepsilon.$$

Oznacza to, że wartości bezwzględne ostatnich składników we wzorze (9), w interesujących nas przypadkach, dla dostatecznie dużych liczb n nie przekraczają $\frac{\varepsilon}{3}$.

Teraz obliczamy S. Mamy

$$\begin{aligned} S = & -k\cdot(\log p + \frac{z}{p}\sqrt{\frac{pq}{n}} - \frac{z^2}{2p^2}\frac{pq}{n} + \frac{s_1}{3} - \log p) \\ & -(n-k)\cdot(\log q - \frac{z}{q}\sqrt{\frac{pq}{n}} - \frac{z^2}{2q^2}\frac{pq}{n} + \frac{s_2}{3} - \log q), \end{aligned}$$

gdzie s_1 i s_2 są liczbami o wartościach bezwzględnych $\leqslant 1$. Zauważmy, że upraszczają się pierwsze i piąte składniki w obu linijkach. Zsumujemy osobno drugie i trzecie składniki. Suma drugich składników jest równa

$$-k\frac{z}{p}\sqrt{\frac{pq}{n}} + (n-k)\frac{z}{q}\sqrt{\frac{pq}{n}} = z\sqrt{\frac{pq}{n}}\left(\frac{n-k}{q} - \frac{k}{p}\right) = z\sqrt{\frac{pq}{n}} \cdot \frac{(n-k)p - kq}{pq} = z \cdot \frac{pn-k}{\sqrt{npq}} = -z^2.$$

Suma trzecich składników wynosi

$$k\frac{z^2}{2p^2}\frac{pq}{n} + (n-k)\frac{z^2}{2q^2}\frac{pq}{n} = \frac{z^2}{2}\left(\frac{kq}{np} + \frac{(n-k)p}{nq}\right).$$

Wyrażenie w nawiasach szacujemy korzystając z nierówności (7). Dobierając odpowiednio N otrzymujemy, że dla $n > N$ i k takich, że $(n,k) \in W$ zachodzą nierówności

$$1 - \varepsilon \leqslant \left(\frac{kq}{np} + \frac{(n-k)p}{nq}\right) \leqslant 1 + \varepsilon.$$

Sumując oszacowania na poszczególne składniki S otrzymujemy, że

$$-\varepsilon - \frac{\varepsilon}{2}z^2 - \frac{1}{2}z^2 \leqslant S \leqslant -\frac{1}{2}z^2 + \frac{\varepsilon}{2}z^2 + \varepsilon \tag{13}$$

dla dostatecznie dużych n.

Lemat 8.4 *Dla dowolnej dodatniej liczby $\varepsilon < 1$ istnieje liczba naturalna N taka, że dla wszystkich $n > N$ i wszystkich K spełniających warunek $(n,k) \in W$ zachodzą nierówności*

$$-\varepsilon - \frac{z^2}{2} \leqslant S \leqslant -\frac{z^2}{2} + \varepsilon.$$

Dowód. Nierówność (8) pozwala oszacować także wyrażenie $\frac{\varepsilon}{2}z^2$ we wzorze (13). Dalej wystarczy skorzystać z nierówności (13) dla odpowiednio dobranej liczby ε. $\square$

Twierdzenie 8.5 *Dla każdego dodatniego $\varepsilon < 1$ istnieje liczba naturalna n_0 taka, że dla wszystkich liczb $n > n_0$ i k spełniających warunek $(n,k) \in W$ zachodzą nierówności*

$$1 - \varepsilon < \frac{\binom{n}{k}p^k q^{n-k}}{\frac{1}{\sqrt{2\pi npq}} \cdot e^{-\frac{(k-np)^2}{2npq}}} < 1 + \varepsilon.$$

Dowód. Ustalamy dodatnie $\varepsilon < 1$. Będą nam potrzebne dwie funkcje pomocnicze

$$k(x) = (1-x)e^{-x} \quad \text{oraz} \quad l(x) = (1+x)e^{x}.$$

Zauważmy, że

$$\lim_{x\to 0} k(x) = \lim_{x\to 0} l(x) = 1.$$

Z definicji granicy, dla ustalonego ε znajdujemy dodatnią liczbę $\delta < 1$ taką, że

$$1 - \varepsilon \leqslant k(\delta) = (1-\delta)e^{-\delta} \text{ oraz } l(\delta) = (1+\delta)e^{\delta} \leqslant 1 + \varepsilon.$$

Teraz korzystamy z lematu 8.3 biorąc za ε z tego lematu liczbę δ i podstawiając wprowadzone oznaczenie S. Otrzymujemy

$$1 - \delta < \frac{\binom{n}{k} p^k q^{n-k}}{\frac{1}{\sqrt{2\pi npq}} e^S} < 1 + \delta.$$

Następnie w podobny sposób korzystamy z nierówności z lematu 8.4. Stąd i z monotoniczności funkcji wykładniczej wnioskujemy, że

$$e^{-\delta} \leqslant \frac{e^S}{e^{-\frac{z^2}{2}}} \leqslant e^{\delta}.$$

Mnożąc otrzymane nierówności przez siebie otrzymujemy, że

$$1 - \varepsilon \leqslant (1-\delta)e^{-\delta} \leqslant \frac{\binom{n}{k} p^k q^{n-k}}{\frac{1}{\sqrt{2\pi npq}} e^S} \cdot \frac{e^S}{e^{-\frac{z^2}{2}}} < (1+\delta)e^{\delta} \leqslant 1 + \varepsilon.$$

Nietrudno zauważyć, że otrzymaliśmy dowodzoną nierówność. □

8.5 Dowód twierdzenia de Moivre'a - Laplace'a

Z twierdzenia 8.5 wynika dla wszystkich par $(n, k) \in W$ i dla dostatecznie dużych n nierówność

$$\frac{1-\varepsilon}{\sqrt{2\pi npq}} \cdot e^{-\frac{(k-np)^2}{2npq}} < \binom{n}{k} p^k q^{n-k} < \frac{1+\varepsilon}{\sqrt{2\pi npq}} \cdot e^{-\frac{(k-np)^2}{2npq}}.$$

Wobec tego

$$\frac{1-\varepsilon}{\sqrt{2\pi}} \cdot \sum_{(n,k)\in W} \frac{1}{\sqrt{npq}} \cdot e^{-\frac{(k-np)^2}{2npq}} < \sum_{(n,k)\in W} \binom{n}{k} p^k q^{n-k} < \frac{1+\varepsilon}{\sqrt{2\pi}} \cdot \sum_{(n,k)\in W} \frac{1}{\sqrt{npq}} \cdot e^{-\frac{(k-np)^2}{2npq}}.$$

Dla dużych n skrajne sumy są sumami występującymi w definicji całki Riemanna funkcji $e^{-\frac{1}{2}z^2}$. Wobec tego, sumy te dążą do odpowiedniej całki. Tak więc

$$\frac{1-\varepsilon}{\sqrt{2\pi}} \cdot \int_A^B e^{-\frac{z^2}{2}} dz \leqslant \lim_{n\to\infty} \sum_{(n,k)\in W} \binom{n}{k} p^k q^{n-k} \leqslant \frac{1+\varepsilon}{\sqrt{2\pi}} \cdot \int_A^B e^{-\frac{z^2}{2}} dz.$$

Suma środkowa jest interesującym nas prawdopodobieństwem:

$$\frac{1-\varepsilon}{\sqrt{2\pi}} \cdot \int_A^B e^{-\frac{z^2}{2}}\, dz \leqslant \lim_{n\to\infty} P_n(\{\omega \in \Omega_n : A \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} \leqslant B\}) \leqslant \frac{1+\varepsilon}{\sqrt{2\pi}} \cdot \int_A^B e^{-\frac{z^2}{2}}\, dz.$$

Ponieważ nierówności te są prawdziwe dla wszystkich dodatnich liczb $\varepsilon < 1$, więc

$$\lim_{n\to\infty} P_n(\{\omega \in \Omega_n : A \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} \leqslant B\}) = \frac{1}{\sqrt{2\pi}} \cdot \int_A^B e^{-\frac{z^2}{2}}\, dz. \ \Box$$

Wniosek 8.6 *Dla każdej liczby rzeczywistej B*

$$\lim_{n\to\infty} P_n(\{\omega \in \Omega_n : \frac{X_n(\omega) - np}{\sqrt{npq}} < B\}) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^B e^{-\frac{z^2}{2}}\, dz = N_{0,1}((-\infty, B)).$$

Dowód. Ustalmy $\varepsilon > 0$. Najpierw znajdujemy liczbę $A > |\, B \,|$ taką, że

$$N_{0,1}([-A, A)) = \frac{1}{\sqrt{2\pi}} \int_{-A}^A e^{-\frac{z^2}{2}}\, dz \geqslant 1 - \frac{\varepsilon}{4}.$$

Dla takiego A mamy też oczywiście

$$N_{0,1}((-\infty, A)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-A} e^{-\frac{z^2}{2}}\, dz \leqslant \frac{\varepsilon}{8}.$$

Korzystając z twierdzenia de Moivre'a - Laplace'a znajdujemy n_0 takie, że dla $n > n_0$

$$P_n(\{\omega \in \Omega_n : -A \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} < A\}) \geqslant 1 - \frac{\varepsilon}{2}.$$

Dla takich n

$$P_n(\{\omega \in \Omega_n : \frac{X_n(\omega) - np}{\sqrt{npq}} < -A\}) \leqslant \frac{\varepsilon}{2}.$$

Zauważmy, że

$$\begin{aligned}
&\left| P_n(\{\omega \in \Omega_n : \frac{X_n(\omega) - np}{\sqrt{npq}} < B\}) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^B e^{-\frac{z^2}{2}}\, dz \right| = \\
&= \left| P_n(\{\omega : \frac{X_n(\omega) - np}{\sqrt{npq}} < -A\}) + P_n(\{\omega : -A \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} < B\}) - N_{0,1}((-\infty, -A)) - N_{0,1}([-A, B)) \right| \\
&\leqslant \left| P_n(\{\omega : -A \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} < B\}) - N_{0,1}([-A, B)) \right| + P_n(\{\omega : \frac{X_n(\omega) - np}{\sqrt{npq}} < -A\}) + N_{0,1}((-\infty, -A)) \\
&\leqslant \left| P_n(\{\omega \in \Omega_n : -A \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} < B\}) - N_{0,1}([-A, B)) \right| + \frac{\varepsilon}{2} + \frac{\varepsilon}{4}.
\end{aligned}$$

Jeżeli teraz po obu stronach nierówności przejdziemy do granicy i znowu skorzystamy z twierdzenia de Moivre'a - Laplace'a, to otrzymamy, że

$$\left| \lim_{n\to\infty} P_n(\{\omega \in \Omega_n : \frac{X_n(\omega) - np}{\sqrt{npq}} < B\}) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{B} e^{-\frac{z^2}{2}}\, dz \right| \leqslant \frac{3\varepsilon}{4}.$$

Ponieważ tak jest dla każdego $\varepsilon > 0$, więc

$$\lim_{n\to\infty} P_n(\{\omega \in \Omega_n : \frac{X_n(\omega) - np}{\sqrt{npq}} < B\}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{B} e^{-\frac{z^2}{2}}\, dz. \ \square$$

8.6 Dwa wnioski z twierdzenia de Moivre'a - Laplace'a

Twierdzenie 8.7 *Granica*

$$\lim_{n\to\infty} P_n(\{\omega \in \Omega_n : A \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} \leqslant B\}) = \frac{1}{\sqrt{2\pi}} \int_{A}^{B} e^{-\frac{z^2}{2}}\, dz$$

z twierdzenia de Moivre'a - Laplace'a jest jednostajna ze względu na A i B.

Dowód. W dowodzie wykorzystamy funkcję zdefiniowaną wzorem

$$f(x) = \int_{-\infty}^{x} e^{-\frac{z^2}{2}}\, dz.$$

Jest to funkcja ciągła i przyjmuje wartości bliskie zarówno 0 jak i 1. Wobec tego przyjmuje wszystkie wartości z przedziału $(0, 1)$.

Weźmy dodatnią liczbę $\varepsilon < 1$. Dla wszystkich dodatnich liczb naturalnych $k \leqslant \frac{4}{\varepsilon}$ bierzemy liczbę x_k taką, że $f(x_k) = \frac{k\varepsilon}{4}$. Ciąg tych liczb jest rosnący oraz zachodzą równości

$$0 < \int_{x_k}^{x_{k+1}} e^{-\frac{z^2}{2}}\, dz = f(x_{k+1}) - f(x_k) = \frac{\varepsilon}{4}.$$

Mamy także

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_1} e^{-\frac{z^2}{2}}\, dz = \frac{\varepsilon}{4} \quad \text{oraz} \quad \int_{x_K}^{\infty} e^{-\frac{z^2}{2}}\, dz \leqslant \frac{\varepsilon}{4}$$

dla $K = \left\lfloor \frac{4}{\varepsilon} \right\rfloor$. Przyjmijmy dla uproszczenia zapisu, że $x_0 = -\infty$ i $x_{K+1} = \infty$. Tak więc cała prosta została rozbita na skończenie wiele odcinków

$$(x_0, x_1), [x_1, x_2), \ldots, [x_{K-1}, x_K), [x_K, x_{K+1}),$$

których miara w sensie rozkładu normalnego jest nie większą niż $\frac{\varepsilon}{4}$. Weźmy teraz liczbę n_0 taką, że dla wszystkich $i \leqslant K$ oraz $n > n_0$ zachodzą nierówności

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_{i+1}} e^{-\frac{z^2}{2}}\, dz - \frac{\varepsilon}{4} \leqslant P_n(\{\omega : \frac{X_n(\omega) - np}{\sqrt{npq}} < x_{i+1}\}) \leqslant \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_{i+1}} e^{-\frac{z^2}{2}}\, dz + \frac{\varepsilon}{4}.$$

Niech a i b będą liczbami $\leqslant K$ takimi, że

$$A \in [x_a, x_{a+1}) \text{ oraz } B \in [x_b, x_{b+1}).$$

Oczywiście, $a \leqslant b$. Dla $n > n_0$ mamy też

$$P_n(\{\omega : A \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} < B\}) \leqslant P_n(\{\omega : x_a \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} < x_{b+1}\}) =$$
$$P_n(\{\omega : \frac{X_n(\omega) - np}{\sqrt{npq}} < x_{b+1}\}) - P_n(\{\omega : \frac{X_n(\omega) - np}{\sqrt{npq}} < x_a\}) \leqslant$$
$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_{b+1}} e^{-\frac{z^2}{2}} dz - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_a} e^{-\frac{z^2}{2}} dz + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{1}{\sqrt{2\pi}} \int_{x_a}^{x_{b+1}} e^{-\frac{z^2}{2}} dz + \frac{\varepsilon}{2} \leqslant$$
$$\frac{1}{\sqrt{2\pi}} \int_{x_a}^{x_{a+1}} e^{-\frac{z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_{A}^{B} e^{-\frac{z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_{x_b}^{x_{b+1}} e^{-\frac{z^2}{2}} dz + \frac{\varepsilon}{2} \leqslant \frac{1}{\sqrt{2\pi}} \int_{A}^{B} e^{-\frac{z^2}{2}} dz + \varepsilon.$$

Analogicznie pokazujemy, że

$$P_n(\{\omega : A \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} < B\}) \geqslant \frac{1}{\sqrt{2\pi}} \int_{A}^{B} e^{-\frac{z^2}{2}} dz - \varepsilon.$$

Teza wynika z udowodnionych nierówności. $\square$

Twierdzenie 8.8 *Dla każdego $\varepsilon > 0$ istnieje n_0 takie, że dla wszystkich $n > n_0$ i dla dowolnych A i B zachodzą nierówności*

$$N_{np,npq}([A, B)) - \varepsilon \leqslant P_n(\{\omega : A \leqslant X_n(\omega) < B\}) \leqslant N_{np,npq}([A, B)) + \varepsilon$$

Dowód. Będziemy posługiwać się funkcją

$$\varphi(x) = \frac{x - np}{\sqrt{npq}}.$$

Jest to funkcja rosnąca. Wobec tego warunki

$$\varphi(A) \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} < \varphi(B) \text{ oraz } A \leqslant X_n(\omega) < B$$

są równoważne.

Z poprzedniego twierdzenia wynika, że dla dostatecznie dużych n i dla wszystkich A i B, a więc także dla $\varphi(A)$ i $\varphi(B)$ zachodzą nierówności

$$\frac{1}{\sqrt{2\pi}} \int_{\varphi(A)}^{\varphi(B)} e^{-\frac{z^2}{2}} dz - \varepsilon \leqslant P_n(\{\omega \in \Omega_n : \varphi(A) \leqslant \frac{X_n(\omega) - np}{\sqrt{npq}} < \varphi(B)\}) \leqslant \frac{1}{\sqrt{2\pi}} \int_{\varphi(A)}^{\varphi(B)} e^{-\frac{z^2}{2}} dz + \varepsilon.$$

Po skorzystaniu z równoważności warunków otrzymujemy, że

$$\frac{1}{\sqrt{2\pi}} \int_{\varphi(A)}^{\varphi(B)} e^{-\frac{z^2}{2}} dz - \varepsilon \leqslant P_n(\{\omega \in \Omega_n : A \leqslant X_n < B\}) \leqslant \frac{1}{\sqrt{2\pi}} \int_{\varphi(A)}^{\varphi(B)} e^{-\frac{z^2}{2}} dz + \varepsilon.$$

Dalej korzystamy z wzoru na całkowanie przez podstawianie:

$$\int_{\varphi(A)}^{\varphi(B)} e^{-\frac{z^2}{2}} dz = \int_{A}^{B} e^{-\frac{\varphi^2(z)}{2}} \varphi'(z) dz = \int_{A}^{B} \frac{1}{\sqrt{npq}} e^{-\frac{(z - np)^2}{2npq}} dz. \ \square$$

9 Literatura

1. William Feller, *Wstęp do rachunku prawdopodobieństwa*, t.1 i 2, PWN, 1966.
2. Mirosław Krzyśko, *Wykłady z teorii prawdopodobieństwa*, WNT, 2000.
3. John Lamperti, *Probability, A survey of the mathematical theory*, 1966, wyd. rosyjskie z 1973.
4. Yakov G. Sinai, *Probability theory, An Introductory Course*, Springer-Verlag, 1992.