

Wykład z Rachunku Prawdopodobieństwa WNE, 2011/2012

1. PODSTAWOWE SCHEMATY KOMBINATORYCZNE

Wariacje z powtórzeniami. Załóżmy, iż mamy zbiór n -elementowy A . Wówczas liczba k -elementowych ciągów o wyrazach ze zbioru A wynosi $n \cdot n \cdot \dots \cdot n = n^k$.

Wariacje bez powtórzeń. Załóżmy, iż mamy zbiór n -elementowy A . Wówczas liczba k -elementowych *różnowartościowych* ciągów o wyrazach ze zbioru A wynosi $n \cdot (n-1) \cdot \dots \cdot (n-k+1) = n!/(n-k)!$, o ile $k \leq n$, i 0 jeśli $k > n$.

Permutacje. Są to wariacje n -elementowe zbioru n -elementowego: inaczej, są to ustawienia elementów zbioru w ciąg. Ich liczba wynosi $n!$.

Kombinacje. Załóżmy, że mamy zbiór n -elementowy A . Wówczas liczba k -elementowych podzbiorów zbioru A wynosi $\binom{n}{k}$, gdzie

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & \text{jeśli } 0 \leq k \leq n, \\ 0 & \text{w p.p.} \end{cases}$$

2. RYS HISTORYCZNY

Motywacje:

- gry hazardowe,
- zjawiska masowe (statystyki urodzeń i zgonów).
- aksjomatyka Kołmogorowa, 1933 r.

3. PRZYKŁADY PROSTYCH MODELI PROBABILISTYCZNYCH: DYSKRETNYCH I CIĄGŁYCH

Przypuśćmy, że wykonujemy eksperyment losowy. Powstaje natychmiast pytanie: w jaki sposób opisać go matematycznie?

Przede wszystkim, na pewno możemy mówić o jego potencjalnych wynikach: *zdarzenia elementarne* to możliwe wyniki tego eksperymentu. Zbiór wszystkich zdarzeń elementarnych oznaczamy literą Ω . Zdarzenie elementarne oznaczamy literą ω .

1. Rzut monetą: możliwe dwa wyniki: $\Omega = \{O, R\}$. $|\Omega| = 2$.
2. Rzut kostką: możliwe sześć wyników: $\Omega = \{1, 2, 3, 4, 5, 6\}$. $|\Omega| = 6$.
3. Rzut dwiema kostkami, patrzymy na sumę oczek: $\Omega = \{2, 3, \dots, 12\}$. Zauważmy, że, intuicyjnie, wyniki *nie są jednakowo prawdopodobne*. Suma 2 zdarza się tylko gdy wypadły dwie 1; a np. suma 7 zdarza się, gdy wypadnie 3 i 4, 4 i 3, 2 i 5, itp. $|\Omega| = 11$.
4. Z talii kart losujemy 5 kart. Wynikiem jest pięcioelementowa kombinacja zbioru kart; zatem Ω to zbiór pięcioelementowych podzbiorów zbioru 52-elementowego. $|\Omega| = \binom{52}{5}$.
5. Rzucamy igłę na stół i mierzymy kąt jaki tworzy z wybraną krawędzią stołu. Wynik to liczba z przedziału $[0, 2\pi)$. $\Omega = [0, 2\pi)$. Jest to przykład ciągłego doświadczenia losowego.

4. σ -CIAŁA

Zdarzenia. Często nie interesuje nas konkretny wynik ω , ale to, czy należy on do wcześniej ustalonego podzbioru A zbioru Ω . Takie podzbiory A nazywamy *zdarzeniami*.

Przykład: Przy rzucie kostką, może nas np. interesować $A = \{1, 3, 5\}$ - zdarzenie polegające na tym, że wypadła nieparzysta liczba oczek.

Jeśli ω - wynik, A - zdarzenie, to:

- jeśli $\omega \in A$, to mówimy, że zaszło A bądź że ω sprzyja A .

- jeśli $\omega \notin A$, to mówimy, że nie zaszło A , bądź że zaszło zdarzenie przeciwne, zdefiniowane jako $A' = \Omega \setminus A$. A' nazywamy też *dopełnieniem* zbioru A .

Na przykład, przy jednokrotnym rzucie kostką może interesować nas wypadnięcie nieparzystej liczby oczek, bądź w przykładzie z talią kart, może nas interesować zdarzenie: „wylosowaliśmy co najmniej 2 asy”.

Szczególne zdarzenia, interpretacje działań/relacji na zdarzeniach:

Ω - zdarzenie pewne,

\emptyset - zdarzenie niemożliwe,

$A \cap B$ - zaszły oba zdarzenia A, B ,

$A \cap B = \emptyset$ - zdarzenia się wykluczają (są rozłączne),

$A \cup B$ - zaszło A lub B ,

A' - nie zaszło A ,

$A \setminus B = A \cap B'$ - zaszło A i nie zaszło B ,

$A \subseteq B$ - A pociąga za sobą B .

Przypuśćmy, że mamy Ω i chcemy zdefiniować sensowną klasę zdarzeń (cokolwiek to znaczy). Naturalny pomysł: rozważać 2^Ω - wszystkie możliwe podzbiory; czasem jednak ta klasa jest zbyt duża i nie da się na niej dobrze pracować.

Rozsądna klasa zdarzeń powinna być zamknięta na branie sumy, iloczynu i zdarzenia przeciwnego. To prowadzi do pojęcia ciała oraz σ -ciała.

Definicja 1. Rodzinę \mathcal{F} podzbiorów Ω nazywamy σ -ciałem, jeśli

- (i) $\emptyset \in \mathcal{F}$,
- (ii) $A \in \mathcal{F} \Rightarrow A' \in \mathcal{F}$,
- (iii) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

5. INTUICJA WIODĄCA DO OKREŚLENIA PRAWDOPODOBIEŃSTWA - CZĘSTOŚĆ ZDARZEŃ

Rozważmy następujący przykład. Jeśli rzucamy (tą samą) monetą wiele razy, to oczekujemy (i rzeczywiście tak bywa), że orzeł pojawi się w przybliżeniu w połowie przypadków. Tak więc „częstościowo”, prawdopodobieństwo wypadnięcia orła to $1/2$. Teraz ogólniej: załóżmy, że wykonujemy eksperyment, w którym zbiór zdarzeń elementarnych to Ω oraz A jest zdarzeniem. Załóżmy, że powtarzamy eksperyment n razy i definiujemy

$$\rho_n(A) = \frac{\text{liczba zajęć } A}{n}.$$

Liczbę tę nazywamy *częstością zdarzenia* A . Gdy n jest duże, spodziewamy się, że $\rho_n(A)$ powinno z grubsza mówić o prawdopodobieństwie A .

Spójrzmy na własności ρ_n : jak łatwo sprawdzić,

- (i) $0 \leq \rho_n(A) \leq 1$,
- (ii) $\rho_n(\Omega) = 1$,
- (iii) $A \cap B = \emptyset \Rightarrow \rho_n(A \cup B) = \rho_n(A) + \rho_n(B)$.

Ponadto, zauważmy, iż $\rho_n(A) = 1 - \rho_n(A')$. Naturalnym pomysłem jest określić prawdopodobieństwo A jako $\lim_{n \rightarrow \infty} \rho_n(A)$. Kłopot: nie wiemy, czy ta granica istnieje.

Może więc spróbujemy z drugiej strony: zdefiniujemy prawdopodobieństwo jako abstrakcyjną funkcję, która ma wszystkie własności (i) – (iii).

6. AKSJOMATYKA KOŁMOGOROWA

Niech (Ω, \mathcal{F}) - ustalone. Wówczas funkcję $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ nazywamy prawdopodobieństwem, jeśli

- (i) $0 \leq \mathbb{P}(A) \leq 1$,
- (ii) $\mathbb{P}(\Omega) = 1$,
- (iii) jeśli $A_1, A_2, \dots \in \mathcal{F}$ są parami rozłączne, to

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Trójkę $(\Omega, \mathcal{F}, \mathbb{P})$ nazywamy *przestrzenią probabilistyczną*.

7. PRZYKŁADY

1. Rzut symetryczną monetą: $\Omega = \{O, R\}$, $\mathcal{F} = 2^\Omega = \{\{O\}, \{R\}, \Omega, \emptyset\}$, $\mathbb{P}(\{O\}) = 1/2$, $\mathbb{P}(\{R\}) = 1/2$, $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$.

2. Rzut niesymetryczną monetą: $\Omega = \{O, R\}$, $\mathcal{F} = 2^\Omega$, $\mathbb{P}(\{O\}) = p$, $\mathbb{P}(\{R\}) = 1 - p$, $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$. Tutaj p jest pewną ustaloną liczbą z przedziału $[0, 1]$.

3. Rzut kostką: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = 2^\Omega$, $\mathbb{P}(A) = |A|/6$.

4. **Schemat klasyczny (prawdopodobieństwo klasyczne)**. Załóżmy, że Ω jest zbiorem skończonym, $\mathcal{F} = 2^\Omega$ i wszystkie zdarzenia elementarne są jednakowo prawdopodobne. Wówczas, jak łatwo sprawdzić, dla $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

5. Z talii 52 kart losujemy jednocześnie pięć kart. Jakie jest prawdopodobieństwo, że wylosujemy cztery asy?

Jak już wiemy, Ω to pięcioelementowe kombinacje zbioru talii kart. Intuicja podpowiada, iż zdarzenia elementarne są równoprawdopodobne, a więc sensownym prawdopodobieństwem na Ω jest prawdopodobieństwo klasyczne.

Niech A - te podbiory, w których są cztery asy:

$$A = \{\{A\clubsuit, A\diamondsuit, A\heartsuit, A\spadesuit, *\} : * - \text{jedna z pozostałych 48 kart}\}.$$

Takich podzbiorów jest 48. A więc $|A| = 48$, $|\Omega| = \binom{52}{5}$.

6. Załóżmy, że $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$ - zbiór co najwyżej przeliczalny oraz p_1, p_2, \dots - liczby nieujemne o sumie 1. Wówczas możemy określić $\mathcal{F} = 2^\Omega$ oraz $\mathbb{P}(\{\omega_i\}) = p_i, i = 1, 2, \dots$. Wówczas, dla $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \sum_i 1_A(\omega_i)p_i,$$

gdzie 1_A to *funkcja wskaźnikowa (charakterystyczna)* bądź *indykator* zbioru A , zdefiniowany wzorem

$$1_A(x) = \begin{cases} 1 & \text{jeśli } x \in A, \\ 0 & \text{jeśli } x \notin A. \end{cases}$$

7. Prawdopodobieństwo geometryczne. W wielu sytuacjach, którymi będziemy się zajmować, doświadczenie losowe ma charakter ciągły. Najprostszym przykładem jest losowanie punktu z otwartego zbioru Ω , leżącego na prostej (lub na płaszczyźnie, czy ogólniej w przestrzeni \mathbb{R}^n) i mającego skończoną długość (pole powierzchni, miarę). Zbiorem takim może być np. odcinek, koło, kwadrat, kula, sześcian. Zgodnie z intuicją naturalnie jest przyjąć, iż prawdopodobieństwo zdarzenia $A \subseteq \Omega$ jest proporcjonalne do jego miary, czyli

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|},$$

gdzie $|\cdot|$ oznacza miarę zbioru. Pojawia się tu pewien techniczny problem, mianowicie jak zdefiniować σ -ciało \mathcal{F} ? Okazuje się, że nie można w naturalny sposób określić długości, pola powierzchni, czy objętości na wszystkich podzbiorach Ω , nie możemy więc przyjąć $\mathcal{F} = 2^\Omega$ i musimy się ograniczyć do mniejszego σ -ciała. Z reguły w takich sytuacjach rozpatruje się tzw. σ -ciało borelowskie $\mathcal{B}(\Omega)$, zdefiniowane jako najmniejsze σ -ciało zawierające wszystkie zbiory otwarte w Ω .

Na przykład, losowanie punktu z koła Ω o promieniu r można opisać przy pomocy przestrzeni probabilistycznej $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$, gdzie dla $A \in \mathcal{B}(\Omega)$,

$$\mathbb{P}(A) = \frac{|A|}{\pi r^2}.$$

W podobny sposób możemy również opisać losowanie punktu np. z okręgu czy sfery.

8. PODSTAWOWE WŁASNOŚCI PRAWDOPODOBIEŃSTWA

Poniżej sformułujemy kilka podstawowych faktów dotyczących prawdopodobieństwa. Przyjmujemy, że $(\Omega, \mathcal{F}, \mathbb{P})$ jest ustaloną przestrzenią probabilistyczną.

Twierdzenie 1. Niech $A, B, A_1, A_2, \dots \in \mathcal{F}$. Wówczas

(i) $\mathbb{P}(\emptyset) = 0$.

(ii) Jeśli A_1, A_2, \dots, A_n są parami rozłączne, to

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i).$$

(iii) $\mathbb{P}(A') = 1 - \mathbb{P}(A)$.

(iv) Jeśli $A \subseteq B$, to $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.

(v) Jeśli $A \subseteq B$, to $\mathbb{P}(A) \leq \mathbb{P}(B)$.

(vi) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

(vii) $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Twierdzenie 2 (Wzór włączeń i wyłączeń). Jeśli $A_1, A_2, \dots, A_n \in \mathcal{F}$, to

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) &= \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \dots \\ &\quad + (-1)^{n+1} \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n). \end{aligned}$$

Definicja 2. Załóżmy, że A_1, A_2, \dots jest ciągiem zdarzeń. Mówimy, że ciąg ten jest wstępujący, jeśli

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$$

oraz że jest zstępujący, jeśli

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$$

Twierdzenie 3 (Reguła ciągłości). Załóżmy, że $(A_n)_{n=1}^{\infty}$ jest ciągiem zdarzeń.

(i) Jeśli ciąg ten jest wstępujący, to

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right).$$

(ii) Jeśli ciąg ten jest zstępujący, to

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right).$$

9. PRAWDOPODOBIENSTWO WARUNKOWE

W praktyce z reguły jesteśmy zainteresowani nie tyle pojedynczym zdarzeniem, co kilkoma zdarzeniami i ich wzajemnymi związkami.

Przykłady:

1. Na podstawie ankiety przeprowadzonej na pewnym zbiorze klientów (oznaczymy go literą Ω) firma fonograficzna posiada dane na temat ich gustów muzycznych. Przypuśćmy, że kierownictwo jest zainteresowane pytaniem jak często fani jazzu lubią także muzykę klasyczną. Jeśli przez J oznaczymy zbiór tych ankietowanych,

którzy są fanami jazzu, a przez K zbiór tych ankietowanych, którzy są fanami muzyki klasycznej, interesująca nas częstość jest równa

$$\frac{|J \cap K|}{|J|} = \frac{|J \cap K|/|\Omega|}{|J|/|\Omega|}.$$

Zauważmy, że wyrażenia w liczniku i mianowniku to częstości poszczególnych zbiorów liczone względem całego zbioru Ω .

2. Przypuśćmy, że suma oczek przy dwóch rzutach kostką wynosi 4. Nie znamy jednak wyników poszczególnych rzutów. Jaka jest szansa zdarzenia $A = \{\text{przy pierwszym rzucie wypadły dwa oczka}\}$?

Informacja którą posiadamy oznacza, że zaszło zdarzenie $B = \{(1, 3), (2, 2), (3, 1)\}$. Intuicja podpowiada nam, że każde z trzech sprzyjających mu zdarzeń elementarnych powinno być tak samo prawdopodobne, a zatem szukane prawdopodobieństwo powinno wynosić $1/3$ (dwójce przy pierwszym rzucie sprzyja tylko jedno zdarzenie elementarne z B). Podobnie spodziewamy się, że wszystkie zdarzenia elementarne na przestrzeni

$$\Omega = \{(a, b) : a, b \in \{1, 2, 3, 4, 5, 6\}\},$$

opisującej dwa rzuty kostką, są jednakowo prawdopodobne. Zatem naturalnym modelem dla naszego doświadczenia jest $(\Omega, 2^\Omega, \mathbb{P})$, gdzie \mathbb{P} jest prawdopodobieństwem klasycznym

$$\mathbb{P}(C) = \frac{|C|}{36}, \text{ dla } C \subseteq \Omega.$$

Zauważmy teraz, że

$$\frac{1}{3} = \frac{1/36}{3/36} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Powyższe przykłady motywują następującą definicję.

Definicja 3. Niech A, B będą dwoma zdarzeniami, przy czym $\mathbb{P}(B) > 0$. Wówczas prawdopodobieństwem warunkowym zdarzenia A pod warunkiem zdarzenia B nazywamy liczbę

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

Uwaga: Pisząc $\mathbb{P}(A|B)$ milcząco zakładamy, że $\mathbb{P}(B) > 0$.

Przy ustalonym zbiorze B , prawdopodobieństwo warunkowe $\mathbb{P}(A|B)$ jako funkcja zbioru $A \in \mathcal{F}$ spełnia aksjomaty Kołmogorowa. W konsekwencji posiada więc wszystkie własności prawdopodobieństwa wprowadzone w paragrafie 8.

Twierdzenie 4 (Wzór łańcuchowy). Dla dowolnych zdarzeń A_1, \dots, A_n , spełniających warunek

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0,$$

zachodzi

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

Przykład: Losujemy po kolei trzy karty bez zwracania. Jakie jest prawdopodobieństwo, że wylosujemy trzy asy?

Niech A_i , $i = 1, 2, 3$, oznacza prawdopodobieństwo, że i -tą wylosowaną kartą jest as. Wiemy, że $\mathbb{P}(A_1) = 4/52$. Jeśli pierwszą wylosowaną kartą jest as, to przed drugim losowaniem w talii znajdują się trzy asy. Ponieważ tym razem losujemy spośród 51 kart, mamy

$$\mathbb{P}(A_2|A_1) = \frac{3}{51}.$$

Analogicznie

$$\mathbb{P}(A_3|A_1 \cap A_2) = \frac{2}{50}.$$

Stosując Twierdzenie 4, otrzymujemy

$$\mathbb{P}(\text{wylosujemy trzy asy}) = \mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50}.$$

W wielu zagadaniach modele probabilistyczne są zadane poprzez specyfikację prawdopodobieństw warunkowych interesujących nas zdarzeń pod warunkiem innych zdarzeń, których prawdopodobieństwa znamy. W takich sytuacjach przydatny jest tzw. wzór na prawdopodobieństwo całkowite. Zanim go sformulujemy, wprowadźmy następującą definicję.

Definicja 4. *Rozbiciem przestrzeni Ω nazywamy dowolną rodzinę zdarzeń $\{H_i\}_{i \in I}$, taką że $H_i \cap H_j = \emptyset$ dla $i \neq j$ oraz $\bigcup_{i \in I} H_i = \Omega$. Jeśli zbiór indeksujący I jest skończony (odp. przeliczalny), to rozbitcie nazywamy skończonym (odp. przeliczalnym).*

Twierdzenie 5 (Wzór na prawdopodobieństwo całkowite). *Dla dowolnego skończonego rozbitcia $\{H_1, H_2, \dots, H_n\}$ zbioru Ω na zbiory o dodatnim prawdopodobieństwie i dowolnego zdarzenia A zachodzi równość*

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|H_i)\mathbb{P}(H_i).$$

Analogiczny wzór zachodzi także dla rozbitcia na przeliczalną liczbę zdarzeń o dodatnim prawdopodobieństwie.

Przykład: Egzamin ustny przeprowadzany jest przez panów Dobrego i Złego. Egzamin u pana Dobrego zdaje 90% studentów, a u pana Złego zaledwie 10%. Jakie jest prawdopodobieństwo, że student zda egzamin jeśli prawdopodobieństwo, że trafi do pana Dobrego wynosi 2/3?

Niech D, Z oznaczają zdarzenia, że student trafił odpowiednio do pana Dobrego lub Złego, zaś OK zdarzenie, że student zda egzamin. Mamy $\mathbb{P}(D) = 2/3$, $\mathbb{P}(Z) = 1/3$ oraz $\mathbb{P}(OK|D) = 9/10$, $\mathbb{P}(OK|Z) = 1/10$. Zatem

$$\mathbb{P}(OK) = \mathbb{P}(OK|D)\mathbb{P}(D) + \mathbb{P}(OK|Z)\mathbb{P}(Z) = \frac{9}{10} \cdot \frac{2}{3} + \frac{1}{10} \cdot \frac{1}{3} = \frac{19}{30}.$$

Kolejne twierdzenie, blisko związane ze wzorem na prawdopodobieństwo całkowite, jest bardzo ważne w zastosowaniach.

Twierdzenie 6 (Wzór Bayesa). Niech $\{H_i\}_{i \in I}$ będzie przeliczalnym rozbiem Ω na zdarzenia o dodatnich prawdopodobieństwach. Wówczas, dla dowolnego zdarzenia A o dodatnim prawdopodobieństwie, zachodzi

$$\mathbb{P}(H_j|A) = \frac{\mathbb{P}(A|H_j)\mathbb{P}(H_j)}{\sum_{i \in I} \mathbb{P}(A|H_i)\mathbb{P}(H_i)}.$$

Przykład: Samochody sprzedawane przez pewną firmę pochodzą z dwóch fabryk: A (40%) oraz B (60%). Co dwudziesty samochód z fabryki A zawiera wadę fabryczną. To samo dotyczy co dziesiątego samochodu z fabryki B . Klient kupuje samochód, który okazuje się być wadliwy. Jakie jest prawdopodobieństwo, że pochodzi z fabryki A ?

Z warunków zadania otrzymujemy, że

$$\begin{aligned} \mathbb{P}(\text{samochód wadliwy}|A) &= \frac{1}{20}, \mathbb{P}(\text{samochód wadliwy}|B) = \frac{1}{10}, \\ \mathbb{P}(A) &= \frac{4}{10}, \mathbb{P}(B) = \frac{6}{10}, \end{aligned}$$

gdzie A, B oznaczają zdarzenia, że samochód pochodzi z fabryki odpowiednio A, B . Z wzoru Bayesa otrzymujemy

$$\begin{aligned} \mathbb{P}(A|\text{samochód wadliwy}) &= \frac{\mathbb{P}(\text{samochód wadliwy}|A)\mathbb{P}(A)}{\mathbb{P}(\text{samochód wadliwy}|A)\mathbb{P}(A) + \mathbb{P}(\text{samochód wadliwy}|B)\mathbb{P}(B)} \\ &= \frac{\frac{1}{20} \cdot \frac{4}{10}}{\frac{1}{20} \cdot \frac{4}{10} + \frac{1}{10} \cdot \frac{6}{10}} = \frac{1}{4}. \end{aligned}$$

10. NIEZALEŻNOŚĆ ZDARZEŃ

Przypuśćmy, że zdarzenia A, B spełniają warunek

$$(1) \quad \mathbb{P}(B|A) = \mathbb{P}(B).$$

Oznacza to, że dodatkowa wiedza, że zaszło zdarzenie A , nie wpływa na prawdopodobieństwo zdarzenia B . Można więc powiedzieć, że zdarzenie B jest niezależne od zdarzenia A . Powyższy warunek zapisuje się równoważnie jako

$$(2) \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

W szczególności widzimy, że jeśli (1) zachodzi oraz $\mathbb{P}(B) > 0$, to

$$\mathbb{P}(A|B) = \mathbb{P}(A),$$

czyli również zdarzenie A nie zależy od zdarzenia B . Zapis (2) ma tę zaletę, że lepiej niż (1) obrazuje symetrię sytuacji, dodatkowo ma sens także dla zdarzeń o zerowym prawdopodobieństwie. Naturalne jest więc przyjąć następującą definicję.

Definicja 5. Zdarzenia A, B nazywamy niezależnymi, jeśli

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Przykłady:

1. Rzucamy kostką. Rozpatrzmy zdarzenia: A - wypadła parzysta liczba oczek, B - liczba wyrzuconych oczek jest mniejsza niż 5, C - liczba wyrzuconych oczek

jest mniejsza niż 6. Oczywiście $\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 2/3$, $\mathbb{P}(C) = 5/6$. Zdarzenia A i B są niezależne natomiast zdarzenia A i C nie są niezależne. Rzeczywiście

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(\text{wypadły 2 lub 4 oczka}) = \frac{1}{3} = \mathbb{P}(A)\mathbb{P}(B), \\ \mathbb{P}(A \cap C) &= \mathbb{P}(A \cap B) = \frac{1}{3} \neq \mathbb{P}(A)\mathbb{P}(C).\end{aligned}$$

2. W ramach nagrody firma wykupiła dla pracowników dwa rodzaje wycieczek, w góry i nad morze. Wśród 12 pracowników rozdzielono w sposób losowy 8 wycieczek nad morze, z czego dwie w lipcu, a sześć w sierpniu oraz cztery wycieczki w góry, jedną w lipcu i trzy w sierpniu. Niech M oznacza zdarzenie, że ustalony pracownik wylosuje wycieczkę nad morze, zaś L - zdarzenie, że ten sam pracownik wylosuje termin lipcowy. Mamy $\mathbb{P}(M) = 8/12$, $\mathbb{P}(L) = 3/12$ oraz $\mathbb{P}(M \cap L) = 2/12$. Ponieważ $8/12 \cdot 3/12 = 2/12$, zdarzenia M i L są niezależne.

3. Losujemy jedną kartę z talii. Zdarzenie A , polegające na wylosowaniu karty starszej niż walet i zdarzenie B , polegające na wylosowaniu karty w kolorze trefl są niezależne. Rzeczywiście, $\mathbb{P}(A) = 12/52$ (wylosowana karta musi być damą, królem lub asem w jednym z czterech możliwych kolorów), $\mathbb{P}(B) = 1/4$ oraz $\mathbb{P}(A \cap B) = \mathbb{P}(\text{wylosowano damę, króla lub asa trefl}) = 3/52 = \mathbb{P}(A)\mathbb{P}(B)$.

Pytanie: Co się zmieni gdy do talii dodamy jednego jokera (przyjmujemy, że joker nie ma żadnego koloru)?

4. Rzucamy dwa razy monetą. Niech O_i oznacza zdarzenie, że w i -tym rzucie wypadł orzeł. Intuicyjnie uważamy te zdarzenia za niezależne (przynajmniej zakładając, że osoba rzucająca monetą nie oszukuje). W klasycznym modelu probabilistycznym dla monety symetrycznej, gdy prawdopodobieństwo każdej z czterech sekwencji (O, O) , (O, R) , (R, O) , (R, R) wynosi $1/4$, łatwo sprawdzić (por. z poprzednim przykładem), że rzeczywiście tak jest ($\mathbb{P}(O_1 \cap O_2) = \mathbb{P}(O_1)\mathbb{P}(O_2)$).

Zastanówmy się więc jak zdefiniować prawdopodobieństwo \mathbb{P} na zbiorze

$$\Omega = \{(O, O), (O, R), (R, O), (R, R)\},$$

tak aby prawdopodobieństwo wyrzucenia orła wynosiło p (zarówno w pierwszym, jak i drugim rzucie), a zdarzenia O_1, O_2 nadal były niezależne. Musimy w tym celu ustalić cztery liczby $p_{(O,O)}, p_{(O,R)}, p_{(R,R)}, p_{(R,O)}$. Chcemy aby

$$\begin{aligned}\mathbb{P}(\{(O, O), (O, R)\}) &= \mathbb{P}(\{(O, O), (R, O)\}) = p \quad \text{oraz} \\ \mathbb{P}(\{(O, O)\}) &= p^2,\end{aligned}$$

skąd dostajemy równania

$$\begin{aligned}p_{(O,O)} + p_{(O,R)} &= p_{(O,O)} + p_{(R,O)} = p, \\ p_{(O,O)} &= p^2.\end{aligned}$$

Zatem $p_{(R,O)} = p_{(R,R)} = p(1-p)$. Ponieważ $p_{(O,O)} + p_{(O,R)} + p_{(R,O)} + p_{(R,R)} = 1$, ostatecznie dostajemy

$$\begin{aligned}p_{(O,O)} &= p^2, \\ p_{(O,R)} &= p_{(R,O)} = p(1-p), \\ p_{(R,R)} &= (1-p)^2.\end{aligned}$$

Można również mówić o niezależności większej liczby zdarzeń. Definicja okazuje się jednak bardziej skomplikowana.

Definicja 6. Zdarzenia A_1, A_2, \dots, A_n nazywamy niezależnymi, jeśli dla dowolnych wskaźników $1 \leq i_1 < i_2 < \dots < i_k \leq n$, $k = 2, 3, \dots, n$, zachodzi równość

$$(3) \quad \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdot \dots \cdot \mathbb{P}(A_{i_k}).$$

Przykłady:

1. Losujemy liczbę od 1 do 90. Rozważmy zdarzenia A - wylosowana liczba jest podzielna przez 2, B - wylosowana liczba jest podzielna przez 3, C - wylosowana liczba jest podzielna przez 5. Wówczas, jak łatwo sprawdzić

$$\mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(B) = 1/3, \quad \mathbb{P}(C) = 1/5$$

oraz

$$\begin{aligned} \mathbb{P}(A \cap B) &= \frac{1}{6} = \mathbb{P}(A)\mathbb{P}(B), \\ \mathbb{P}(A \cap C) &= \frac{1}{10} = \mathbb{P}(A)\mathbb{P}(C), \\ \mathbb{P}(B \cap C) &= \frac{1}{15} = \mathbb{P}(B)\mathbb{P}(C), \\ \mathbb{P}(A \cap B \cap C) &= \frac{1}{30} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C). \end{aligned}$$

Zdarzenia A, B, C są zatem niezależne.

2. Można się zastanawiać, czy powyżej musieliśmy sprawdzać prawdopodobieństwa wszystkich czterech iloczynów zbiorów. Okazuje się, że tak, co ilustruje następujący przykład.

Trzech współlokatorów (Bartek, Czarek i Darek) decyduje się oddać butelki do sklepu. Zadanie wymaga udziału dwóch osób. Przygotowują więc cztery losy $\{\text{Bartek, Czarek, Darek, Za tydzień}\}$, aby zadecydować czy dwóch z nich zda butelki, a wylosowany zostanie w domu, czy też odłożą problem na przyszły tydzień. Rozważmy zdarzenia

$B = \{\text{Bartek, Za tydzień}\}$ - Bartek zostanie w domu,

$C = \{\text{Czarek, Za tydzień}\}$ - Czarek zostanie w domu,

$D = \{\text{Darek, Za tydzień}\}$ - Darek zostanie w domu.

Prawdopodobieństwo każdego ze zdarzeń B, C, D wynosi $1/2$. Ponadto

$$\begin{aligned} \mathbb{P}(B \cap C) &= \frac{1}{4} = \mathbb{P}(B)\mathbb{P}(C), \\ \mathbb{P}(B \cap D) &= \frac{1}{4} = \mathbb{P}(B)\mathbb{P}(D), \\ \mathbb{P}(C \cap D) &= \frac{1}{4} = \mathbb{P}(C)\mathbb{P}(D). \end{aligned}$$

Zatem każde dwa spośród zdarzeń B, C, D są niezależne (w takiej sytuacji mówimy, że zdarzenia B, C, D są niezależne parami). Zdarzenia B, C, D nie są jednak niezależne, gdyż

$$\mathbb{P}(B \cap C \cap D) = \mathbb{P}(\{\text{Za tydzień}\}) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(B)\mathbb{P}(C)\mathbb{P}(D).$$

Twierdzenie 7. Rozważmy zdarzenia A_1, A_2, \dots, A_n i oznaczmy $A_i^0 = A_i, A_i^1 = A_i'$. Wówczas następujące warunki są równoważne:

- (i) zdarzenia A_1, A_2, \dots, A_n są niezależne,
- (ii) dla każdego ciągu $\varepsilon_1, \dots, \varepsilon_n$, gdzie $\varepsilon_i \in \{0, 1\}$ ($i = 1, \dots, n$), zdarzenia $B_1 = A_1^{\varepsilon_1}, \dots, B_n = A_n^{\varepsilon_n}$, są niezależne,
- (iii) dla każdego ciągu $\varepsilon_1, \dots, \varepsilon_n$, gdzie $\varepsilon_i \in \{0, 1\}$ ($i = 1, \dots, n$), zachodzi

$$\mathbb{P}(A_1^{\varepsilon_1} \cap \dots \cap A_n^{\varepsilon_n}) = \mathbb{P}(A_1^{\varepsilon_1}) \cdot \dots \cdot \mathbb{P}(A_n^{\varepsilon_n}).$$

W szczególności, z powyższego twierdzenia wynika, że jeśli zdarzenia A, B są niezależne, to niezależne są także zdarzenia A', B' . Fakt ten pozwala uprościć nieco rachunki w przykładzie 4 powyżej.

11. SCHEMAT BERNOULLIEGO

Definicja 7. *Schematem Bernoulliego nazywamy ciąg niezależnych powtórzeń tego samego doświadczenia, w którym są możliwe dwa wyniki: jeden z nich nazywamy sukcesem (i prawdopodobieństwo jego zajścia oznaczamy przez p), a drugie - porażką (jego prawdopodobieństwo wynosi $q = 1 - p$). Pojedyncze doświadczenie nazywamy próbą Bernoulliego.*

Schemat Bernoulliego jest jednoznacznie określony przez podanie liczby prób (oznaczanej dalej literą n) i prawdopodobieństwa sukcesu p . Można też rozpatrywać schematy Bernoulliego z nieskończoną liczbą prób.

Przykłady:

1. Rzucamy 10 razy prawidłową monetą. Próbą Bernoulliego jest pojedynczy rzut monetą, jako sukces przyjmujemy wyrzucenie orła. Mamy $n = 10$, $p = 1/2$.

2. Rzucamy 5 razy prawidłową kostką. Próbą Bernoulliego jest pojedynczy rzut kostką, jako sukces przyjmujemy wyrzucenie co najwyżej 2 oczek. Mamy $n = 5$, $p = 1/3$.

3. Z urny, w której znajduje się 5 białych i 4 czarne kule, losujemy 20 razy ze zwracaniem po 2 kule. Próbą Bernoulliego jest pojedyncze losowanie dwóch kul, jako sukces bierzemy wylosowanie dwóch białych kul. Mamy $n = 20$, $p = \binom{5}{2} / \binom{9}{2}$.

Łatwo podać przestrzeń probabilistyczną modelującą schemat Bernoulliego składającego się z n prób i prawdopodobieństwie sukcesu p . Mianowicie,

$$\Omega = \{(a_1, a_2, \dots, a_n) : a_i \in \{0, 1\}, i = 1, 2, \dots, n\},$$

gdzie $a_i = 1$ (odp., $a_i = 0$) interpretujemy jako sukces (odp., porażkę) w i -tej próbie, $i = 1, 2, \dots, n$. Ponadto, bierzemy $\mathcal{F} = 2^\Omega$. Aby określić prawdopodobieństwo na (Ω, \mathcal{F}) , wystarczy określić je na zdarzeniach jednoelementowych (patrz przykład 6 ze strony 4). Kładziemy

$$\mathbb{P}(\{(a_1, a_2, \dots, a_n)\}) = p^{\sum_{i=1}^n a_i} (1-p)^{n-\sum_{i=1}^n a_i}.$$

Stąd łatwo wynika, iż prawdopodobieństwo uzyskania dokładnie k sukcesów w schemacie Bernoulliego składającego się z n prób wynosi

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

Przykłady:

1. Rzucamy 10 razy kostką. Jakie jest prawdopodobieństwo tego, że szóstka wypadnie raz lub dwa razy?

Mamy tu do czynienia ze schematem Bernoulliego składającego się z 10 prób. Próba Bernoulliego jest pojedynczy rzut kostką, a sukcesem jest wyrzucenie 6 oczek; zatem $p = 1/6$. Wobec tego

$$\begin{aligned} \mathbb{P}(\text{szóstka wypadnie raz lub dwa razy}) &= \mathbb{P}(\text{jeden sukces}) + \mathbb{P}(\text{dwa sukcesy}) \\ &= \binom{10}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^9 + \binom{10}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^8. \end{aligned}$$

2. Dany jest schemat Bernoulliego składający się z n prób, o prawdopodobieństwie sukcesu p . Jaka jest najbardziej prawdopodobna liczba sukcesów?

Oznaczmy

$$p_k = \mathbb{P}(\text{mamy dokładnie } k \text{ sukcesów}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Mamy

$$\frac{p_{k+1}}{p_k} = \frac{\binom{n}{k+1} p^{k+1} (1-p)^{n-(k+1)}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{(n-k)p}{(k+1)(1-p)}.$$

Powyższe wyrażenie jest większe niż 1 wtedy i tylko wtedy, gdy $k < (n+1)p - 1$; jest zaś mniejsze niż 1 wtedy i tylko wtedy, gdy $k > (n+1)p - 1$. Innymi słowy, do momentu $k = (n+1)p$ liczby p_k rosną, a potem maleją. Daje to następującą odpowiedź. Jeśli $(n+1)p$ jest liczbą całkowitą, to dwie liczby sukcesów są najbardziej prawdopodobne: $(n+1)p - 1$ oraz $(n+1)p$. Jeśli zaś $(n+1)p$ nie jest liczbą całkowitą, to najbardziej prawdopodobną liczbą sukcesów jest $\lfloor (n+1)p \rfloor$.

W przypadku, gdy liczba prób w schemacie Bernoulliego jest duża, obliczanie prawdopodobieństwa danej liczby sukcesów jest kłopotliwe. W przypadku gdy np jest „umiarkowane”, dobre przybliżenie takiego prawdopodobieństwa daje następujące twierdzenie.

Twierdzenie 8 (Poissona). *Jeśli $p_n \in [0, 1]$, $\lim_{n \rightarrow \infty} np_n = \lambda > 0$, to dla $k = 0, 1, 2, \dots$,*

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Powstaje naturalne pytanie, na ile powyższe przybliżenie jest „dobre”. Odpowiedź jest zawarta w następującym twierdzeniu.

Twierdzenie 9 (Oszacowanie błędu w przybliżeniu poissonowskim). *Niech S_n oznacza liczbę sukcesów w schemacie Bernoulliego składającym się z n prób i prawdopodobieństwie sukcesu p . Oznaczmy $\lambda = np$. Dla dowolnego $A \subset \{0, 1, 2, \dots\}$,*

$$\left| \mathbb{P}(S_n \in A) - \sum_{k \in A} \frac{\lambda^k}{k!} e^{-\lambda} \right| \leq \frac{\lambda^2}{n}.$$

Przykłady:

1. W urnie znajduje się 999 czarnych i 1 biała kula. Wyznaczyć przybliżone prawdopodobieństwo tego, że losując 500 razy ze zwracaniem wylosujemy 2 razy białą kulę.

Mamy tu do czynienia ze schematem 500 prób Bernoulliego (z których każda to pojedyncze losowanie z urny), o prawdopodobieństwie sukcesu $p = 1/1000$. Liczba prób $n = 500$ jest duża, $\lambda = np = 1/2$ jest umiarkowane, a więc na mocy twierdzenia

Poissona, szukane prawdopodobieństwo jest w przybliżeniu równe $\frac{(1/2)^2}{2!}e^{-1/2} = 0,076\dots$. Ponadto, jak widać z powyższego twierdzenia, błąd oszacowania jest nie większy niż $\lambda^2/n = 1/2000 = 0,0005$.

2. Artykuł liczy 10^5 znaków. Podczas wprowadzania artykułu do komputera, prawdopodobieństwo iż dany znak zostanie wpisany błędnie wynosi 0,0001. Jakie jest prawdopodobieństwo, że w artykule są co najmniej 2 błędy?

Widzimy, iż mamy do czynienia ze schematem Bernoulliego składającym się z $n = 10^5$ prób (k -ta z nich odpowiada wprowadzeniu k -tego znaku artykułu). Prawdopodobieństwo sukcesu (wprowadzenia znaku *błędnie*) wynosi $p = 0,0001$. Mamy, iż n jest duże, a $\lambda = np = 10$ jest umiarkowane; stąd możemy używać twierdzenia Poissona. Łatwiej jest pracować ze zdarzeniem przeciwnym do rozważanego: w artykule jest co najwyżej 1 błąd. Prawdopodobieństwo tego zdarzenia wynosi w przybliżeniu

$$\frac{10^0}{0!}e^{-10} + \frac{10^1}{1!}e^{-10} = 11e^{-10} = 0,0005\dots,$$

a więc rozważane w przykładzie prawdopodobieństwo wynosi około 0,9995. Błąd przybliżenia szacuje się przez $\lambda^2/n = 0,001$.

3. Z przedziału $[0, 2]$ wybieramy losowo 100 punktów. Jakie jest prawdopodobieństwo tego, że co najmniej jeden z nich będzie należał do odcinka $[0, 1/4]$?

Mamy schemat $n = 100$ prób Bernoulliego z prawdopodobieństwem sukcesu (wpadnięcie losowanego punktu do $[0, 1/4]$) wynoszącym $p = 1/8$. Mamy $\lambda = np = 12,5$ i zdarzenie przeciwne do badanego ma w przybliżeniu prawdopodobieństwo $e^{-12,5} = 0,000004\dots$. Błąd przybliżenia szacuje się przez $\lambda^2/n = 1,5625$. Widać więc, że otrzymany wynik jest bezwartościowy. Jest tak dlatego, iż λ , w porównaniu do n , nie jest „umiarkowane”.

12. ZMIENNE LOSOWE JEDNOWYMIAROWE

Jak już wiemy, matematycznym opisem doświadczenia losowego jest przestrzeń probabilistyczna $(\Omega, \mathcal{F}, \mathbb{P})$. Często jednak nie interesuje nas konkretny wynik $\omega \in \Omega$, ale pewne charakterystyki liczbowe wyniku. Na przykład, przy rzucie dwoma kostkami może nas interesować suma oczek; przy nieskończonym ciągu rzutów monetą może nas interesować numer losowania, w którym orzeł pojawił się po raz pierwszy, itp. Innymi słowy, często obiektem naszych zainteresowań jest pewna funkcja X określona na Ω , przyjmująca wartości rzeczywiste. Przy badaniu takiej funkcji, naturalnym pytaniem jest np. pytanie o prawdopodobieństwo tego, że $X \leq a$ (por. powyższe przykłady). W szczególności oznacza to, iż „ X nie przekracza a ” jest zdarzeniem, tzn.

$$X^{-1}((-\infty, a]) = \{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}.$$

Prowadzi to do następującego pojęcia.

Definicja 8. Funkcję $X : \Omega \rightarrow \mathbb{R}$ nazywamy zmienną losową o wartościach w \mathbb{R} , jeśli dla dowolnego $a \in \mathbb{R}$ zbiór $X^{-1}((-\infty, a])$ jest zdarzeniem, czyli $X^{-1}((-\infty, a]) \in \mathcal{F}$.

Uwaga: Gdy Ω jest zbiorem co najwyżej przeliczalnym i $\mathcal{F} = 2^\Omega$, to każda funkcja $X : \Omega \rightarrow \mathbb{R}$ jest zmienną losową.

Przykłady:

1. Rzucamy dwa razy monetą, X - liczba wyrzuconych orłów. Mamy $\Omega = \{(O, O), (O, R), (R, O), (R, R)\}$ i $X((O, O)) = 2$, $X((O, R)) = X((R, O)) = 1$, $X((R, R)) = 0$.

2. Rzucamy dwa razy kostką, X - suma oczek. Mamy $\Omega = \{(a, b) : a, b \in \{1, 2, 3, 4, 5, 6\}\}$, $X((a, b)) = a + b$.

3. Z odcinka $[0, 3]$ wybieramy punkt, X - jego odległość od najbliższej liczby całkowitej. Wówczas $\Omega = [0, 3]$ i dla $\omega \in \Omega$,

$$X(\omega) = \begin{cases} \omega & \text{jeśli } \omega \in [0, 1/2], \\ |\omega - 1| & \text{jeśli } \omega \in (1/2, 3/2], \\ |\omega - 2| & \text{jeśli } \omega \in (3/2, 5/2], \\ 3 - \omega & \text{jeśli } \omega \in (5/2, 3]. \end{cases}$$

Na zmiennych losowych (określonych na tej samej przestrzeni probabilistycznej) można wykonywać wszelkie (rozsądne...) działania: dodawanie, odejmowanie, mnożenie, dzielenie (o ile nie dzielimy przez 0) i jako wynik otrzymujemy nowe zmienne losowe. Ponadto, jeśli X jest zmienną losową, a $f : \mathbb{R} \rightarrow \mathbb{R}$ jest funkcją borelowską, to $f(X)$ też jest zmienną losową. Np., jeśli X, Y są zmiennymi losowymi, to $Z_1 = \sin X$, $Z_2 = 3 \sin X + Y^2$, $Z_3 = \frac{X}{Y^2+1}$ także są zmiennymi losowymi.

Przechodzimy teraz do pojęcia rozkładu zmiennej losowej. Zaczniemy od kilku przykładów.

1. Rzucamy trzy razy symetryczną monetą. Niech X oznacza liczbę wyrzuconych orłów. Korzystając ze schematu Bernoulliego obliczamy, iż

$$\mathbb{P}(X = 0) = \frac{1}{8}, \mathbb{P}(X = 1) = \frac{3}{8}, \mathbb{P}(X = 2) = \frac{3}{8}, \mathbb{P}(X = 3) = \frac{1}{8}.$$

Widzimy więc, że 0 oraz 3 są przyjmowane z prawdopodobieństwem $1/8$, a 1 i 2 - z prawdopodobieństwem $3/8$. Widać, że dostajemy pewien rozkład prawdopodobieństwa na prostej.

Niech teraz Y - liczba wyrzuconych reszek. Wówczas tak samo: 0 oraz 3 są przyjmowane przez zmienną Y z prawdopodobieństwem $1/8$, a 1 i 2 - z prawdopodobieństwem $3/8$. Tak więc dostajemy to samo prawdopodobieństwo na prostej.

2. Z koła o promieniu 1 losujemy punkt. Niech X oznacza odległość tego punktu od środka koła. Wówczas X przyjmuje wartości z przedziału $[0, 1]$. Dla $a \in [0, 1]$ mamy

$$\mathbb{P}(X \in [0, a]) = \frac{\pi a^2}{\pi} = a^2,$$

a więc potrafimy „mierzyć wielkość” przedziałów $[0, a]$. Okazuje się, iż podaną funkcję można rozszerzyć do prawdopodobieństwa określonego na prostej. Zależy ono oczywiście od zmiennej X .

Z powyższych dwóch przykładów widać, iż przy ustalonej zmiennej losowej X , prawdopodobieństwo z wyjściowej przestrzeni probabilistycznej daje się „przetransportować” do prawdopodobieństwa μ_X na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Prowadzi to do pojęcia *rozkładu zmiennej losowej*.

Definicja 9. *Rozkładem zmiennej losowej rzeczywistej X nazywamy prawdopodobieństwo μ_X na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, dane wzorem*

$$\mu_X(A) = \mathbb{P}(X \in A).$$

Uwaga: Istnieją różne zmienne losowe mające ten sam rozkład. Por. przykład 1 powyżej.

Przykłady:

1. Rzucamy raz kostką. Niech X oznacza liczbę oczek. Wówczas μ_X jest takim prawdopodobieństwem *skoncentrowanym* na zbiorze $\{1, 2, 3, 4, 5, 6\}$, że

$$\mu_X(\{k\}) = \frac{1}{6}.$$

Tak więc, dla $A \in \mathcal{B}(\mathbb{R})$,

$$\mu_X(A) = \frac{1}{6} \sum_{k=1}^6 1_A(k).$$

2. Powyższy rozkład jest przykładem *rozkładu dyskretnego*. Rozkład na prostej rzeczywistej nazwiemy dyskretnym, jeśli istnieje co najwyżej przeliczalny zbiór S taki, że $\mu(S) = 1$. Rozkład taki jest jednoznacznie wyznaczony przez masy (prawdopodobieństwa) punktów należących do S (ściślej, jednoelementowych podzbiorów S): istotnie, dla dowolnego $A \in \mathcal{B}(\mathbb{R})$,

$$\mu(A) = \sum_{k \in A} \mu(\{k\}).$$

3. Rozkład Bernoulliego $B(n, p)$. Jest to rozkład zmiennej losowej X określonej jako liczba sukcesów w schemacie Bernoulliego składającego się z n prób o prawdopodobieństwie sukcesu p . Dany jest on poprzez

$$\mu(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

4. Rozkład geometryczny z parametrem $p \in (0, 1)$, ozn. $\text{Geom}(p)$. Jest to rozkład zmiennej losowej X określonej jako numer próby, w której sukces pojawił się po raz pierwszy. Jest to rozkład skoncentrowany na zbiorze $\{1, 2, \dots, \infty\}$. Ponadto, mamy

$$\mu_X(\{k\}) = (1-p)^{k-1} p, \quad k = 1, 2, \dots$$

oraz

$$\mu_X(\{\infty\}) = 1 - \sum_{k=1}^{\infty} \mu_X(\{k\}) = 0.$$

Czasami rozkładem geometrycznym nazywamy rozkład zmiennej $Y = X - 1$, określony przez

$$\mu_Y(\{k\}) = (1-p)^k p, \quad k = 0, 1, 2, \dots$$

5. Rozkład Poissona z parametrem $\lambda > 0$, ozn. $\text{Pois}(\lambda)$. Jest to taki rozkład skoncentrowany na liczbach całkowitych nieujemnych, że

$$\mu(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Jak wiadomo z twierdzenia Poissona, jest to rozkład graniczny, będący granicą rozkładów Bernoulliego.

6. Przykład rozkładu ciągłego: rozkład jednostajny na odcinku $[a, b]$, ozn. $\mathcal{U}(a, b)$. Załóżmy, że losujemy liczbę X z odcinka $[a, b]$. Wówczas, z prawdopodobieństwa geometrycznego, mamy, dla przedziału $[c, d] \subset [a, b]$,

$$\mu_X([c, d]) = \mathbb{P}(X \in [c, d]) = \frac{|[c, d]|}{|[a, b]|} = \frac{d - c}{b - a} = \int_c^d \frac{1}{b - a} dx = \int_{[c, d]} \frac{1}{b - a} dx.$$

Ogólniej, jeśli A jest borelowskim podzbiorem $[a, b]$, to

$$\mu_X(A) = \mathbb{P}(X \in A) = \frac{|A|}{|[a, b]|} = \frac{1}{b - a} |A| = \int_A \frac{1}{b - a} dx.$$

Jeszcze ogólniej, gdy $A \subset \mathbb{R}$, to bierzemy $\mu_X(A) = \mu_X(A \cap [a, b])$.

7. Inny przykład rozkładu ciągłego. Załóżmy, że rzucamy monetą, dla której prawdopodobieństwo wypadnięcia orła wynosi $1/3$. Dalej, jeśli wypadnie orzeł, to losujemy punkt X z odcinka $[-2, 0)$, natomiast gdy wypadnie reszka - losujemy punkt X z odcinka $[0, 3]$. Argumentując jak w poprzednim przykładzie mamy, iż dla borelowskiego podzbioru $[0, 3]$,

$$\mu_X(A) = \mathbb{P}(X \in A) = \int_A \frac{2}{3} \cdot \frac{1}{3 - 0} dx,$$

a dla borelowskiego podzbioru A odcinka $[-2, 0)$,

$$\mu_X(A) = \int_A \frac{1}{3} \cdot \frac{1}{0 - (-2)} dx.$$

Ogólnie, gdy A jest podzbiorem borelowskim prostej, to

$$\mu_X(A) = \int_A g(x) dx,$$

gdzie

$$g(x) = \begin{cases} \frac{1}{6} & \text{jeśli } x \in [-2, 0), \\ \frac{2}{9} & \text{jeśli } x \in [0, 3], \\ 0 & \text{w pozostałych przypadkach.} \end{cases}$$

Powyższe dwa przykłady to przykłady rozkładów z gęstością bądź rozkładów ciągłych.

Definicja 10. *Zmienna losowa X ma rozkład ciągły, jeśli istnieje taka funkcja $g : \mathbb{R} \rightarrow \mathbb{R}_+$, że dla dowolnego zbioru $A \in \mathcal{B}(\mathbb{R})$,*

$$\mu_X(A) = \mathbb{P}(X \in A) = \int_A g(x) dx.$$

Wówczas funkcję g nazywamy gęstością rozkładu zmiennej X bądź gęstością zmiennej X .

Uwaga: Gęstość jednoznacznie wyznacza rozkład.

Przykłady - ciąg dalszy:

8. Przykład 6 możemy więc zapisać następująco: rozkład jednostajny $\mathcal{U}(a, b)$ to rozkład z gęstością

$$g(x) = \frac{1}{b - a} 1_{[a, b]}(x).$$

9. Rozkład wykładniczy z parametrem $\lambda > 0$, ozn. $\text{Exp}(\lambda)$. Jest to rozkład z gęstością

$$g(x) = \lambda e^{-\lambda x} 1_{[0, \infty)}(x).$$

10. Standardowy rozkład normalny, ozn. $\mathcal{N}(0, 1)$. Jest to rozkład o gęstości

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

Ogólniej, dla $a \in \mathbb{R}$ oraz $\sigma > 0$ definiujemy rozkład normalny o parametrach a, σ^2 (ozn. $\mathcal{N}(a, \sigma^2)$) jako rozkład o gęstości

$$g_{a, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

Dodatkowo dla $\sigma = 0$, definiujemy $\mathcal{N}(a, 0)$ jako rozkład jednopunktowy δ_a (tzw. delta Diraca w a), zadany wzorem

$$\delta_a(A) = 1_A(a) = \begin{cases} 1 & \text{gdy } a \in A \\ 0 & \text{w p.p.} \end{cases}$$

Jak widzimy $\mathcal{N}(a, \sigma^2)$ jest rozkładem ciągłym dla $\sigma > 0$ i dyskretnym dla $\sigma = 0$.

Uwaga: Rozkłady normalne należą do najważniejszych rozkładów w rachunku prawdopodobieństwa. Pojawiają się one niezwykle często w zastosowaniach, ze względu na fakt, że wiele występujących w przyrodzie wielkości ma rozkład w przybliżeniu normalny. Wykres gęstości rozkładu normalnego ciągłego to charakterystyczna krzywa o kształcie „dzwonu”, znana chociażby z opracowań popularnych, gdzie ilustruje np. rozkład wzrostu, wagi, ilorazu inteligencji czy innych cech w populacji. W dalszej części wykładu poznamy tzw. Centralne Twierdzenie Graniczne, które stanowi matematyczne wyjaśnienie faktu pojawiania się gęstości normalnej w tak wielu, często dość odległych problemach.

13. DYSTRYBUANTA ZMIENNEJ LOSOWEJ

Jak już wspomniano w poprzednim rozdziale, z reguły jesteśmy zainteresowani zdarzeniami typu $\{\omega \in \Omega: X(\omega) \leq a\} = \{X \leq a\}$, gdzie X jest zmienną losową, zaś a – liczbą rzeczywistą. Zdarzenia tego typu mają podstawowe znaczenie dla badania zmiennych losowych, w szczególności, jak zobaczymy nieco później, znajomość prawdopodobieństwa $\mathbb{P}(X \leq a)$ dla wszystkich $a \in \mathbb{R}$ wyznacza jednoznacznie rozkład zmiennej. Dlatego też wprowadza się następującą definicję.

Definicja 11. *Dystrybuantą zmiennej losowej $X: \Omega \rightarrow \mathbb{R}$ nazywamy funkcję $F_X: \mathbb{R} \rightarrow [0, 1]$ daną wzorem*

$$F_X(t) = \mathbb{P}(X \leq t).$$

Uwaga: Dystrybuanta zależy jedynie od rozkładu zmiennej losowej X , a zatem jest sens mówić o dystrybuancie rozkładu (a nie zmiennej).

Przykłady

1. Dystrybuanta zmiennej X o rozkładzie δ_a (czyli przyjmującej z prawdopodobieństwem 1 wartość a) jest dana wzorem

$$F_X(t) = \begin{cases} 0 & \text{dla } t < a \\ 1 & \text{dla } t \geq a. \end{cases}$$

2. Dystrybuanta zmiennej dwupunktowej, przyjmującej wartości 1, -1, każdą z prawdopodobieństwem 1/2 jest funkcja

$$F(t) = \begin{cases} 0 & \text{dla } t \in (-\infty, -1) \\ 1/2 & \text{dla } t \in [-1, 1) \\ 1 & \text{dla } t \in [1, \infty). \end{cases}$$

3. Jeśli Y jest zmienną o rozkładzie wykładniczym z parametrem 1, czyli o gęstości $g_Y(t) = e^{-t}1_{[0, \infty)}(t)$, to

$$F_Y(t) = \mathbb{P}(Y \leq t) = \int_{-\infty}^t g(x)dx = [-e^{-x}1_{[0, \infty)}(x)]_{x=-\infty}^{x=t} = (1 - e^{-t})1_{[0, \infty)}(t).$$

Powyższe przykłady sugerują, że dystrybuantami zmiennych losowych mogą być tylko funkcje szczególnego typu. Mówi o tym poniższe twierdzenie.

Twierdzenie 10. *Dystrybuanta F_X zmiennej losowej X ma następujące własności:*

- (i) F_X jest niemalejąca,
- (ii) $\lim_{t \rightarrow \infty} F_X(t) = 1$, $\lim_{t \rightarrow -\infty} F_X(t) = 0$,
- (iii) F_X jest prawostronnie ciągła.

Uwaga Czasami w literaturze, szczególnie tej nieco starszej, definiuje się dystrybuantę wzorem $F_X(t) = \mathbb{P}(X < t)$ (czyli używając ostrej nierówności). Tak zdefiniowana dystrybuanta posiada własności (i), (ii), ale własność (iii) zostaje zastąpiona warunkiem lewostronnej ciągłości.

Okazuje się, że powyższe twierdzenie można odwrócić, mianowicie każda funkcja spełniająca warunki (i)–(iii) jest dystrybuantą pewnej zmiennej losowej.

Twierdzenie 11. *Jeśli funkcja $F: \mathbb{R} \rightarrow \mathbb{R}$ spełnia warunki (i)–(iii), to istnieje przestrzeń probabilistyczna $(\Omega, \mathcal{F}, \mathbb{P})$ oraz zmienna losowa $X: \Omega \rightarrow \mathbb{R}$, taka że F jest dystrybuantą X . Co więcej rozkład zmiennej X jest wyznaczony jednoznacznie.*

Zatem w dystrybuancie zmiennej X „zakodowane” są wszystkie informacje o jej rozkładzie, w szczególności powinniśmy móc odczytać z niej czy zmienna X ma gęstość albo czy X jest zmienną dyskretną.

Przykład: Rozważmy dyskretną zmienną losową, przyjmującą wartości $t_1 < t_2 < \dots < t_n$, przy czym $\mathbb{P}(X = t_i) = p_i$ (zakładamy że zmienna nie przyjmuje żadnych innych wartości, czyli $\sum_{i=1}^n p_i = 1$). Wówczas dla $t < t_1$ mamy $F_X(t) = 0$, dla $t \geq t_n$ mamy $F_X(t) = 1$, zaś dla $t \in [t_j, t_{j+1})$ zachodzi $F_X(t) = \sum_{i=1}^j p_i$.

W szczególności widzimy, że F_X jest ciągła poza punktami t_i oraz posiada granice lewostronne dla każdego $t \in \mathbb{R}$. Oznaczmy $F_X(t-) = \lim_{x \rightarrow t-} F_X(x)$. Mamy

$$F_X(t_i) - F_X(t_i-) = p_i = \mathbb{P}(X = t_i),$$

oraz dla $t \notin \{t_1, t_2, \dots, t_n\}$,

$$F_X(t) - F_X(t-) = 0 = \mathbb{P}(X = t).$$

Okazuje się, że jest to ogólny fakt.

Twierdzenie 12. *Jeśli F_X jest dystrybuantą zmiennej losowej X , to dla $t \in \mathbb{R}$ zachodzi*

$$F_X(t-) = \mathbb{P}(X < t)$$

oraz

$$F_X(t) - F_X(t-) = \mathbb{P}(X = t).$$

W szczególności, jeśli F_X jest ciągła w punkcie t , to $\mathbb{P}(X = t) = 0$.

W przypadku rozkładów ciągłych, dystrybuanta może być użyta do znalezienia gęstości.

Przykład: Niech X będzie zmienną o rozkładzie $Exp(1)$, czyli z gęstością $g(x) = e^{-x}1_{[0,\infty)}(x)$. Wówczas dla $t \in \mathbb{R}$ mamy

$$F_X(t) = (1 - e^{-t})1_{[0,\infty)}(t).$$

Zauważmy, że dla $t \neq 0$ mamy $F'_X(t) = g(t)$. Nie jest to jednak prawdą dla $t = 0$, gdyż $F'_X(t)$ nie jest różniczkowalna w zerze.

W ogólności mamy następujące twierdzenie, które w wielu sytuacjach pozwala obliczyć gęstość zmiennej losowej, gdy znana jest jej dystrybuanta.

Twierdzenie 13. Niech F będzie dystrybuantą zmiennej losowej X .

1. Jeśli F nie jest ciągła, to X nie ma rozkładu ciągłego (tzn. nie ma gęstości).
2. Załóżmy, że F jest funkcją ciągłą. Jeśli F jest różniczkowalna poza skończonym zbiorem punktów, to funkcja

$$g(t) = \begin{cases} F'(t) & \text{jeśli } F'(t) \text{ istnieje,} \\ 0 & \text{w p.p.,} \end{cases}$$

jest gęstością zmiennej X .

Przykłady:

1. Rozważmy zmienną losową X o dystrybuancie

$$F(t) = \begin{cases} 0 & \text{dla } t \in (-\infty, 0), \\ 2t & \text{dla } t \in [0, 1/2), \\ 1 & \text{dla } t \in [1/2, \infty). \end{cases}$$

Funkcja F jest różniczkowalna wszędzie poza punktami $t = 0$ i $t = 1/2$. Ponadto $F'(t) = 0$ dla $t \in (-\infty, 0) \cup (1/2, \infty)$ oraz $F'(t) = 2$ dla $t \in (0, 1/2)$. Zatem funkcja

$$g(t) = 2 \cdot 1_{(0,1/2)}(t)$$

jest gęstością zmiennej X .

2. Należy podkreślić, że istnieją rozkłady, które nie są ani ciągłe ani dyskretne. Przykładowo rozkład μ , dany wzorem

$$\mu(A) = \frac{1}{2}|A \cap (0, 1)| + \frac{1}{2}1_A(3).$$

Dystrybuanta tego rozkładu to

$$F(t) = \begin{cases} 0 & \text{dla } t \in (-\infty, 0), \\ \frac{t}{2} & \text{dla } t \in [0, 1), \\ \frac{1}{2} & \text{dla } t \in [1, 3), \\ 1 & \text{dla } t \in [3, \infty) \end{cases}$$

Jak łatwo sprawdzić korzystając ze wzoru na prawdopodobieństwo całkowite, rozkład μ opisuje doświadczenie: „rzucamy symetryczną monetą; jeśli wypadnie orzeł

zwracamy jako wynik 3, w przeciwnym wypadku jako wynik zwracamy liczbę wylosowaną z przedziału $(0, 1)$ ”.

Jak wiemy, jeżeli X jest zmienną losową, a φ funkcją borelowską, to $Y = \varphi(X)$ też jest zmienną losową. Następane twierdzenia dotyczą zależności między gęstością zmiennej X oraz zmiennej Y , gdy funkcja φ jest dostatecznie regularna.

Twierdzenie 14. *Jeżeli X jest zmienną losową o gęstości f oraz X przyjmuje wartości w przedziale (a, b) , zaś funkcja $\varphi: (a, b) \rightarrow \mathbb{R}$ jest klasy C^1 i $\varphi'(x) \neq 0$ dla $x \in (a, b)$, to zmienna losowa $Y = \varphi(X)$ ma rozkład ciągły o gęstości*

$$g(y) = f(h(y))|h'(y)|1_{\varphi((a,b))}(y),$$

gdzie $h(s) = \varphi^{-1}(s)$.

Przykład: Zmienna X ma rozkład jednostajny na odcinku $(0, 4)$. Znaleźć rozkład zmiennej $Y = \sqrt{X}$.

Używając notacji z twierdzenia, mamy $a = 0, b = 4, f(x) = \frac{1}{4}1_{(0,4)}(x)$ oraz $\varphi(x) = \sqrt{x}$. Zatem $h(x) = x^2, \varphi((a, b)) = (0, 2)$. Gęstość Y dana jest wzorem

$$g(y) = \frac{1}{4}1_{(0,4)}(y^2) \cdot 2y \cdot 1_{(0,2)}(y) = \frac{1}{2}y1_{(0,2)}(y).$$

Rozważania dotyczące gęstości i dystrybuanty zakończymy definicją tzw. kwantyli, które odgrywają istotną rolę w statystyce.

Definicja 12. *Niech X będzie zmienną losową, zaś $p \in [0, 1]$. Kwantylem rzędu p zmiennej X nazywamy dowolną liczbę x_p , taką że*

$$\mathbb{P}(X \leq x_p) = F_X(x_p) \geq p$$

oraz

$$\mathbb{P}(X \geq x_p) \geq 1 - p.$$

Kwantyl rzędu $1/2$ nazywamy także medianą.

Przykłady:

1. Jeśli X jest zmienną przyjmującą dwie wartości $1, -1$, każdą z prawdopodobieństwem $1/2$, to dowolna liczba z przedziału $[-1, 1]$ jest medianą zmiennej X . Dla $p \in (0, 1/2)$ zmienna X ma jeden kwantyl równy -1 zaś dla $p \in (1/2, 1)$, jeden kwantyl, równy 1 . Kwantylami rzędu 0 są wszystkie liczby z przedziału $(-\infty, -1]$ zaś kwantylami rzędu 1 , liczby z przedziału $[1, \infty)$.

2. Standardowa zmienna normalna ma jedną medianę równą 0 . Podobnie, dla dowolnego $p \in (0, 1)$, zmienna ta ma dokładnie jeden kwantyl rzędu p , wyznaczony przez równość

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_p} e^{-x^2/2} dx = p.$$

14. PARAMETRY ROZKŁADÓW

14.1. **Wartość oczekiwana.** Zaczniemy od następującego przykładu.

Przykład. Załóżmy, iż ktoś proponuje nam następującą grę: rzucamy raz kostką, i jeśli wypadnie 1 oczko, to dostajemy 100 zł, natomiast w przeciwnym razie musimy zapłacić 30 zł. Czy w taką grę opłaca się grać? Czy na dłuższą metę wygrywamy?

Jeśli zagramy n razy w powyższą grę, to jedynka wypada średnio w $n/6$ wypadkach, a więc nasza wygrana po n grach to średnio

$$\frac{n}{6} \cdot 100 - \frac{5n}{6} \cdot 30 = -\frac{50n}{6} < 0,$$

a więc nie powinniśmy grać. Dodatkowo, jeśli X jest naszą wygraną w pojedynczej grze, to spodziewamy się, iż średnia X wynosi

$$\frac{1}{6} \cdot 100 + \frac{5}{6} \cdot (-30) = -\frac{50}{6} < 0.$$

Prowadzi to do następującej definicji.

Definicja 13. Załóżmy, że X jest zmienną losową o rozkładzie dyskretnym, skoncentrowanym na zbiorze $S \subset \mathbb{R}$ i niech $p_x = \mathbb{P}(X = x)$ dla $x \in S$. Mówimy, że wartość oczekiwana zmiennej losowej X jest skończona (bądź że zmienna losowa X jest całkowalna), jeśli $\sum_{x \in S} |x|p_x < \infty$. Wówczas określamy wartość oczekiwaną zmiennej X jako

$$\mathbb{E}X = \sum_{x \in S} xp_x.$$

Uwagi:

1. Wartość oczekiwana zmiennej losowej to, intuicyjnie, jej średnia wartość. Czasami, zamiast „wartość oczekiwana X ” będziemy mówić „średnia X ”.

2. Jeśli zbiór wartości zmiennej X jest skończony, to wartość oczekiwana zmiennej X jest skończona - sumy pojawiające się w definicji zawierają skończoną liczbę składników.

3. Wartość oczekiwana zmiennej losowej zależy tylko od rozkładu tej zmiennej.

Przykłady:

1. Jeśli X jest stała: $\mathbb{P}(X = a) = 1$ dla pewnego $a \in \mathbb{R}$, to $\mathbb{E}X = a \cdot 1 = a$.

2. Rzucamy raz kostką. Niech X oznacza liczbę wyrzuconych oczek. Wówczas $\mathbb{P}(X = k) = 1/6$ dla $k = 1, 2, \dots, 6$ i

$$\mathbb{E}X = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3\frac{1}{2}.$$

3. Załóżmy, że zmienna X ma rozkład Bernoulliego z parametrami n, p . Wówczas

$$\begin{aligned} \mathbb{E}X &= \sum_{k=0}^n k\mathbb{P}(X = k) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np. \end{aligned}$$

4. Załóżmy, że zmienna losowa X ma rozkład na $\{1, 2, \dots\}$ dany przez

$$\mathbb{P}(X = k) = \frac{1}{k(k+1)}, \quad k = 1, 2, \dots$$

Wówczas wartość oczekiwana X nie istnieje: mamy

$$\sum_{k=1}^{\infty} k\mathbb{P}(X = k) = \sum_{k=1}^{\infty} \frac{1}{k+1} = \infty.$$

5. Załóżmy, że zmienna losowa X ma rozkład na liczbach całkowitych różnych od 0, zadany przez

$$\mathbb{P}(X = k) = \frac{1}{2|k|(|k| + 1)}, \quad k \in \mathbb{Z}, k \neq 0.$$

Wówczas X nie jest całkowalna: mamy

$$\sum_{k \neq 0} |k| \mathbb{P}(X = k) = \sum_{k \neq 0} \frac{1}{2(|k| + 1)} = \infty.$$

Przejdźmy teraz do zmiennych losowych o rozkładach ciągłych.

Definicja 14. Załóżmy, że zmienna losowa X ma rozkład z gęstością g . Jeśli

$$\int_{\mathbb{R}} |x|g(x)dx < \infty,$$

to mówimy, że wartość oczekiwana X istnieje (bądź że zmienna losowa X jest całkowalna). Definiujemy wówczas wartość oczekiwaną X jako

$$\mathbb{E}X = \int_{\mathbb{R}} xg(x)dx.$$

Uwaga: Wartość oczekiwana zależy tylko od rozkładu zmiennej X .

Uwaga: Jeśli zmienna losowa X jest ograniczona, tzn. z prawdopodobieństwem 1 przyjmuje wartości z pewnego ograniczonego przedziału (a, b) , to istnieje jej wartość oczekiwana: istotnie,

$$\int_{\mathbb{R}} |x|g(x)dx \leq \int_{\mathbb{R}} \max\{|a|, |b|\}g(x)dx = \max\{|a|, |b|\}.$$

Oznaczenie: Czasami, zamiast mówić „ X jest całkowalną zmienną losową”, będziemy pisać „ $\mathbb{E}|X| < \infty$ ”.

Przykłady:

1. Załóżmy, że X ma rozkład jednostajny na odcinku (a, b) . Wówczas, jak wynika z powyższej uwagi, X jest całkowalna. Ponadto

$$\mathbb{E}X = \int_{\mathbb{R}} xg(x)dx = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

2. Załóżmy, że X ma rozkład $\mathcal{N}(0, 1)$. Wówczas

$$\int_{\mathbb{R}} |x| \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x \exp(-x^2/2) dx = \frac{2}{\sqrt{2\pi}} (-e^{-x^2/2}) \Big|_0^{\infty} = \frac{2}{\sqrt{2\pi}},$$

a więc wartość oczekiwana X jest skończona. Wynosi ona

$$\int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx = 0.$$

Twierdzenie 15 (Własności wartości oczekiwanej). Załóżmy, że X i Y są całkowalnymi zmiennymi losowymi.

(i) Jeśli $X \geq 0$, to $\mathbb{E}X \geq 0$.

(ii) Jeśli $X \leq Y$, to $\mathbb{E}X \leq \mathbb{E}Y$.

(iii) Mamy $\mathbb{E}X \leq \mathbb{E}|X|$.

(iv) Wartość oczekiwana jest operatorem liniowym: jeśli $a, b \in \mathbb{R}$, to zmienna $aX + bY$ jest zmienną całkowalną i

$$\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y.$$

(v) Jeśli $X = 1_A$, to $\mathbb{E}X = \mathbb{P}(A)$.

Uwaga: Własność (iv) uogólnia się, poprzez prostą indukcję, do następującej: jeśli X_1, X_2, \dots, X_n są całkowalnymi zmiennymi losowymi i $a_1, a_2, \dots, a_n \in \mathbb{R}$, to zmienna $a_1X_1 + a_2X_2 + \dots + a_nX_n$ też jest zmienną całkowalną i

$$\mathbb{E}(a_1X_1 + a_2X_2 + \dots + a_nX_n) = a_1\mathbb{E}X_1 + a_2\mathbb{E}X_2 + \dots + a_n\mathbb{E}X_n.$$

W szczególności,

$$\mathbb{E}(X_1 + X_2 + \dots + X_n) = \mathbb{E}X_1 + \mathbb{E}X_2 + \dots + \mathbb{E}X_n.$$

Przykłady:

1. Rzucamy 100 razy kostką i niech X oznacza sumę wyrzuconych oczek. Wówczas obliczenie wartości oczekiwanej X z definicji jest praktycznie niemożliwe - wymaga to wyznaczenia rozkładu zmiennej X . Ale jeśli zauważymy, że $X = X_1 + X_2 + \dots + X_{100}$, gdzie X_i to liczba oczek w i -tym rzucie, to mamy, iż

$$\mathbb{E}X = \mathbb{E}X_1 + \mathbb{E}X_2 + \dots + \mathbb{E}X_{100} = 100 \cdot 3\frac{1}{2} = 350.$$

2. W urnie znajduje się 5 białych i 10 czarnych kul. Losujemy ze zwracaniem 50 razy po jednej kuli. Niech X oznacza liczbę losowań, w których wyciągnięto białą kulę. Tak jak wyżej, wyznaczenie wartości oczekiwanej X bezpośrednio z definicji jest niesłychanie żmudne. Jeśli natomiast określimy

$$X = X_1 + X_2 + \dots + X_{50},$$

gdzie

$$X_i = 1_{\{\text{w } i\text{-tym losowaniu wyciągnięto białą kulę}\}} \\ = \begin{cases} 1 & \text{jeśli w } i\text{-tym losowaniu wyciągnięto białą kulę,} \\ 0 & \text{jeśli w } i\text{-tym losowaniu wyciągnięto czarną kulę,} \end{cases}$$

to mamy

$$\mathbb{E}X = \mathbb{E}X_1 + \mathbb{E}X_2 + \dots + \mathbb{E}X_{50} = 50 \cdot \mathbb{P}(\text{wyciągnięto białą kulę}) = \frac{50}{3}.$$

Przejdźmy teraz do sytuacji, gdy chcemy obliczyć wartość oczekiwaną funkcji pewnej zmiennej losowej.

Twierdzenie 16. Załóżmy, że $\phi : \mathbb{R} \rightarrow \mathbb{R}$ jest pewną funkcją borelowską.

(i) Załóżmy, że X ma rozkład dyskretny na zbiorze S i $p_x = \mathbb{P}(X = x)$ dla $x \in S$. Wówczas zmienna losowa $\phi(X)$ jest całkowalna wtedy i tylko wtedy, gdy $\sum_{x \in S} |\phi(x)|p_x < \infty$; wartość oczekiwana $\phi(X)$ wynosi wtedy

$$\mathbb{E}\phi(X) = \sum_{x \in S} \phi(x)p_x.$$

(ii) Załóżmy, że X ma rozkład ciągły z gęstością g . Wówczas zmienna losowa $\phi(X)$ jest całkowalna wtedy i tylko wtedy, gdy $\int_{\mathbb{R}} |\phi(x)|g(x)dx < \infty$; wartość oczekiwana wynosi wówczas

$$\mathbb{E}\phi(X) = \int_{\mathbb{R}} \phi(x)g(x)dx.$$

Przykłady:

1. Rzucamy raz kostką. Niech X oznacza liczbę wyrzuconych oczek. Wówczas

$$\mathbb{E}X^2 = \sum_{k=1}^6 k^2 \frac{1}{6} = \frac{91}{6}.$$

2. Z przedziału $[0, \pi/2]$ wybieramy losowo kąt X . Wówczas wartość oczekiwana sinusa tego kąta wynosi

$$\mathbb{E} \sin X = \int_0^{\pi/2} \sin x \cdot \frac{2}{\pi} dx = \frac{2}{\pi}.$$

Wartość oczekiwaną możemy też, w niektórych przypadkach, prosto wyrazić poprzez dystrybuantę F (a raczej funkcję $1 - F$). Zaczniemy od następującego przykładu.

Przykład: Załóżmy, że X jest całkowalną, dyskretną zmienną losową skoncentrowaną na liczbach całkowitych nieujemnych. Wówczas

$$\mathbb{E}X = \sum_{k=0}^{\infty} k\mathbb{P}(X = k) = \sum_{k=1}^{\infty} k\mathbb{P}(X = k).$$

Wyrazy powyższego szeregu możemy w następujący sposób ustawić „w trójkątną macierz”:

$$\begin{array}{cccc} \mathbb{P}(X = 1) & & & \\ \mathbb{P}(X = 2) & \mathbb{P}(X = 2) & & \\ \mathbb{P}(X = 3) & \mathbb{P}(X = 3) & \mathbb{P}(X = 3) & \\ \mathbb{P}(X = 4) & \mathbb{P}(X = 4) & \mathbb{P}(X = 4) & \mathbb{P}(X = 4) \\ \dots & & & \end{array}$$

W powyższym szeregu, sumowanie odbywa się najpierw wierszami, a następnie dodajemy do siebie otrzymane sumy. Ponieważ szereg ten posiada tylko nieujemne wyrazy, więc kolejność sumowania można zmieniać i nie ma to wpływu na wynik. Spróbujmy więc najpierw zsumować liczby występujące w poszczególnych kolumnach, a następnie dodać te sumy do siebie.

Suma liczb w pierwszej kolumnie to

$$\mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \dots = \mathbb{P}(X \geq 1) = \mathbb{P}(X > 0).$$

Dodając wyrazy stojące w drugiej kolumnie dostajemy

$$\mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \dots = \mathbb{P}(X \geq 2) = \mathbb{P}(X > 1),$$

itd, następne sumy będą wynosić $\mathbb{P}(X \geq 3) = \mathbb{P}(X > 2)$, $\mathbb{P}(X \geq 4) = \mathbb{P}(X > 3)$, ... Po zsumowaniu ich musimy dostać tyle, ile poprzednio, czyli $\mathbb{E}X$. Udowodniliśmy zatem

Twierdzenie 17. *Jeśli X jest jak wyżej, to*

$$\mathbb{E}X = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k) = \sum_{k=0}^{\infty} \mathbb{P}(X > k).$$

Poniższe twierdzenie stanowi rozszerzenie tego rezultatu. Jest ono prawdziwe dla dowolnych zmiennych losowych (także takich, których rozkład nie jest ani dyskretny, ani ciągły).

Twierdzenie 18. Niech X będzie zmienną losową nieujemną.

(i) Jeśli

$$\int_0^{\infty} \mathbb{P}(X > t) dt < \infty,$$

to X jest całkowalna i powyższa całka to wartość oczekiwana X .

(ii) Jeśli $p \in (0, \infty)$ i

$$p \int_0^{\infty} t^{p-1} \mathbb{P}(X > t) dt < \infty,$$

to X^p jest całkowalna i powyższa całka to wartość oczekiwana X^p .

Przykłady:

1. Załóżmy, że zmienna losowa X ma rozkład skoncentrowany na zbiorze $\{1, 2, \dots\}$, taki, że

$$\mathbb{P}(X = k) = \frac{2k + 1}{[k(k + 1)]^2}, \quad k = 1, 2, \dots$$

Wówczas

$$\mathbb{P}(X = k) = \frac{1}{k^2} - \frac{1}{(k + 1)^2} \quad \text{i} \quad \mathbb{P}(X \geq k) = \frac{1}{k^2},$$

a więc

$$\mathbb{E}X = \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}.$$

2. Załóżmy, że zmienna losowa X ma rozkład wykładniczy z parametrem $\lambda > 0$. Mamy, dla $t \geq 0$,

$$\mathbb{P}(X > t) = 1 - F_X(t) = e^{-\lambda t},$$

skąd wynika, iż

$$\mathbb{E}X = \int_0^{\infty} e^{-\lambda t} dt = \left(-\frac{1}{\lambda} e^{-\lambda t} \right) \Big|_0^{\infty} = \frac{1}{\lambda}.$$

3. Załóżmy, że zmienna losowa X ma rozkład z gęstością

$$g(x) = \frac{2}{x^2} 1_{[2, \infty)}(x).$$

Weźmy teraz liczbę $p \in (0, \infty)$ i zastanówmy się nad istnieniem wartości oczekiwanej zmiennej X^p . Przede wszystkim widzimy, iż

$$\mathbb{P}(X > t) = \begin{cases} 1 & \text{jeśli } t < 2, \\ \frac{2}{t} & \text{jeśli } t \geq 2. \end{cases}$$

Tak więc, na mocy powyższego twierdzenia, musimy zbadać całkę

$$\begin{aligned} p \int_0^{\infty} t^{p-1} \mathbb{P}(X > t) dt &= \int_0^2 p t^{p-1} dt + p \int_2^{\infty} t^{p-1} \mathbb{P}(X > t) dt \\ &= 2^p + 2p \int_2^{\infty} t^{p-2} dt. \end{aligned}$$

Powyższa całka jest zbieżna wtedy i tylko wtedy, gdy $p < 1$, i wynosi wówczas $2^p p / (1 - p)$. Tak więc wartość oczekiwana X^p istnieje wtedy i tylko wtedy, gdy $p < 1$ i wynosi $2^p / (1 - p)$.

4. Ostatni z rozważanych tu przykładów porusza problem wyznaczania wartości oczekiwanej zmiennej, której rozkład nie jest ani ciągły, ani dyskretny. Załóżmy,

że X ma rozkład jednostajny na $[0, 2]$ i obliczmy $\mathbb{E} \min\{X, 1\}$. Zmienna $\{X, 1\}$ ma rozkład mieszany. Mamy, iż dla $t \geq 0$,

$$\mathbb{P}(\min\{X, 1\} > t) = \mathbb{P}(X > t, 1 > t) = \begin{cases} 1 - \frac{t}{2} & \text{dla } t < 1, \\ 0 & \text{dla } t \geq 1. \end{cases}$$

Zatem

$$\mathbb{E} \min\{X, 1\} = \int_0^\infty \mathbb{P}(\min\{X, 1\} > t) dt = \int_0^1 \left(1 - \frac{t}{2}\right) dt = 3/4$$

i ogólniej, dla $p \in (0, \infty)$,

$$\mathbb{E} \min\{X, 1\}^p = p \int_0^1 t^{p-1} \left(1 - \frac{t}{2}\right) dt = \frac{p+2}{2(p+1)}.$$

Zadanie to można też było rozwiązać w inny sposób, stosując wzór na wartość oczekiwaną funkcji zmiennej losowej. Niech $\phi : \mathbb{R} \rightarrow \mathbb{R}$ będzie dana wzorem $\phi(x) = \min\{x, 1\}$. Wówczas

$$\begin{aligned} \mathbb{E} \min\{X, 1\} &= \mathbb{E}\phi(X) = \int_{\mathbb{R}} \phi(x)g(x)dx = \int_0^2 \min\{x, 1\} \cdot \frac{1}{2}dx \\ &= \int_0^1 x \cdot \frac{1}{2}dx + \int_1^2 1 \cdot \frac{1}{2}dx = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}. \end{aligned}$$

Podobnie, biorąc $\phi(x) = x^p$, $p \in (0, \infty)$,

$$\begin{aligned} \mathbb{E} \min\{X, 1\}^p &= \mathbb{E}\phi(X) = \int_0^2 \min\{x, 1\}^p \cdot \frac{1}{2}dx = \int_0^1 x^p \cdot \frac{1}{2}dx + \int_1^2 1 \cdot \frac{1}{2}dx \\ &= \frac{1}{2(p+1)} + \frac{1}{2} = \frac{p+2}{2(p+1)}. \end{aligned}$$

14.2. Wariancja. Kolejnym ważnym parametrem związanym z rozkładem zmiennej losowej jest jego wariancja.

Definicja 15. Załóżmy, że X jest zmienną losową spełniającą warunek $\mathbb{E}|X| < \infty$ oraz $\mathbb{E}(X - \mathbb{E}X)^2 < \infty$. Wówczas wariancją zmiennej losowej X nazywamy liczbę

$$D^2X = \text{Var}X = \mathbb{E}(X - \mathbb{E}X)^2.$$

Odchyleniem standardowym (rozkładu) zmiennej X nazywamy pierwiastek z wariancji:

$$\sigma_X = \sqrt{D^2X}.$$

Uwagi:

1. Aby określić wariancję, wystarczy zakładać, że $\mathbb{E}X^2 < \infty$ (mówimy wówczas, że X jest całkowalna z kwadratem). Pociąga to za sobą żadaną skończoność obu powyższych wartości oczekiwanych.

2. Jeśli zmienna losowa X jest ograniczona, to jej wariancja jest skończona.

3. Wariancję można wyrazić innym wzorem, często bardziej przydatnym w konkretnych obliczeniach:

$$D^2X = \text{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

4. Wariancja zależy tylko od rozkładu zmiennej losowej.

Wariancja zmiennej losowej to średnia kwadratu odchylenia od średniej. Tak więc, intuicyjnie, jeśli zmienna X posiada małą wariancję, to spodziewamy się, że przyjmuje ona wartości dość blisko swojej średniej; natomiast, gdy wariancja jest duża, to zmienna posiada „duży rozrzut” bądź „duże wahanie”.

Zilustrujemy to na następującym przykładzie.

Przykład: Załóżmy, że zmienna X ma rozkład

$$\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

Ponadto, niech Y ma rozkład skoncentrowany na $\{-100, 100\}$, taki, że

$$\mathbb{P}(Y = -100) = \mathbb{P}(Y = 100) = \frac{1}{2}.$$

Wówczas $\mathbb{E}X = \frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot 1 = 0$ oraz, analogicznie, $\mathbb{E}Y = 0$; tak więc obie te zmienne mają tę samą średnią. W oczywisty sposób Y „ma większy rozrzut”. I rzeczywiście

$$D^2X = \mathbb{E}(X - 0)^2 = \mathbb{E}X^2 = 1, \quad D^2Y = \mathbb{E}(Y - 0)^2 = 10000.$$

Dalsze przykłady:

1. Rzucamy kostką i niech X oznacza liczbę wyrzuconych oczek. Wówczas, jak już wiemy,

$$\mathbb{E}X = 3\frac{1}{2} \quad \text{oraz} \quad \mathbb{E}X^2 = \frac{91}{6},$$

zatem

$$\text{Var}X = \frac{91}{6} - \left(3\frac{1}{2}\right)^2 = 15\frac{1}{6} - 12\frac{1}{4} = 2\frac{11}{12}, \quad \sigma_X = 1,7078\dots$$

2. Przypuśćmy, że X ma rozkład jednostajny na odcinku $[a, b]$. Jak wiemy, $\mathbb{E}X = (a + b)/2$; ponadto,

$$\mathbb{E}X^2 = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3} = \frac{b^2 + ab + a^2}{3},$$

a zatem

$$D^2X = \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}, \quad \sigma_X = \frac{b-a}{2\sqrt{3}}.$$

Twierdzenie 19 (Własności wariancji). *Załóżmy, że zmienna X jest całkowalna z kwadratem. Wówczas*

a) $\text{Var}X \geq 0$, przy czym równość ma miejsce, gdy X jest stała z prawdopodobieństwem 1, tzn. istnieje taka liczba $a \in \mathbb{R}$, że $\mathbb{P}(X = a) = 1$.

b) $\text{Var}(bX) = b^2 \text{Var}X$ dla dowolnej liczby $b \in \mathbb{R}$.

c) $\text{Var}(X + c) = \text{Var}X$ dla dowolnej liczby $c \in \mathbb{R}$.

Przykład: parametry rozkładu normalnego. Niech $m \in \mathbb{R}$, $\sigma > 0$ i załóżmy, że X ma rozkład $\mathcal{N}(0, 1)$. Wówczas

$$\mathbb{P}(\sigma X + m \leq t) = \mathbb{P}\left(X \leq \frac{t-m}{\sigma}\right) = \int_{-\infty}^{(t-m)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Stosując podstawienie $x = (y - m)/\sigma$, $dx = dy/\sigma$, dostajemy

$$F_{\sigma X + m}(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y - m)^2}{2\sigma^2}\right) dy.$$

Stąd od razu widać, że $\sigma X + m$ ma rozkład normalny $\mathcal{N}(m, \sigma^2)$.

Na mocy powyższych twierdzeń mamy, iż

$$\mathbb{E}(\sigma X + m) = \sigma \mathbb{E}X + m = m,$$

$$D^2(\sigma X + m) = D^2(\sigma X) = \sigma^2 D^2 X = \sigma^2,$$

gdzież, całkując przez części,

$$\begin{aligned} D^2 X &= \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}X^2 = \int_{\mathbb{R}} x \cdot \frac{x}{\sqrt{2\pi}} \exp(-x^2/2) dx \\ &= \int_{\mathbb{R}} x \left(-\frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \right)' dx \\ &= -x \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \Big|_{-\infty}^{\infty} - \int_{\mathbb{R}} \left(-\frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \right) dx = 0 + 1 = 1. \end{aligned}$$

Wobec tego, parametry m , σ rozkładu normalnego to jego średnia i odchylenie standardowe.

Zdefiniujmy teraz parametry rozkładów, które grają ważną rolę w statystyce. Niech X będzie pewną zmienną losową.

Definicja 16. Dla $p \in (0, \infty)$, momentem (absolutnym) rzędu p zmiennej X nazywamy liczbę $\mathbb{E}|X|^p$, o ile wartość oczekiwana jest skończona (w przeciwnym razie przyjmujemy, że moment jest nieskończony).

Definicja 17. Załóżmy, że $\mathbb{E}|X|^3 < \infty$. Współczynnikiem asymetrii (skośności) zmiennej X nazywamy liczbę

$$\alpha_3 = \frac{\mathbb{E}(X - \mathbb{E}X)^3}{(D^2 X)^{3/2}} = \frac{\mathbb{E}(X - \mathbb{E}X)^3}{\sigma_X^3}.$$

Definicja 18. Załóżmy, że $\mathbb{E}|X|^4 < \infty$. Kurtozą (współczynnikiem spłaszczenia) zmiennej X nazywamy liczbę

$$\alpha_4 = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{\sigma_X^4} - 3.$$

15. CHARAKTERYSTYKI LICZBOWE PRÓBK

W praktyce nie są nam znane ani wszystkie wartości ani nawet rozkłady zmiennych losowych. Przykładowo, analitycy chcący poznać rozkład miesięcznych wydatków na rozrywkę wśród mieszkańców Warszawy, nie mają możliwości zebrania oraz przeanalizowania danych dotyczących każdego mieszkańca i muszą bazować na wynikach ankiety przeprowadzonej na losowej próbie mieszkańców stolicy. Również fizycy czy inżynierowie dokonując obciążonego losowym błędem pomiaru pewnej wielkości fizycznej, nie znają dokładnego rozkładu błędu (który może być traktowany jako cecha charakterystyczna urządzenia pomiarowego), dlatego często dokonują wielokrotnych pomiarów tej samej wielkości, aby na ich podstawie uzyskać jej jak najlepsze przybliżenie. W obu sytuacjach informacja, która jest dostępna, to tzw. próbka, czyli ciąg liczb X_1, X_2, \dots, X_n z którego chcielibyśmy odzyskać

informacje na temat interesującej nas zmiennej losowej X . Sposobami pobierania próbek oraz wnioskowania na ich podstawie zajmuje się statystyka matematyczna, nie będziemy więc w tym momencie zgłębiać tego zagadnienia. Ograniczymy się jedynie do informacji, że wiele metod bazuje na charakterystykach liczbowych próbki, analogicznych do zdefiniowanych w poprzednich rozdziałach charakterystyk zmiennych losowych.

Zauważmy, że z próbką X_1, X_2, \dots, X_n możemy związać rozkład prawdopodobieństwa na prostej (tzw. rozkład empiryczny), zdefiniowany jako

$$(4) \quad \mu_n(A) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}(A) = \frac{|\{i \leq n: X_i \in A\}|}{n},$$

informujący nas jaka część wszystkich obserwacji znajduje się w zbiorze A . Intuicyjnie można się spodziewać, że rozkład ten (przynajmniej dla odpowiednio dużych n) powinien dość dobrze przybliżać rozkład interesującej nas zmiennej losowej (zjawisko to wyjaśnimy dokładniej na kolejnych wykładach, gdy będziemy mówić o tzw. prawach wielkich liczb). Analogicznie, charakterystyki liczbowe rozkładu μ możemy uznać za dostępne nam przybliżenia odpowiednich charakterystyk nieznanego rozkładu zmiennej losowej.

Definicja 19. Dystrybuantą empiryczną próbki X_1, X_2, \dots, X_n nazywamy funkcję $F: \mathbb{R} \rightarrow [0, 1]$, zadaną wzorem

$$F_n(t) = \mu_n((-\infty, t]) = \frac{|\{i \leq n: X_i \leq t\}|}{n}.$$

Dystrybuanta empiryczna jest zatem dystrybuantą rozkładu empirycznego próbki.

Mając dystrybuantę empiryczną możemy w szczególności zdefiniować odpowiadające jej kwantyle.

Definicja 20. Kwantylem rzędu p z próbki X_1, \dots, X_n nazywamy dowolną liczbę x_p , taką że

$$\begin{aligned} \mu_n((-\infty, x_p]) &\geq p \\ \mu_n([x_p, \infty)) &\geq 1 - p. \end{aligned}$$

Kwantyl rzędu $1/2$ nazywamy także medianą.

Uwaga: Często, jeśli odpowiedni kwantyl nie jest zdefiniowany jednoznacznie, tzn. istnieje nieskończenie wiele liczb spełniających powyższe nierówności, wyróżniamy najmniejszą z nich.

Przykład:

Przypuśćmy, że nasza próbka składa się z liczb 2, 3, 5, 9. Rozkład empiryczny jest dany wzorem

$$\mu(A) = \frac{1}{4}1_A(2) + \frac{1}{4}1_A(3) + \frac{1}{4}1_A(5) + \frac{1}{4}1_A(9),$$

zaś dystrybuanta empiryczna,

$$F(t) = \begin{cases} 0 & \text{dla } t \in (-\infty, 2) \\ \frac{1}{4} & \text{dla } t \in [2, 3) \\ \frac{1}{2} & \text{dla } t \in [3, 5) \\ \frac{3}{4} & \text{dla } t \in [5, 9) \\ 1 & \text{dla } t \in [9, \infty). \end{cases}$$

Medianą jest dowolna liczba z przedziału $[3, 5]$, zaś kwantylem rzędu $3/4$ jest dowolna liczba z przedziału $[5, 9]$.

Podobnie możemy zdefiniować średnią i wariancję z próbki.

Definicja 21. Średnią z próbki X_1, X_2, \dots, X_n nazywamy liczbę

$$m = \frac{X_1 + X_2 + \dots + X_n}{n},$$

czyli średnią arytmetyczną liczb X_1, X_2, \dots, X_n .

Definicja 22. Wariancją z próbki X_1, X_2, \dots, X_n nazywamy liczbę

$$s^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2,$$

gdzie m jest średnią z próbki.

Uwagi:

1. Tak jak w przypadku dystrybuanty empirycznej i kwantyli, średnia i wariancja z próbki to po prostu wartość oczekiwana i wariancja rozkładu empirycznego.

2. W praktyce (np. w popularnych arkuszach kalkulacyjnych) do przybliżania wariancji zmiennej losowej na podstawie próbki często używa się raczej wyrażenia

$$\frac{1}{n-1} \sum_{i=1}^n (X_i - m)^2,$$

ponieważ jednak wyjaśnienie dlaczego w mianowniku pojawia się liczba $(n-1)$ wymagałoby wprowadzenia dodatkowego aparatu statystyki matematycznej, nie będziemy wnikać w szczegóły.

Przykład: Średnia empiryczna próbki 2, 4, 2, 6 wynosi

$$m = \frac{1}{4}(2 + 4 + 2 + 6) = 3.5,$$

zaś wariancja

$$s^2 = \frac{1}{4}[(2 - 3.5)^2 + (4 - 3.5)^2 + (2 - 3.5)^2 + (6 - 3.5)^2] = 2.75.$$

16. ROZKŁAD ŁĄCZNY ZMIENNYCH LOSOWYCH

W interesujących modelach probabilistycznych mamy z reguły do czynienia z kilkoma powiązаныmi ze sobą zmiennymi losowymi. Aby opisać ich wzajemne zależności wygodnie jest rozpatrywać je łącznie i traktować jako wektor losowy (zmienną losową o wartościach w przestrzeni wielowymiarowej). Jeśli na przykład w modelu występują zmienne X_1, X_2, \dots, X_n (gdzie $X_i: \Omega \rightarrow \mathbb{R}$), możemy utworzyć z

nich pojedynczy wektor losowy $X = (X_1, X_2, \dots, X_n)$ o wartościach w przestrzeni \mathbb{R}^n . Na zmienne losowe wielowymiarowe tego typu możemy rozszerzyć niektóre (choć nie wszystkie!) znane nam już definicje.

Definicja 23. Rozkładem wektora losowego $X = (X_1, X_2, \dots, X_n)$ nazywamy prawdopodobieństwo μ_X na $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, dane wzorem

$$\mu_X(A) = \mathbb{P}(X \in A)$$

dla $A \in \mathcal{B}(\mathbb{R}^n)$.

Uwagi:

1. Powyższa definicja jest analogiczna do definicji rozkładu rzeczywistej zmiennej losowej.

2. Rozkład wektora $X = (X_1, X_2, \dots, X_n)$ nazywamy również łącznym rozkładem zmiennych X_1, X_2, \dots, X_n . Możemy z niego odtworzyć rozkłady poszczególnych zmiennych X_i podstawiając za A odpowiednio dobrany zbiór. Jeśli na przykład chcemy obliczyć $\mu_{X_i}(B) = \mathbb{P}(X_i \in B)$ dla pewnego zbioru $B \subseteq \mathbb{R}$, definiujemy

$$A = \underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{i-1} \times B \times \underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{n-i}$$

i korzystamy z równości

$$\mathbb{P}(X_i \in B) = \mathbb{P}((X_1, X_2, \dots, X_n) \in A) = \mu_X(A).$$

3. Rozkłady zmiennych X_1, \dots, X_n nazywane są rozkładami brzegowymi wektora losowego X .

Przykład: Rzucamy dwukrotnie symetryczną monetą. Niech X_i ($i = 1, 2$) będą zmiennymi losowymi przyjmującymi wartość 1, jeśli w i -tym rzucie wypadł orzeł oraz 0 w przeciwnym przypadku. Mamy zatem

$$\mu_{X_1} = \mu_{X_2} = \frac{1}{2}\delta_0(A) + \frac{1}{2}\delta_1(A).$$

Rozkład łączny zmiennych X_1, X_2 dany jest wzorem

$$\mu_{(X_1, X_2)}(A) = \frac{1}{4}(\delta_{(0,0)}(A) + \delta_{(0,1)}(A) + \delta_{(1,0)}(A) + \delta_{(1,1)}(A))$$

dla $A \subseteq \mathbb{R}^2$, borelowskiego. Rozkład ten skoncentrowany jest w punktach $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$. Niech teraz X_3 będzie zmienną losową daną wzorem $X_3 = 1 - X_1$. Jak łatwo zauważyć, $\mu_{X_3} = \mu_{X_1} = \mu_{X_2}$. Tym niemniej rozkład $\mu_{(X_1, X_3)}$ jest skoncentrowany w dwóch punktach $(0, 1)$, $(1, 0)$; dokładniej

$$\mu_{(X_1, X_3)} = \frac{1}{2}\delta_{(0,1)} + \frac{1}{2}\delta_{(1,0)} \neq \mu_{(X_1, X_2)}.$$

Jak zatem widzimy, rozkład łączny zawiera w sobie dużo więcej informacji na temat zmiennych X_1, X_2, \dots, X_n niż rozkłady brzegowe.

W dalszej części ograniczymy się do badania rozkładów łącznych dwóch zmiennych losowych. Warto jednak podkreślić, iż podane definicje w naturalny sposób przenoszą się na rozkłady łączne większej liczby zmiennych.

Zacniemy od zdefiniowania dwuwymiarowych rozkładów dyskretnych i ciągłych. Definicje te są analogiczne do znanych nam już definicji dla zmiennych jednowymiarowych.

Definicja 24. Wektor losowy (X, Y) nazwiemy dyskretnym, jeśli istnieje przeliczalny zbiór $S \subseteq \mathbb{R}^2$, taki że $\mu_{(X, Y)}(S) = 1$.

Definicja 25. Wektor losowy (X, Y) nazwiemy ciągłym, jeśli istnieje gęstość, czyli funkcja $g: \mathbb{R}^2 \rightarrow [0, \infty)$, taka że dla każdego zbioru $A \in \mathcal{B}(\mathbb{R}^2)$,

$$\mu_{(X, Y)}(A) = \iint_A g(x, y) dx dy.$$

Przykłady:

1. Wektory (X_1, X_2) oraz (X_1, X_3) z poprzedniego przykładu mają rozkłady dyskretnie.

2. Jak łatwo sprawdzić, funkcja $g(x, y) = \frac{1}{2\pi} \exp(-(x^2 + y^2)/2)$ jest gęstością rozkładu prawdopodobieństwa na \mathbb{R}^2 .

3. Losujemy punkt (X, Y) z koła D o środku w punkcie $(2, 2)$ i promieniu R . Wektor (X, Y) jest wówczas wektorem ciągłym o gęstości

$$g(x, y) = \frac{1}{\pi R^2} 1_D(x, y) = \begin{cases} \frac{1}{\pi R^2} & \text{jeśli } (x-2)^2 + (y-2)^2 \leq R^2, \\ 0 & \text{w p.p.} \end{cases}$$

Twierdzenie 20. Załóżmy, że (X, Y) ma rozkład z gęstością g . Wówczas rozkłady brzegowe zmiennych X oraz Y także są ciągłe, co więcej, odpowiednie gęstości wynoszą

$$g_X(x) = \int_{\mathbb{R}} g(x, y) dy, \quad g_Y(y) = \int_{\mathbb{R}} g(x, y) dx.$$

Ogólniej, jeśli n -wymiarowy wektor ma gęstość g , to jego k -ta współrzędna ma rozkład z gęstością g_k , którą uzyskujemy odcałkowując g po wszystkich zmiennych różnych niż x_k :

$$g_k(x_k) = \int_{\mathbb{R}^{n-1}} g(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_{k-1} dx_{k+1} \dots dx_n.$$

Podamy kilka użytecznych faktów - uogólnień dobrze znanych nam wzorów z przypadku jednowymiarowego.

Twierdzenie 21. (i) Jeśli (X, Y) jest dyskretną zmienną losową skoncentrowaną na zbiorze S i $\phi: \mathbb{R}^2 \rightarrow \mathbb{R}$ jest funkcją borelowską, to

$$\mathbb{E}\phi(X, Y) = \sum_{(x, y) \in S} \phi(x, y) \mathbb{P}((X, Y) = (x, y))$$

(o ile wartość oczekiwana istnieje).

(ii) Jeśli (X, Y) jest zmienną losową o rozkładzie ciągłym (z gęstością g) i $\phi: \mathbb{R}^2 \rightarrow \mathbb{R}$ jest funkcją borelowską, to

$$\mathbb{E}\phi(X, Y) = \iint_{\mathbb{R}^2} \phi(x, y) g(x, y) dx dy$$

(o ile wartość oczekiwana istnieje).

Przykłady:

1. Załóżmy, że (X, Y) ma rozkład dany przez

$$\mathbb{P}((X, Y) = (k, l)) = 2^{-k-l}, \quad k, l = 1, 2, \dots,$$

i spróbujmy wyznaczyć $\mathbb{E}(X + Y)$.

Stosujemy powyższe twierdzenie dla $\phi(x, y) = x + y$. Mamy więc

$$\mathbb{E}(X + Y) = \sum_{k,l=1}^{\infty} (k+l)2^{-k-l} = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} (k+l)2^{-k-l}.$$

Zajmijmy się najpierw wewnętrzną sumą. Zauważmy, iż dla $x \in [0, 1)$ zachodzi wzór

$$\sum_{n=1}^{\infty} nx^{n-1} = \left(\sum_{n=1}^{\infty} x^n \right)' = \left(\frac{x}{1-x} \right)' = \frac{1}{(1-x)^2}.$$

Wobec tego

$$\begin{aligned} \sum_{l=1}^{\infty} (k+l)2^{-k-l} &= \sum_{l=1}^{\infty} k2^{-k-l} + \sum_{l=1}^{\infty} l2^{-k-l} \\ &= k2^{-k} + 2^{-k-1} \sum_{l=1}^{\infty} l \cdot \left(\frac{1}{2} \right)^{l-1} \\ &= k2^{-k} + 2^{-k-1} \cdot 4 = k2^{-k} + 2^{-k+1}, \end{aligned}$$

a zatem

$$\mathbb{E}(X + Y) = \sum_{k=1}^{\infty} (k2^{-k} + 2^{-k+1}) = \frac{1}{2} \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2} \right)^{k-1} + 2 = \frac{1}{2} \cdot 4 + 2 = 4.$$

2. Załóżmy, że (X, Y) ma rozkład z gęstością

$$g(x, y) = 24xy \cdot 1_{\{(x,y): x \geq 0, y \geq 0, x+y \leq 1\}}.$$

Obliczymy $\mathbb{E}(X^2 + 1)$. Otóż, stosując powyższe twierdzenie dla $\phi(x, y) = x^2 + 1$ i rozbijając całość łączną na całki iterowane, dostajemy

$$\mathbb{E}(X^2 + 1) = \int \int_{\mathbb{R}^2} \phi(x, y)g(x, y)dxdy = \int_0^1 \int_0^{1-x} (x^2 + 1) \cdot 24xy dydx.$$

Zajmijmy się najpierw wewnętrzną całką. Mamy

$$\int_0^{1-x} (x^2 + 1) \cdot 24xy dy = 24(x^2 + 1)x \cdot \frac{(1-x)^2}{2} = 12x^5 - 24x^4 + 24x^3 - 24x^2 + 12x.$$

Całkując powyższy wielomian po przedziale $[0, 1]$ dostajemy $\mathbb{E}(X^2 + 1) = 1, 2$.

Zdefiniujemy teraz dwie ważne charakterystyki liczbowe par zmiennych losowych.

Definicja 26. Niech (X, Y) będzie dwuwymiarową zmienną losową o całkowalnych współrzędnych (czyli, innymi słowy, założmy, że X, Y są rzeczywistymi, całkowalnymi zmiennymi losowymi) takich, że $\mathbb{E}|XY| < \infty$. Kowariancją zmiennych X, Y nazywamy liczbę

$$\text{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y).$$

Jeśli dodatkowo $\text{Var}X, \text{Var}Y > 0$, definiujemy współczynnik korelacji liniowej zmiennych X, Y , jako

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}X \cdot \text{Var}Y}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Uwaga: Kowariancja i współczynnik korelacji zmiennych X, Y nie zmieniają się, gdy te zmienne „przesuniemy”, tzn., jeśli $X_1 = X + a$ oraz $Y_1 = Y + b$, gdzie $a, b \in \mathbb{R}$, to $\text{Cov}(X, Y) = \text{Cov}(X_1, Y_1)$ oraz $\rho(X, Y) = \rho(X_1, Y_1)$. Jest to natychmiastowy wniosek z liniowości wartości oczekiwanej.

Przypuśćmy teraz, że zmienne losowe X, Y określone są na dyskretnej przestrzeni probabilistycznej $\Omega = \{1, 2, \dots, n\}$ z prawdopodobieństwem klasycznym $\mathbb{P}(A) = |A|/n$. Z definicji, zmienne losowe X, Y to funkcje na Ω , a zatem możemy je interpretować jako ciągi liczb: $X = (x_1, x_2, \dots, x_n)$, $Y = (y_1, y_2, \dots, y_n)$ (gdzie $x_j = X(j)$, $y_j = Y(j)$, $j = 1, 2, \dots, n$). Spróbujmy podać geometryczną interpretację współczynnika korelacji zmiennych X, Y . Na podstawie powyższej uwagi o przesunięciach możemy założyć, że $\mathbb{E}X = (x_1 + \dots + x_n)/n = 0$, $\mathbb{E}Y = (y_1 + \dots + y_n)/n = 0$. Wówczas

$$\rho(X, Y) = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i}{\sqrt{n^{-1} \sum_{i=1}^n x_i^2} \sqrt{n^{-1} \sum_{i=1}^n y_i^2}} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}.$$

Korzystając z algebry liniowej i interpretując ciągi $(x_i), (y_i)$ jako wektory w przestrzeni \mathbb{R}^n , widzimy, że współczynnik korelacji to cosinus kąta między tymi wektorami. W szczególności $|\rho(X, Y)| \leq 1$. Co więcej, jeśli $\rho(X, Y)$ jest bliskie 1, to zmienne X, Y (po przeskalowaniu) są w pewnym sensie bliskie, jeśli zaś $\rho(X, Y)$ jest bliskie -1 , odpowiadające im wektory wskazują w „przeciwnych kierunkach”. Takie intuicyjne rozumienie współczynnika korelacji przenosi się na przypadek dowolnych zmiennych losowych, dzięki tzw. nierówności Schwarz’a – twierdzeniu o bardzo wielu zastosowaniach w różnych działach matematyki.

Twierdzenie 22. Niech $X, Y: \Omega \rightarrow \mathbb{R}$ będą zmiennymi losowymi takimi, że $\mathbb{E}X^2 < \infty$, $\mathbb{E}Y^2 < \infty$. Wówczas

$$|\mathbb{E}XY| \leq (\mathbb{E}X^2)^{1/2} (\mathbb{E}Y^2)^{1/2}.$$

Co więcej, równość w powyższej nierówności zachodzi wtedy i tylko wtedy, gdy istnieją takie dwie liczby a, b (z których co najmniej jedna jest różna od zera), że $\mathbb{P}(aX = bY) = 1$.

Jako wniosek z nierówności Schwarz’a, dostajemy

Twierdzenie 23. Jeśli $X, Y: \Omega \rightarrow \mathbb{R}$ są zmiennymi losowymi o skończonej i niezerowej wariancji, to $|\rho(X, Y)| \leq 1$. Ponadto, jeśli $|\rho(X, Y)| = 1$, to istnieją takie liczby $a, b \in \mathbb{R}$, że $Y = aX + b$.

Można też wprowadzić dystrybuantę zmiennej dwuwymiarowej (X, Y) .

Definicja 27. Dystrybuantą dwuwymiarowej zmiennej losowej (X, Y) nazywamy funkcję $F_{(X, Y)}: \mathbb{R}^2 \rightarrow [0, 1]$ daną wzorem

$$F_{(X, Y)}(s, t) = \mathbb{P}(X \leq s, Y \leq t).$$

Wprowadzimy teraz odpowiedniki wielowymiarowe wartości oczekiwanej i wariancji.

Definicja 28. Jeśli (X, Y) jest dwuwymiarową zmienną losową, to:

a) jeśli X oraz Y są całkowalne, to wartością oczekiwaną $\mathbb{E}(X, Y)$ zmiennej (X, Y) nazywamy wektor $(\mathbb{E}X, \mathbb{E}Y)$.

b) jeśli X oraz Y są całkowalne z kwadratem, to macierzą kowariancji zmiennej (X, Y) nazywamy macierz

$$\begin{bmatrix} \text{Var}X & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \text{Var}Y \end{bmatrix}.$$

Analogiczne wzory zachodzą dla zmiennych losowych o wartościach w \mathbb{R}^d , $d \geq 3$: wartość oczekiwaną definiujemy jako wektor $(\mathbb{E}X_1, \mathbb{E}X_2, \dots, \mathbb{E}X_d)$, a macierz kowariancji - jako $(\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d}$.

Wartość oczekiwana i macierz kowariancji zachowują się w sposób regularny przy przekształceniach liniowych. Mamy następujący fakt.

Twierdzenie 24. Załóżmy, że $X = (X_1, X_2, \dots, X_n)$ jest zmienną losową, a A - macierzą $m \times n$.

(i) Jeśli X posiada skończoną wartość oczekiwaną, to zmienna AX też ma skończoną wartość oczekiwaną i $\mathbb{E}(AX) = A\mathbb{E}X$.

(ii) Jeśli istnieje macierz kowariancji Q_X zmiennej X , to istnieje też macierz kowariancji zmiennej AX i wynosi ona $Q_{AX} = A Q_X A^t$.

Twierdzenie 25. Załóżmy, że X jest zmienną losową o rozkładzie ciągłym w \mathbb{R}^n , m jest pewnym wektorem w \mathbb{R}^n , a T jest pewnym nieosobliwym przekształceniem \mathbb{R}^n w siebie (tzn. $\det T \neq 0$, gdzie T utożsamiamy z jego macierzą). Wówczas zmienna $TX + m$ ma gęstość

$$g_{TX+m}(x) = \frac{1}{|\det T|} g_X(T^{-1}(x - m)).$$

17. NIEZALEŻNE ZMIENNE LOSOWE

Definicja 29. Zmienne losowe $X_1, \dots, X_n: \Omega \rightarrow \mathbb{R}$ nazywamy niezależnymi, jeśli dla dowolnego ciągu B_1, B_2, \dots, B_n zbiorów borelowskich zachodzi

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1)\mathbb{P}(X_2 \in B_2)\dots\mathbb{P}(X_n \in B_n).$$

Uwagi:

1. Łatwo wykazać, że zmienne X_1, X_2, \dots, X_n są niezależne wtedy i tylko wtedy, gdy dla dowolnych zbiorów borelowskich B_1, B_2, \dots, B_n zdarzenia $\{X_1 \in B_1\}, \{X_2 \in B_2\}, \dots, \{X_n \in B_n\}$ są łącznie niezależne.

2. Podobnie jak w przypadku niezależności zdarzeń, należy pamiętać, że niezależność parami zmiennych X_1, X_2, \dots, X_n nie implikuje niezależności łącznej tych zmiennych. Przykładowo, jeśli rzucamy dwa razy monetą oraz X_i ($i = 1, 2$) przyjmują wartość 1, jeśli w i -tym rzucie wypadł orzeł oraz wartość -1 , jeśli w i -tym rzucie wypadła reszka, natomiast $X_3 = X_1 \cdot X_2$, to $\{X_1, X_2\}, \{X_1, X_3\}, \{X_2, X_3\}$ stanowią pary niezależnych zmiennych losowych, jednak X_1, X_2, X_3 oczywiście nie są niezależne.

Sprawdzanie niezależności zmiennych losowych wprost z definicji jest dość skomplikowane. Szczęśliwie, w przypadku zmiennych dyskretnych i ciągłych można podać prostsze kryteria niezależności.

Twierdzenie 26. Niech X_1, X_2, \dots, X_n będą dyskretnymi zmiennymi losowymi, zaś S_{X_i} takimi zbiorami przeliczalnymi, że $\mathbb{P}(X_i \in S_{X_i}) = 1$. Wówczas zmienne X_1, X_2, \dots, X_n są niezależne, wtedy i tylko wtedy, gdy dla każdego ciągu x_1, x_2, \dots, x_n takiego, że $x_i \in S_{X_i}$, $i = 1, 2, \dots, n$, zachodzi równość

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) \cdot \dots \cdot \mathbb{P}(X_n = x_n).$$

Przykłady:

1. Przeprowadzamy n prób Bernoulliego z prawdopodobieństwem sukcesu p . Niech

$$X_i = 1_{\{\text{sukces w } i\text{-tym rzucie}\}} = \begin{cases} 1 & \text{jeśli w } i\text{-tym rzucie zaszedł sukces,} \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Wówczas zmienne X_1, \dots, X_n są niezależne. Aby to wykazać, wystarczy sprawdzić warunek z twierdzenia dla każdego ciągu binarnego. Niech więc $x_1, \dots, x_n \in \{0, 1\}$. Wówczas z definicji schematu Bernoulliego,

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = p^{\sum_{j=1}^n x_j} (1-p)^{n-\sum_{j=1}^n x_j}$$

oraz

$$\begin{aligned} & \mathbb{P}(X_i = x_i) \\ &= \sum_{y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n=0}^1 \mathbb{P}(X_j = y_j \text{ dla } j \neq i, X_i = x_i) \\ &= \sum_{y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n=0}^1 p^{x_i} (1-p)^{1-x_i} \prod_{j \neq i} p^{y_j} (1-p)^{1-y_j} = p^{x_i} (1-p)^{1-x_i} \end{aligned}$$

(gdyż $p^1(1-p)^0 + p^0(1-p)^1 = 1$), a zatem

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) \cdot \dots \cdot \mathbb{P}(X_n = x_n).$$

2. Rzucamy n razy kostką. Niech X_i oznacza liczbę oczek wyrzuconych w i -tym rzucie. Wówczas dla dowolnego ciągu x_1, x_2, \dots, x_n , gdzie $x_i \in \{1, 2, \dots, 6\}$, mamy

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{1}{6^n}.$$

Z drugiej strony, jeśli A to zbiór tych ciągów $\omega = (\omega_1, \omega_2, \dots, \omega_n) \in \{1, 2, \dots, 6\}^n$, że $\omega_i = x_i$, to

$$\mathbb{P}(X_i = x_i) = \frac{|A|}{6^n} = \frac{6^{n-1}}{6^n} = \frac{1}{6}.$$

Zatem dla dowolnego ciągu x_1, x_2, \dots, x_n jak wyżej,

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) \cdot \dots \cdot \mathbb{P}(X_n = x_n),$$

czyli zmienne X_1, X_2, \dots, X_n są niezależne.

Odpowiednie twierdzenie dla zmiennych ciągłych wyraża się poprzez gęstości i rozkład łączny zmiennych losowych.

Twierdzenie 27. Niech $X_1, X_2, \dots, X_n: \Omega \rightarrow \mathbb{R}$ będą ciągłymi zmiennymi losowymi o gęstościach g_1, g_2, \dots, g_n , odpowiednio. Wówczas zmienne X_1, X_2, \dots, X_n są niezależne wtedy i tylko wtedy, gdy funkcja $g: \mathbb{R}^n \rightarrow [0, \infty)$, dana wzorem

$$g(x_1, x_2, \dots, x_n) = g_1(x_1)g_2(x_2) \cdot \dots \cdot g_n(x_n),$$

jest gęstością rozkładu $\mu_{(X_1, X_2, \dots, X_n)}$.

Przykłady

1. Niech (X, Y) będzie punktem wylosowanym z koła o środku w punkcie $(0, 0)$ i promieniu 1. Jak łatwo sprawdzić, wektor losowy (X, Y) ma wówczas gęstość

$$f(x, y) = \begin{cases} \frac{1}{\pi} & \text{jeśli } x^2 + y^2 \leq 1, \\ 0 & \text{jeśli } x^2 + y^2 > 1. \end{cases}$$

Zatem dla $t \in (0, 1)$,

$$\begin{aligned} F_X(t) &= \mathbb{P}(X \leq t) = \mathbb{P}\left((X, Y) \in (-\infty, t) \times \mathbb{R}\right) = \int_{-\infty}^t \int_{-\infty}^{\infty} f(x, y) dy dx \\ &= \frac{1}{\pi} \int_{-1}^t \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy dx = \frac{2}{\pi} \int_{-1}^t \sqrt{1-x^2} dx \end{aligned}$$

oraz $F_X(t) = 0$ dla $t \leq -1$ i $F_X(t) = 1$ dla $t \geq 1$. Różniczkując dystrybuantę widzimy, że zmienna X ma gęstość $g(x) = 2\pi^{-1}\sqrt{1-x^2}1_{(-1,1)}(x)$. Przez symetrię, g jest także gęstością zmiennej Y . Ponieważ $g(x)g(y) \neq f(x, y)$, zmienne X, Y nie są niezależne (co jest intuicyjnie jasne, skoro spełniają warunek $X^2 + Y^2 \leq 1$ i przyjmują dowolne wartości z przedziału $[-1, 1]$).

2. Niech teraz (X, Y) będzie punktem wylosowanym z kwadratu o wierzchołkach w punktach $(\pm 1, \pm 1)$. Wektor (X, Y) ma gęstość

$$f(x, y) = \frac{1}{4}1_{[-1,1]}(x)1_{[-1,1]}(y).$$

Łatwo sprawdzić całkując powyższą gęstość po odpowiednich współrzędnych, że zmienne X, Y mają tę samą gęstość $g(x) = 2^{-1}1_{[-1,1]}(x)$. Mamy $f(x, y) = g(x)g(y)$, zatem zmienne X, Y są niezależne.

3. Niech (X, Y) będzie dwuwymiarową zmienną losową o gęstości

$$f(x, y) = e^{-y}1_{\{0 \leq x \leq 1, y \geq 0\}}.$$

Wówczas zmienna X spełnia

$$\begin{aligned} \mu_X(A) &= \mathbb{P}(X \in A) = \mathbb{P}((X, Y) \in A \times \mathbb{R}) \\ &= \int_A \int_{\mathbb{R}} 1_{[0,1]}(x)e^{-y}1_{[0,\infty)}(y) dy dx = \int_A 1_{[0,1]}(x) dx, \end{aligned}$$

a zatem ma rozkład o gęstości $g_1(x) = 1_{[0,1]}(x)$ (czyli rozkład jednostajny $\mathcal{U}(0, 1)$). Podobnie, zmienna Y spełnia

$$\begin{aligned} \mu_Y(A) &= \mathbb{P}(Y \in A) = \mathbb{P}((X, Y) \in \mathbb{R} \times A) \\ &= \int_{\mathbb{R}} \int_A 1_{[0,1]}(x)e^{-y}1_{[0,\infty)}(y) dy dx = \int_A e^{-y}1_{[0,\infty)}(y) dy, \end{aligned}$$

czyli ma gęstość $g_2(y) = e^{-y}1_{[0,\infty)}(y)$ (rozkład wykładniczy $\text{Exp}(1)$). Ponieważ $f(x, y) = g_1(x)g_2(y)$, to zmienne X, Y są niezależne.

Poniższe dość skomplikowane w zapisie twierdzenie mówi, że funkcje od niezależnych zmiennych losowych pozostają niezależne. Przykładowo, jeśli X, Y, T, Z są niezależnymi zmiennymi losowymi, zaś $f: \mathbb{R}^2 \rightarrow \mathbb{R}, g, h: \mathbb{R} \rightarrow \mathbb{R}$ są funkcjami borelowskimi, to zmienne $f(X, Y), g(T), h(Z)$ są niezależne.

Twierdzenie 28. *Rozważmy zmienne losowe*

$$X_{1,1}, X_{1,2}, \dots, X_{1,k_1}, X_{2,1}, X_{2,2}, \dots, X_{2,k_2}, \dots, X_{n,1}, X_{n,2}, \dots, X_{n,k_n}$$

oraz funkcje borelowskie $\varphi_i: \mathbb{R}^{k_i} \rightarrow \mathbb{R}$, $i = 1, 2, \dots, n$. *Jeśli powyższe zmienne są niezależne, to również zmienne*

$$Y_1 = \varphi_1(X_{1,1}, X_{1,2}, \dots, X_{1,k_1}),$$

$$Y_2 = \varphi_2(X_{2,1}, X_{2,2}, \dots, X_{2,k_2}),$$

...

$$Y_n = \varphi_n(X_{n,1}, X_{n,2}, \dots, X_{n,k_n})$$

są niezależne.

Wrócimy teraz do charakterystyk liczbowych zmiennych losowych, by przekonać się, że niezależność może nam w istotny sposób pomóc w ich obliczaniu.

Twierdzenie 29. *Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi, posiadającymi wartość oczekiwaną. Wówczas zmienna $X_1 \cdot X_2 \cdot \dots \cdot X_n$ również ma wartość oczekiwaną i zachodzi równość*

$$\mathbb{E}(X_1 X_2 \cdot \dots \cdot X_n) = \mathbb{E}X_1 \cdot \mathbb{E}X_2 \cdot \dots \cdot \mathbb{E}X_n.$$

Przykład: Rzucamy 100 razy kostką. Niech X_i oznacza liczbę oczek wyrzucanych w i -tym rzucie oraz $X = X_1 \cdot X_2 \cdot \dots \cdot X_n$. Zmienne X_i są niezależne i $\mathbb{E}X_i = 3.5$, a zatem $\mathbb{E}X = (3.5)^{100}$.

Zauważmy, że jeśli zmienne X, Y są niezależne, to „przesunięte” zmienne $X - \mathbb{E}X$, $Y - \mathbb{E}Y$ także są niezależne. Zatem

$$\text{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = [\mathbb{E}(X - \mathbb{E}X)] \cdot [\mathbb{E}(Y - \mathbb{E}Y)] = 0,$$

gdyż $\mathbb{E}(X - \mathbb{E}X) = \mathbb{E}X - \mathbb{E}(\mathbb{E}X) = \mathbb{E}X - \mathbb{E}X = 0$. Udowodniliśmy zatem następujące twierdzenie:

Twierdzenie 30. *Jeśli X, Y są niezależnymi zmiennymi losowymi takimi, że zachodzi $\mathbb{E}|XY| < \infty$, to $\text{Cov}(X, Y) = 0$. Jeśli dodatkowo zmienne X, Y mają skończoną, niezerową wariancję, to zeruje się także ich współczynnik korelacji,*

$$\rho(X, Y) = 0.$$

Definicja 30. *Zmienne losowe X, Y nazwiemy nieskorelowanymi, jeżeli*

$$\rho(X, Y) = 0.$$

Uwaga: Jak zatem widzimy, niezależne zmienne losowe są zawsze nieskorelowane. Implikacja odwrotna nie musi jednak zachodzić.

Przykład: Jak już wiemy, zmienna losowa (X, Y) o rozkładzie jednostajnym na kole jednostkowym nie ma niezależnych współrzędnych. Ale, z drugiej strony,

$$\mathbb{E}X = \mathbb{E}Y = \int_{-1}^1 x \frac{2}{\pi} \sqrt{1-x^2} dx = 0$$

oraz

$$\text{Cov}(X, Y) = \mathbb{E}XY - 0 \cdot 0 = \mathbb{E}XY = \frac{1}{\pi} \int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} xy dy dx = 0,$$

gdyż wewnętrzna całka wynosi 0. Ponieważ (X, Y) jest ograniczona, więc jej współrzędne są całkowalne z kwadratem; mamy zatem także $\rho(X, Y) = 0$. Zmienne X, Y są więc nieskorelowane, ale są zależne.

Twierdzenie 31. *Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi o skończonej wariancji. Wówczas zmienna $X_1 + X_2 + \dots + X_n$ także ma skończoną wariancję oraz*

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n).$$

Dowód: Ponieważ wariancja nie zmienia się na skutek przesuwania zmiennych, wystarczy udowodnić powyższą równość dla *scentrowanych* zmiennych, tzn. o średniej 0. Mamy wówczas

$$\begin{aligned} \text{Var}(X_1 + X_2 + \dots + X_n) &= \mathbb{E}(X_1 + X_2 + \dots + X_n)^2 \\ &= \mathbb{E}X_1^2 + \mathbb{E}X_2^2 + \dots + \mathbb{E}X_n^2 + 2 \sum_{i < j} \mathbb{E}X_i X_j \\ &= \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n) + 2 \sum_{i < j} \text{Cov}(X_i, X_j). \end{aligned}$$

Powyższy wzór jest prawdziwy dla dowolnych zmiennych losowych (całkowalnych z kwadratem). W naszym przypadku dysponujemy dodatkowym warunkiem, iż zmienne są niezależne; pociąga to za sobą, iż wszystkie kowariancje są równe 0. Stąd teza. \square

Przykład: Rzucamy 100 razy kostką. Niech X oznacza łączną liczbę wyrzuczonych oczek. Możemy zapisać $X = \sum_{i=1}^{100} X_i$, gdzie X_i – liczba oczek wyrzuczona w i -tym rzucie. Jak wiemy, zachodzi równość

$$\text{Var}X_i = \frac{35}{12},$$

skąd (i z niezależności zmiennych X_i),

$$\text{Var}X = \frac{3500}{12}.$$

Podamy tu jeszcze jeden ważny fakt dotyczący rozkładu sum niezależnych zmiennych losowych.

Twierdzenie 32. *Załóżmy, że X, Y są niezależnymi zmiennymi losowymi o rozkładach ciągłych z gęstościami g_X, g_Y . Wówczas zmienna $X + Y$ ma rozkład z gęstością będącą splotem funkcji g_X i g_Y , tzn.*

$$g_{X+Y}(x) = g_X * g_Y(x) = \int_{\mathbb{R}} g_X(y)g_Y(x-y)dy = \int_{\mathbb{R}} g_X(x-y)g_Y(y)dy.$$

Przykłady zastosowania powyższego twierdzenia omówimy w dalszej części wykładu.

Na zakończenie tego rozdziału omówimy nieco dokładniej wielowymiarowy analog rozkładu normalnego. Zaczniemy od przypadku dwuwymiarowego. Standardowym rozkładem normalnym w \mathbb{R}^2 jest rozkład zmiennej (X_1, X_2) , gdzie X_1, X_2 są niezależne i mają rozkłady $\mathcal{N}(0, 1)$. Taki rozkład ma gęstość

$$g(x, y) = \frac{1}{2\pi} \exp \left[-\frac{1}{2}(x^2 + y^2) \right].$$

Ogólniej, niech $m = (m_1, m_2)$ będzie wektorem w \mathbb{R}^2 oraz A będzie symetryczną i dodatnio określoną macierzą 2×2 , (ostatni warunek oznacza, iż $a_{11} > 0$ oraz $\det A > 0$). Określmy

$$g(x, y) = \frac{\sqrt{\det A}}{2\pi} \exp\left(-\frac{1}{2}\langle A(x - m_1, y - m_2), (x - m_1, y - m_2) \rangle\right),$$

gdzie $\langle x, y \rangle$ oznacza iloczyn skalarny wektorów $x, y \in \mathbb{R}^2$. Bardziej wprost, jeśli

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

to funkcja g dana jest wzorem

$$g(x, y) = \frac{\sqrt{a_{11}a_{22} - a_{21}^2}}{2\pi} \cdot \exp\left(-\frac{1}{2}(a_{11}(x - m_1)^2 + 2a_{12}(x - m_1)(y - m_2) + a_{22}(y - m_2)^2)\right).$$

Wówczas g jest gęstością rozkładu o średniej (m_1, m_2) i macierzy kowariancji $Q = A^{-1}$. Rozkład ten nazywamy *rozkładem normalnym o średniej m i macierzy kowariancji Q* .

Okazuje się, że dwuwymiarowe rozkłady normalne dają się otrzymać poprzez liniowe (a raczej afiniczne) przekształcenie $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ standardowej zmiennej (X_1, X_2) ; istotnie, mając daną macierz A i wektor m jak wyżej, istnieje przekształcenie liniowe S takie, że $A = SS^t$ i jako szukane przekształcenie afiniczne wystarczy wziąć $T = SX + m$.

Ogólna definicja jest następująca. Niech $m = (m_1, m_2, \dots, m_n)$ będzie wektorem w \mathbb{R}^n , a A będzie dodatnio określoną macierzą $n \times n$ (tzn. dla dowolnego wektora $x \in \mathbb{R}^n$ mamy $\langle Ax, x \rangle > 0$, gdzie tym razem $\langle \cdot, \cdot \rangle$ oznacza iloczyn skalarny w \mathbb{R}^n). Rozkład o gęstości

$$g(x) = \frac{\sqrt{\det A}}{(2\pi)^{n/2}} \exp\left(-\frac{\langle A(x - m), (x - m) \rangle}{2}\right), \quad x \in \mathbb{R}^n,$$

nazwiemy rozkładem normalnym; jego średnią jest m i macierz kowariancji wynosi $Q = A^{-1}$.

Podajmy kilka własności wielowymiarowego rozkładu normalnego.

(i) Rozkład normalny jest jednoznacznie wyznaczony przez swoją średnią i macierz kowariancji.

(ii) Jeśli X ma rozkład normalny w \mathbb{R}^n , o średniej m i macierzy kowariancji Q , k jest wektorem w \mathbb{R}^d i T jest macierzą $d \times n$, to $TX + k$ ma rozkład normalny w \mathbb{R}^d , o średniej $Tm + k$ i macierzy kowariancji TQT^t .

Kolejną własność zapiszemy jako twierdzenie.

Twierdzenie 33. *Załóżmy, że $X = (X_1, X_2, \dots, X_n)$ ma rozkład normalny i zmienne X_1, X_2, \dots, X_n są nieskorelowane. Wówczas są one niezależne.*

Dowód: Jeśli zmienne X_i są nieskorelowane, to macierz kowariancji Q jest diagonalna:

$$Q = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}.$$

Wobec tego

$$A = Q^{-1} = \begin{bmatrix} 1/\sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1/\sigma_n^2 \end{bmatrix}$$

oraz $\det A = 1/(\sigma_1\sigma_2\dots\sigma_n)^2$, a więc

$$\begin{aligned} g(x) &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(x_1 - m_1)^2}{2\sigma_1^2}\right] \cdot \dots \cdot \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{(x_n - m_n)^2}{2\sigma_n^2}\right] \\ &= g_1(x_1)g_2(x_2)\dots g_n(x_n). \end{aligned}$$

Zatem gęstość łączna jest iloczynem gęstości brzegowych; stąd niezależność współrzędnych. \square

18. ZAGADNIENIE REGRESJI LINIOWEJ

Rozważmy tutaj - pokrótce - pewien problem, grający ważną rolę w zastosowaniach. Załóżmy, że mamy zmienne losowe X, Y całkowalne z kwadratem i znamy ich łączny rozkład. Ponadto, przypuśćmy, iż obserwujemy wartości zmiennej X , a zmienna Y jest trudniejsza - bądź niemożliwa - do zmierzenia. Powstaje więc interesujące zagadnienie optymalnego przybliżania zmiennej Y za pomocą zmiennej X . Oczywiście, musimy odpowiednio postawić ten problem; będziemy szukać optymalnego przybliżenia *liniowego*, tzn. postaci $aX + b$, $a, b \in \mathbb{R}$, a błąd będziemy mierzyć w sensie średniokwadratowym. Innymi słowy, szukamy stałych $a, b \in \mathbb{R}$, dla których wielkość $f(a, b) = \mathbb{E}(Y - aX - b)^2$ jest najmniejsza.

Aby rozwiązać ten problem, zauważmy, iż przy ustalonym a , funkcja $b \mapsto f(a, b)$ jest trójmianem kwadratowym, który przyjmuje swoją najmniejszą wartość w punkcie $\mathbb{E}(Y - aX)$. Wystarczy więc wyznaczyć najmniejszą wartość funkcji

$$h(a) = f(a, \mathbb{E}(Y - aX)) = \mathbb{E}(Y - \mathbb{E}Y - a(X - \mathbb{E}X))^2 = a^2 \text{Var}X - 2a \text{Cov}(X, Y) + \text{Var}Y.$$

Jeśli zmienna X jest stała p.n. (czyli $\text{Var}X=0$), to wówczas h jest funkcją stałą i widać, że optymalnym liniowym estymatorem zmiennej Y jest jej średnia: $aX + b = aX + (\mathbb{E}Y - a\mathbb{E}X) = \mathbb{E}Y$. Jeśli zaś $\text{Var}X \neq 0$, to h jest trójmianem kwadratowym zmiennej a , przyjmującym swoją najmniejszą wartość w punkcie

$$a = \frac{\text{Cov}(X, Y)}{\text{Var}X}$$

i wówczas

$$b = \mathbb{E}Y - \mathbb{E}X \cdot \frac{\text{Cov}(X, Y)}{\text{Var}X}.$$

Uwagi:

1. Widać, że do powyższych obliczeń nie potrzebowaliśmy całej wiedzy o rozkładzie łącznym zmiennych (X, Y) . Wystarczy nam znajomość średnich i wariancji zmiennych X, Y oraz ich kowariancji.

2. W praktyce nie znamy tych wielkości, dysponujemy tylko pewną próbką danych. Wówczas możemy przeprowadzić powyższe rozumowanie w oparciu o wielkości empiryczne.

19. PRZEGLĄD WAŻNIEJSZYCH ROZKŁADÓW PRAWDOPODOBIENSTWA

19.1. Rozkład wykładniczy dwustronny (rozkład Laplace'a) z parametrem $\lambda > 0$. Jest to „usymetrycznienie” rozkładu wykładniczego: jeśli X_1, X_2 są niezależnymi zmiennymi losowymi o rozkładzie $\text{Exp}(\lambda)$, to ich różnica ma dwustronny rozkład wykładniczy. Rozkład ten posiada gęstość

$$g(x) = \frac{1}{2}\lambda e^{-\lambda|x|}.$$

Mamy $\mathbb{E}X = 0$, $\text{Var}X = 2/\lambda^2$. Ponadto, X ma rozkład symetryczny, tzn. X ma ten sam rozkład, co $-X$.

19.2. Rozkład Gamma $\Gamma(a, b)$, $a, b > 0$. Zaczniemy od definicji funkcji Γ . Jest to funkcja określona na półprostej dodatniej, dana wzorem

$$\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt.$$

Funkcja Γ uogólnia silnię; mamy mianowicie $\Gamma(n) = (n-1)!$ dla $n = 1, 2, \dots$

Niech teraz a, b będą liczbami dodatnimi. Zmienna losowa X ma rozkład $\Gamma(a, b)$, jeśli ma gęstość

$$g_{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \mathbf{1}_{(0,\infty)}(x).$$

Szczególne przypadki:

- (i) jeśli wziąć $a = 1$, dostajemy rozkład wykładniczy z parametrem b .
- (ii) jeśli a jest liczbą całkowitą, to $\Gamma(a, b)$ nazywamy czasem *rozkładem Erlanga*.
- (iii) jeśli n jest liczbą całkowitą dodatnią, to $\Gamma(n/2, 1/2)$ nazywamy *rozkładem χ_n^2 - chi kwadrat z n stopniami swobody*.

Łatwo policzyć, że $\mathbb{E}X = a/b$ oraz $\text{Var}X = a/b^2$.

Twierdzenie 34. *Załóżmy, że zmienne losowe X_1, X_2, \dots, X_n są niezależne i mają rozkłady $\Gamma(a_1, b), \Gamma(a_2, b), \dots, \Gamma(a_n, b)$, odpowiednio. Wówczas zmienna $X_1 + X_2 + \dots + X_n$ ma rozkład $\Gamma(a_1 + a_2 + \dots + a_n, b)$.*

Dowód: Ze względu na indukcję, wystarczy udowodnić tezę dla $n = 2$. Gęstość rozkładu $X_1 + X_2$ zadana jest przez splot

$$\begin{aligned} g_{X_1+X_2}(x) &= \int_{\mathbb{R}} \frac{b^{a_1}}{\Gamma(a_1)} y^{a_1-1} e^{-by} \mathbf{1}_{(0,\infty)}(y) \frac{b^{a_2}}{\Gamma(a_2)} (x-y)^{a_2-1} e^{-b(x-y)} \mathbf{1}_{(0,\infty)}(x-y) dy \\ &= \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)} e^{-bx} \int_0^x y^{a_1-1} (x-y)^{a_2-1} dy. \end{aligned}$$

Po podstawieniu w całce $y = xt$ dostajemy, iż

$$g_{X_1+X_2}(x) = C \cdot x^{a_1+a_2-1} e^{-bx} \mathbf{1}_{(0,\infty)}(x)$$

dla pewnej stałej C . To zaś oznacza już, iż $X_1 + X_2$ ma rozkład $\Gamma(a_1 + a_2, b)$; istotnie, ponieważ $g_{X_1+X_2}$ jest gęstością, to całka z niej wynosi 1; zatem stała C musi wynosić

$$\frac{b^{a_1+a_2}}{\Gamma(a_1 + a_2)}. \quad \square$$

Jako wniosek dostajemy, iż suma n niezależnych zmiennych losowych o rozkładzie wykładniczym z parametrem λ ma rozkład Erlanga.

Kolejny fakt ilustruje ciekawy związek między rozkładem chi kwadrat a rozkładem normalnym.

Twierdzenie 35. *Załóżmy, że X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o rozkładzie $\mathcal{N}(0, 1)$. Wówczas zmienna $X_1^2 + X_2^2 + \dots + X_n^2$ ma rozkład chi kwadrat z n stopniami swobody.*

Dowód: Wystarczy udowodnić, że zmienna X_1^2 ma rozkład $\Gamma(1/2, 1/2)$ i skorzystać z poprzedniego twierdzenia. To zaś uzyskujemy prosto przez obliczenie dystrybuanty X_1^2 i jej zróżniczkowanie. \square

Rozkład chi kwadrat pojawia się w naturalny sposób przy badaniu średniej i wariancji z próbki o rozkładzie normalnym, co ma dość duże znaczenie w statystyce.

Twierdzenie 36. *Jeżeli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o rozkładzie $\mathcal{N}(0, 1)$ oraz m, s^2 oznaczają odpowiednio średnią i wariancję z próby:*

$$m = \frac{1}{n} \sum_{i=1}^n X_i$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2,$$

to \sqrt{nm} ma rozkład $\mathcal{N}(0, 1)$, ns^2 ma rozkład χ_{n-1}^2 oraz zmienne m, s^2 są niezależne.

19.3. Rozkład lognormalny (logarytmicznie normalny) $L(m, \sigma)$, $m \in \mathbb{R}$, $\sigma > 0$. Jest to rozkład zmiennej $Y = e^X$, gdzie $X \sim \mathcal{N}(m, \sigma^2)$. Ma on gęstość

$$g(x) = \frac{1}{\sqrt{2\pi\sigma x}} \exp\left(-\frac{(\ln x - m)^2}{2\sigma^2}\right) 1_{(0, \infty)}(x).$$

Rozkłady lognormalne służą w ekonomii np. do modelowania cen akcji. W nieco nieprecyzyjnym ujęciu, rozkłady te pojawiają się przy modelowaniu iloczynów dużej liczby zmiennych, z których każda jest blisko 1.

Można policzyć, iż $\mathbb{E}Y = \exp(m + \frac{1}{2}\sigma^2)$, $\text{Var}Y = (\exp(\sigma^2) - 1) \exp(2m + \sigma^2)$.

Twierdzenie 37. *Jeśli Y_1, Y_2, \dots, Y_n są niezależnymi zmiennymi losowymi o rozkładach lognormalnych $L(m_1, \sigma_1), L(m_2, \sigma_2), \dots, L(m_n, \sigma_n)$, to ich iloczyn ma rozkład lognormalny $L(m_1 + m_2 + \dots + m_n, \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2})$.*

Wynika to natychmiast z analogicznej własności rozkładu normalnego.

19.4. Rozkład Cauchy'ego $\text{Cau}(a, m)$, $a > 0$, $m \in \mathbb{R}$. Jest to rozkład z gęstością

$$g(x) = \frac{a}{\pi((x - m)^2 + a^2)}.$$

Najczęściej spotyka się rozkład $\text{Cau}(1, 0)$ (standardowy rozkład Cauchy'ego).

Twierdzenie 38. *Jeśli X_1, X_2, \dots, X_n są niezależnymi zmiennymi o standardowym rozkładzie Cauchy'ego, to ich suma ma ten sam rozkład, co zmienna nX_1 .*

Rozkład Cauchy'ego nie posiada skończonej wartości oczekiwanej.

19.5. **Rozkład Pareto.** Jest to rozkład o gęstości

$$g_{a,b}(x) = \frac{ab^a}{x^{a+1}} 1_{[b,\infty)}(x), \quad a, b > 0.$$

Służą on do modelowania dochodów bądź wielkości miast, średnic meteorytów, itp.

Mamy $\mathbb{E}X = ab/(a-1)$ dla $a > 1$, $\text{Var}X = ab^2/[(a-1)^2(a-2)]$ dla $a > 2$.

19.6. **Rozkład F -Snedecora,** $F(d_1, d_2)$, $d_1, d_2 = 1, 2, \dots$. Jest to ważny rozkład w statystyce. Zmienna losowa X ma rozkład $F(d_1, d_2)$, jeśli $X = \frac{Y_1}{d_1} / \frac{Y_2}{d_2}$, gdzie Y_1, Y_2 są niezależne i Y_i ma rozkład $\chi_{d_i}^2$, $i = 1, 2$.

Rozkład $F(d_1, d_2)$ posiada gęstość

$$g_{d_1, d_2}(x) = \frac{\left(\frac{d_1 x}{d_1 x + d_2}\right)^{d_1/2} \left(1 - \frac{d_1 x}{d_1 x + d_2}\right)^{d_2/2}}{x B(d_1/2, d_2/2)} 1_{(0, \infty)}(x),$$

gdzie B oznacza funkcję beta:

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad a, b > 0.$$

Średnia wynosi $d_2/(d_2-2)$ dla $d_2 > 2$, a wariancja

$$\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)} \quad \text{dla } d_2 > 4.$$

19.7. **Rozkład t -Studenta.** Jest to rozkład w naturalny sposób pojawiający się w statystyce. Rozkład t -Studenta o n stopniach swobody to rozkład zmiennej $\sqrt{n}X/\sqrt{Y_n}$, gdzie X, Y_n są niezależne, $X \sim \mathcal{N}(0, 1)$ i $Y_n \sim \chi_n^2$. Ma on gęstość

$$g_n(x) = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma(\frac{1}{2}(n+1))}{\Gamma(\frac{1}{2}n)} \left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)}, \quad n = 1, 2, \dots$$

Dla $n = 1$ dostajemy standardowy rozkład Cauchy'ego.

Wartość oczekiwana nie istnieje dla $n = 1$, wariancja nie istnieje dla $n = 1, 2$. W pozostałych przypadkach, wartość oczekiwana jest równa 0, a wariancja wynosi $n/(n-2)$.

19.8. **Rozkład Weibulla.** Ma on gęstość

$$g(x) = \alpha\beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} 1_{(0, \infty)}(x), \quad \alpha, \beta > 0.$$

Rozkład ten służy do modelowania czasu „życia” rozmaitych obiektów technicznych.

Jeśli X ma ten rozkład, to

$$\mathbb{E}X = \beta\Gamma\left(1 + \frac{1}{\alpha}\right), \quad \text{Var}X = \beta^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \left(\Gamma\left(1 + \frac{1}{\alpha}\right)\right)^2\right].$$

20. WARUNKOWA WARTOŚĆ OCZEKIWANA

Jak już się przekonaliśmy, dodatkowa wiedza na temat wyniku eksperymentu losowego wymusza na nas przejście do prawdopodobieństwa warunkowego i może istotnie zmienić nasze oczekiwania co do możliwych wyników doświadczenia. Zjawisko to przejawia się nie tylko na poziomie prawdopodobieństw poszczególnych zdarzeń, dotyczy także wartości oczekiwanej zmiennych losowych. Przykładowo, jeśli rzucamy dwa razy kostką, wartość oczekiwana łącznej wyrzuconej liczby oczek wynosi 7. Gdy jednak wiemy, że w pierwszym rzucie wypadły dwa oczka, naturalne jest przyjąć, że wartość oczekiwana łącznej liczby oczek wynosi $2 + 3.5 = 5.5$ (czyli

jest sumą wyrzuconej dotychczas liczby oczek i wartości oczekiwanej liczby oczek wyrzuconych w drugim rzucie).

Podobnie, losując punkt (X, Y) z kwadratu o wierzchołkach $(0, 0)$, $(1, 0)$, $(1, 1)$, $(0, 1)$ mamy $\mathbb{E}XY = \mathbb{E}X \cdot \mathbb{E}Y = 1/4$ (zmiennie X, Y są niezależne, o wspólnym rozkładzie $\mathcal{U}([0, 1])$). Jeśli jednak znamy wartość jednej ze współrzędnych, np. wiemy, że $X = 1/3$, rozsądnie byłoby przyjąć, że wartość oczekiwana iloczynu współrzędnych wynosi $\frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$.

Powyższe przykłady prowadzą do pojęcia *warunkowej wartości oczekiwanej*, niezwykle przydatnego w działach rachunku prawdopodobieństwa. Wyczerpujące omówienie wszystkich własności tego obiektu wykracza znacznie poza zakres wykładu, skoncentrujemy się więc, podobnie jak przy omawianiu rozkładów zmiennych losowych, na dwóch najważniejszych przypadkach: dyskretnym i ciągłym.

Rozpatrzmy zatem dwuwymiarowy wektor losowy (X, Y) i założmy na początek, że ma on rozkład dyskretny. Weźmy $x \in \mathbb{R}$ takie, że $\mathbb{P}(X = x) > 0$. Funkcja zbioru $A \in \mathcal{B}(\mathbb{R})$, dana wzorem $\mu(A) = \mathbb{P}(Y \in A | X = x)$ jest prawdopodobieństwem (spełnia aksjomaty Kołmogorowa) - jest to po prostu prawdopodobieństwo warunkowe względem zdarzenia $\{X = x\}$. Możemy zatem zdefiniować *warunkową wartość oczekiwaną zmiennej Y pod warunkiem $X = x$* po prostu jako wartość oczekiwaną rozkładu prawdopodobieństwa μ .

Definicja 31. Niech (X, Y) będzie dwuwymiarową zmienną losową o rozkładzie dyskretnym, spełniającą $\mathbb{E}|Y| < \infty$. Dla dowolnej liczby $x \in \mathbb{R}$, takiej że $\mathbb{P}(X = x) > 0$, definiujemy *warunkową wartość oczekiwaną Y pod warunkiem $X = x$* (ozn. $\mathbb{E}(Y|X = x)$) jako *wartość oczekiwaną rzeczywistej zmiennej losowej o rozkładzie μ , danym wzorem*

$$\mu(A) = \mathbb{P}(Y \in A | X = x).$$

Zatem, jeśli S jest zbiorem tych $y \in \mathbb{R}$, że $\mathbb{P}(Y = y | X = x) > 0$ (równoważnie, $\mathbb{P}((X, Y) = (x, y)) > 0$), mamy

$$\mathbb{E}(Y|X = x) = \sum_{y \in S} y \mathbb{P}(Y = y | X = x).$$

Zachodzi też odpowiednik znanego nam twierdzenia o wartości oczekiwanej funkcji zmiennych losowych.

Twierdzenie 39. Jeśli (X, Y) jest zmienną losową o rozkładzie dyskretnym, zaś $\varphi: \mathbb{R} \rightarrow \mathbb{R}$, funkcją borelowską, taką że $\mathbb{E}|\varphi(Y)| < \infty$, to dla dowolnego x , takiego że $\mathbb{P}(X = x) > 0$, zachodzi równość

$$\mathbb{E}(\varphi(Y)|X = x) = \sum_{y \in S} \varphi(y) \mathbb{P}(Y = y | X = x),$$

gdzie, jak w poprzedniej definicji, $S = \{y \in \mathbb{R}: \mathbb{P}((X, Y) = (x, y)) > 0\}$.

Przykłady:

1. Niech (X, Y) będzie wektorem losowym skupionym na zbiorze $\{(0, 0)$, $(0, 1)$, $(1, 0)\}$, o rozkładzie zadany przez wagi $p_{(0,0)} = 1/2$, $p_{(0,1)} = p_{(1,0)} = 1/4$. Mamy

$$\mathbb{E}Y = \frac{3}{4} \cdot 0 + \frac{1}{4} \cdot 1 = \frac{1}{4}.$$

Ponadto $\mathbb{P}(Y = 0 | X = 0) = 2/3$ oraz $\mathbb{P}(Y = 1 | X = 0) = 1/3$. Zatem

$$\mathbb{E}(Y|X = 0) = \frac{2}{3} \cdot 0 + \frac{1}{3} \cdot 1 = \frac{1}{3}.$$

Analogicznie $\mathbb{P}(Y = 0|X = 1) = 1$, a więc

$$\mathbb{E}(Y|X = 1) = 0 \cdot 1 = 0.$$

Ten wynik jest intuicyjnie zupełnie oczywisty: jeżeli $X = 1$, wartość zmiennej Y jest już zdeterminowana (równa 0).

2. Jeżeli zmienna losowa Y jest funkcją zmiennej X , tzn. $Y = f(X)$, to zgodnie z intuicją, $\mathbb{E}(Y|X = x) = f(x)$. Rzeczywiście, zbiór S z twierdzenia zawiera tylko jedną liczbę $y = f(x)$ oraz $\mathbb{P}(Y = y|X = x) = 1$.

3. Niech Y będzie zmienną losową skupioną na zbiorze $0, 1, -1, 2, -2$ i rozkładzie zadany przez wagi

$$p_0 = \frac{1}{3}, p_1 = p_{-1} = \frac{1}{6}, p_2 = \frac{1}{4}, p_{-2} = \frac{1}{12}.$$

Rozważmy zmienną losową $X = |Y|$. Mamy

$$\mathbb{P}(Y = 0|X = 0) = 1,$$

więc $\mathbb{E}(Y|X = 0) = 0$. Ponadto

$$\mathbb{P}(Y = 1|X = 1) = \mathbb{P}(Y = -1|X = 1) = \frac{1}{2},$$

więc również

$$\mathbb{E}(Y|X = 1) = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0.$$

Z kolei $\mathbb{P}(Y = -2|X = 2) = 1/4$ oraz $\mathbb{P}(Y = 2|X = 2) = 3/4$, zatem

$$\mathbb{E}(Y|X = 2) = (-2) \cdot \frac{1}{4} + 2 \cdot \frac{3}{4} = 1.$$

Przejdźmy teraz do zmiennych X, Y o łącznym rozkładzie ciągłym i niech $g: \mathbb{R}^2 \rightarrow [0, \infty)$ będzie gęstością (rozkładu) wektora (X, Y) . Chcielibyśmy zdefiniować w sensowny sposób $\mathbb{E}(Y|X = x)$. Sytuacja jest jednak bardziej skomplikowana niż w przypadku dyskretnym, gdyż dla każdego $x \in \mathbb{R}$, $\mathbb{P}(X = x) = 0$, a więc prawdopodobieństwo warunkowe $\mathbb{P}(\cdot|X = x)$ nie jest dobrze określone. Problem ten pokonamy, definiując tzw. *gęstość warunkową* zmiennej Y pod warunkiem zmiennej X .

Definicja 32. Niech (X, Y) będzie dwuwymiarowym wektorem losowym o gęstości $g: \mathbb{R}^2 \rightarrow [0, \infty)$. Niech $g_X(x) = \int_{-\infty}^{\infty} g(x, y) dy > 0$ będzie gęstością zmiennej X . Dla $x \in \mathbb{R}$, definiujemy gęstość warunkową zmiennej Y pod warunkiem $X = x$ jako funkcję daną wzorem

$$g_{Y|X}(y|x) = \begin{cases} \frac{g(x, y)}{g_X(x)} & \text{jeśli } g_X(x) > 0 \\ f(y) & \text{w przeciwnym przypadku,} \end{cases}$$

gdzie $f: \mathbb{R} \rightarrow [0, \infty)$ jest dowolną ustaloną gęstością prawdopodobieństwa.

Uwagi

1. Gęstość warunkowa może być postrzegana jako ciągły odpowiednik prawdopodobieństwa warunkowego. Całka w mianowniku ma charakter normujący. Intuicyjnie, jeśli wiemy, że $X = x$, oczekujemy, że warunkowa gęstość zmiennej Y w punkcie y , powinna być proporcjonalna do $g(x, y)$. Stała proporcjonalności powinna być taka, by gęstość warunkowa całkowała się do 1, co już ją determinuje.

2. Gęstość warunkowa nie jest wyznaczona jednoznacznie, nie tylko ze względu na dowolność wyboru gęstości f w definicji, ale także ze względu na to, że gęstość wektora losowego (X, Y) nie musi być wyznaczona jednoznacznie. W rzeczywistości jednak nie stanowi to problemu, gdyż gęstość warunkowa jest wyznaczona jednoznacznie na prawie całej prostej (nie dysponujemy odpowiednim aparatem matematycznym, aby sprecyzować to pojęcie) i w problemach, które będziemy rozpatrywać nie będzie miało znaczenia, której „wersji” gęstości warunkowej używamy.

Przykłady:

1. Niech (X, Y) ma rozkład jednostajny na kwadracie o wierzchołkach $(1, 0)$, $(0, 1)$, $(-1, 0)$, $(0, -1)$. Gęstość wektora (X, Y) to

$$g(x, y) = \frac{1}{2} 1_{\{|x|+|y|\leq 1\}}(x, y).$$

Dla $x \in (-1, 1)$ mamy $\int_{-\infty}^{\infty} g(x, y) dy = (1 - |x|)$. Gęstość warunkowa wynosi zatem

$$g_{Y|X}(y|x) = \frac{1_{\{|y|\leq 1-|x|\}}(x, y)}{2(1 - |x|)},$$

dla $x \in (-1, 1)$. Dla $x \notin (-1, 1)$ za gęstość warunkową podstawiamy dowolną ustaloną gęstość prawdopodobieństwa na \mathbb{R} . Korzystając z tak obliczonej gęstości warunkowej możemy teraz nadać sens np. prawdopodobieństwu warunkowemu $\mathbb{P}(Y \geq \frac{1}{2} | X = x)$. Zauważmy, że nie możemy skorzystać z klasycznej definicji (przeszkodą jest równość $\mathbb{P}(X = x) = 0$), tym niemniej intuicyjnie poprawne jest przyjęcie, że

$$\mathbb{P}(Y \geq \frac{1}{2} | X = x) = \int_{1/2}^{\infty} g_{Y|X}(y|x) dy = \begin{cases} \frac{1/2-|x|}{1-|x|} & \text{dla } |x| \leq 1/2 \\ 0 & \text{dla } x \in (-1, 1) \setminus (-1/2, 1/2). \end{cases}$$

Dla pozostałych wartości x nie definiujemy powyższego prawdopodobieństwa warunkowego. Możemy też przyjąć, że jest ono równe $\int_{1/2}^{\infty} f(y) dy$ bądź dowolnej liczbie z przedziału $[0, 1]$ (wybór nie ma znaczenia, zmienna X przyjmuje wartość spoza zbioru $(-1, 1)$ z prawdopodobieństwem 0, więc w praktyce to prawdopodobieństwo warunkowe pozostanie dla nas tylko „ozdobnikiem”).

2. Losujemy liczbę Λ z przedziału $(0, 1)$, a następnie liczbę X z rozkładu $\text{Exp}(\Lambda)$. Wyznamy gęstość zmiennych (Λ, X) oraz X . Mamy, iż gęstość rozkładu zmiennej losowej Λ jest równa $g_{\Lambda}(\lambda) = 1_{(0,1)}(\lambda)$, natomiast gęstość warunkowa zmiennej X pod warunkiem $\Lambda = \lambda$ dana jest wzorem $g_{X|\Lambda}(x|\lambda) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x)$. Stąd

$$g_{(\Lambda, X)}(\lambda, x) = g_{X|\Lambda}(x|\lambda) g_{\Lambda}(\lambda) = \lambda e^{-\lambda x} 1_{(0,1)}(\lambda) 1_{(0,\infty)}(x).$$

Znając gęstość łączną wektora (Λ, X) , możemy obliczyć g_X całkując po zmiennej λ . Otrzymujemy

$$\begin{aligned} g_X(x) &= \int_{-\infty}^{\infty} g_{(\Lambda, X)}(\lambda, x) d\lambda = 1_{(0,\infty)}(x) \int_0^1 \lambda e^{-\lambda x} d\lambda \\ &= 1_{(0,\infty)}(x) \left[-\frac{\lambda}{x} e^{-\lambda x} - \frac{1}{x^2} e^{-\lambda x} \right] \Big|_0^1 = 1_{(0,\infty)}(x) \left(\frac{1}{x^2} - \frac{1}{x} e^{-x} - \frac{1}{x^2} e^{-x} \right). \end{aligned}$$

Możemy teraz zdefiniować warunkową wartość oczekiwaną w przypadku ciągłym. Zauważmy, że dla tych x , dla których gęstość warunkowa jest dobrze określona, jest ona gęstością pewnego rozkładu prawdopodobieństwa na prostej (jest nieujemna i całkuje się do 1). Możemy więc postąpić podobnie jak w przypadku rozkładów dyskretnych i zdefiniować warunkową wartość oczekiwaną jako wartość oczekiwaną tego nowego rozkładu.

Definicja 33. Niech (X, Y) będzie dwuwymiarowym wektorem losowym o gęstości $g: \mathbb{R}^2 \rightarrow [0, \infty)$, takim że $\mathbb{E}|Y| < \infty$. Dla $x \in \mathbb{R}$ definiujemy warunkową wartość oczekiwaną zmiennej Y pod warunkiem $X = x$ (ozn. $\mathbb{E}(Y|X = x)$) jako wartość oczekiwaną rzeczywistej zmiennej losowej o gęstości $f_x(y) = g_{Y|X}(y|X = x)$. Zatem

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} yg_{Y|X}(y|x)dy.$$

Również w przypadku ciągłym mamy warunkowy odpowiednik twierdzenia o wartości oczekiwanej funkcji zmiennych losowych.

Twierdzenie 40. Jeśli (X, Y) jest dwuwymiarową zmienną losową posiadającą gęstość $g: \mathbb{R}^2 \rightarrow [0, \infty)$, zaś $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ jest funkcją borelowską taką że $\mathbb{E}|\varphi(Y)| < \infty$, to dla dowolnego $x \in \mathbb{R}$,

$$\mathbb{E}(\varphi(Y)|X = x) = \int_{-\infty}^{\infty} \varphi(y)g_{Y|X}(y|x)dy.$$

Przykłady:

1. Wektor losowy (X, Y) ma rozkład o gęstości

$$g(x, y) = \frac{1}{2(|x| + 1)} e^{-|x|} 1_{(|x|, 2|x|+1)}(y).$$

Mamy $g_X(x) = \int_{-\infty}^{\infty} g(x, y)dy = 2^{-1}e^{-|x|}$. Zatem

$$g_{Y|X}(y|x) = \frac{g(x, y)}{g_X(x)} = \frac{1}{|x| + 1} 1_{(|x|, 2|x|+1)}(y).$$

Innymi słowy, rozkład zmiennej Y pod warunkiem $X = x$ to rozkład jednostajny na odcinku $(|x|, 2|x| + 1)$. Wobec tego,

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} yg_{Y|X}(y|x)dy = \frac{3|x| + 1}{2}.$$

Czasami wygodnie jest traktować warunkową wartość oczekiwaną jako zmienną losową. Dlatego wprowadzimy następującą definicję.

Definicja 34. Niech X, Y będą zmiennymi losowymi na tej samej przestrzeni probabilistycznej. Załóżmy, że $\mathbb{E}|Y| < \infty$. Warunkową wartość oczekiwaną zmiennej Y pod warunkiem X (ozn. $\mathbb{E}(Y|X)$) nazywamy zmienną losową, daną wzorem

$$\mathbb{E}(Y|X) = m(X),$$

gdzie $m(x) = \mathbb{E}(Y|X = x)$.

Definicja 35. Niech X będzie zmienną losową. Dla dowolnego zdarzenia $A \in \mathcal{F}$, definiujemy $\mathbb{P}(A|X) = \mathbb{E}(1_A|X)$.

Przykłady:

1. Zmienna $\mathbb{E}(Y|X)$ jest zatem funkcją zmiennej X . W przykładzie powyżej, gdy (X, Y) ma gęstość

$$g(x, y) = \frac{1}{2(|x| + 1)} e^{-|x|} 1_{(|x|, 2|x|+1)}(y),$$

mamy $m(x) = \frac{3|x|+1}{2}$, a zatem $\mathbb{E}(Y|X) = \frac{3|X|+1}{2}$.

2. W rozpatrywanym już przez nas przykładzie w przypadku dyskretnym, gdy Y ma rozkład wyznaczony przez wagi

$$p_0 = \frac{1}{3}, \quad p_1 = p_{-1} = \frac{1}{6}, \quad p_2 = \frac{1}{4}, \quad p_{-2} = \frac{1}{12},$$

a $X = |Y|$, mieliśmy $\mathbb{E}(Y|X = 0) = 0$, $\mathbb{E}(Y|X = 1) = 0$, $\mathbb{E}(Y|X = 2) = 1$. Możemy zatem napisać, że $m(x) = \mathbb{E}(Y|X = x) = x(x - 1)/2$ dla $x \in \{0, 1, 2\}$ (pozostałe wartości x nas nie interesują, gdyż X jest skupiona na zbiorze $\{0, 1, 2\}$). Zatem

$$\mathbb{E}(Y|X) = m(X) = \frac{X(X - 1)}{2}.$$

Warunkowa wartość oczekiwana ma wiele własności podobnych do „zwykłej” wartości oczekiwanej. Podsumujemy je w następującym twierdzeniu.

Twierdzenie 41. *Niech $X, Y, Z: \Omega \rightarrow \mathbb{R}$ będą zmiennymi losowymi, przy czym $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$. Wówczas*

- (i) *Jeśli $X \geq 0$, to $\mathbb{E}(X|Z) \geq 0$.*
- (ii) *$|\mathbb{E}(X|Z)| \leq \mathbb{E}(|X||Z)$.*
- (iii) *Dla dowolnych $a, b \in \mathbb{R}$ mamy $\mathbb{E}(aX + bY|Z) = a\mathbb{E}(X|Z) + b\mathbb{E}(Y|Z)$.*

Następne twierdzenie podaje kilka dalszych, użytecznych własności warunkowej wartości oczekiwanej.

Twierdzenie 42. *Niech $X, Y: \Omega \rightarrow \mathbb{R}$ będą zmiennymi losowymi. Jeśli Y jest całkowalna, to*

- (i) *$\mathbb{E}|\mathbb{E}(Y|X)| < \infty$ oraz $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}Y$.*
- (ii) *Jeśli X, Y są niezależne, to $\mathbb{E}(Y|X) = \mathbb{E}Y$.*
- (iii) *Jeśli $h(X)$ jest ograniczoną zmienną losową, to*

$$\mathbb{E}(h(X)Y|X) = h(X)\mathbb{E}(Y|X).$$

Ostatnia własność pozwala uprościć obliczanie warunkowej wartości oczekiwanej. Aby to zilustrować, powróćmy do przykładu wektora (X, Y) o gęstości

$$g(x, y) = \frac{1}{2(|x| + 1)} e^{-|x|} 1_{(|x|, 2|x|+1)}(y).$$

Wiemy, że $\mathbb{E}(Y|X) = (3|X| + 1)/2$. Na przykład, otrzymujemy stąd równość $\mathbb{E}(\sin(X)Y|X) = (3|X| + 1) \sin(X)/2$.

Na zakończenie podamy zastosowanie warunkowej wartości oczekiwanej do zagadnienia prognozy. Wyobraźmy sobie, że pewne zjawisko jest opisane wektorem losowym (X, Y) , przy czym Y jest całkowalna z kwadratem. W praktyce zaobserwować możemy jedynie X (lub też X obserwujemy wcześniej niż Y), jesteśmy jednak zainteresowani wartością zmiennej Y (dla przykładu X, Y mogą oznaczać temperaturę odpowiednio dziś i jutro w południe). Chcielibyśmy więc dysponować

regułą, która pozwoli nam przybliżyć Y przy pomocy X . Jest to sytuacja analogiczna do znanego nam już zagadnienia regresji liniowej, tym razem jednak nie chcemy ograniczać się do przybliżeń postaci $aX + b$, zamiast tego dopuszczamy przybliżenie dowolną funkcją borelowską zmiennej X , czyli zmienną losową postaci $\varphi(X)$, gdzie $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ to funkcja borelowska. Błąd mierzymy ponownie w sensie średniokwadratowym: formalnie, chcemy znaleźć funkcję φ tak, aby zminimalizować wielkość

$$\mathbb{E}(Y - \varphi(X))^2.$$

Okazuje się, że optymalną funkcją φ jest $\varphi^*(x) = \mathbb{E}(Y|X = x)$. Dokładniej, zachodzi następujące twierdzenie.

Twierdzenie 43. *Niech $X, Y: \Omega \rightarrow \mathbb{R}$ będą zmiennymi losowymi, $\mathbb{E}Y^2 < \infty$, wówczas funkcja $\varphi^*: \mathbb{R} \rightarrow \mathbb{R}$, dana wzorem $\varphi^*(x) = \mathbb{E}(Y|X = x)$ spełnia*

$$\mathbb{E}(Y - \varphi^*(X))^2 = \min\{\mathbb{E}(Y - \varphi(X))^2: \varphi: \mathbb{R} \rightarrow \mathbb{R} \text{ - borelowska}\}.$$

21. NIERÓWNOŚĆ CZEBYSZEWA, W STRONĘ PRAW WIELKICH LICZB

W rachunku prawdopodobieństwa czy statystyce często nie jest możliwe dokładne wyliczenie pewnych interesujących nas wartości. Niemniej w wielu zastosowaniach istotna jest nie tyle dokładna wartość co jej oszacowanie. Przykładowo, graczka może interesować czy prawdopodobieństwo, że przegra jest mniejsze niż pewna z góry ustalona liczba α i na tej podstawie może podjąć decyzję czy wziąć udział w grze. Podobnie, w badaniach statystycznych czy przy pomiarach wielkości fizycznych ważne jest oszacowanie prawdopodobieństwa, że błąd wyniesie więcej niż interesująca nas dokładność. Innymi słowy, jeśli przez X oznaczymy losowy błąd danej metody pomiarowej, a przez x żadaną precyzję pomiaru, jesteśmy zainteresowani nierównościami postaci

$$\mathbb{P}(X \geq x) \leq \alpha.$$

Konkretna metoda pomiaru czy procedura statystyczna może zostać uznana za wiarygodną jeżeli α jest odpowiednio małe (dobór α zależy istotnie od konkretnego problemu).

Podstawowym narzędziem matematycznym służącym do uzyskiwania nierówności powyższego typu jest następujące twierdzenie

Twierdzenie 44 (Nierówność Czebyszewa). *Dla dowolnej nieujemnej zmiennej losowej X oraz dla każdego $\varepsilon > 0$,*

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}X}{\varepsilon}.$$

Dowód. Mamy

$$X \geq X1_{\{X \geq \varepsilon\}} \geq \varepsilon 1_{\{X \geq \varepsilon\}}.$$

Biorąc teraz wartości oczekiwane, otrzymujemy

$$\mathbb{E}X \geq \mathbb{E}(\varepsilon 1_{\{X \geq \varepsilon\}}) = \varepsilon \mathbb{P}(X \geq \varepsilon),$$

skąd natychmiast wynika żądana nierówność. \square

Nierówność Czebyszewa, choć niezwykle prosta, ma bardzo dużo zastosowań. Jej siła wynika między innymi z faktu, że możemy zastosować ją nie tylko do zmiennej losowej X , którą jesteśmy zainteresowani, ale także do zmiennych postaci $f(X)$, uzyskując nowe nierówności. Ilustruje to poniższe twierdzenie. Aby je uzyskać, stosujemy nierówność Czebyszewa kolejno do zmiennych $|X|^p$, $(X - \mathbb{E}X)^2$, $e^{\lambda X}$.

Twierdzenie 45. Niech X będzie zmienną losową.

a) **Nierówność Markowa.** Dla dowolnej liczby $p > 0$ oraz dowolnego $\varepsilon > 0$,

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}|X|^p}{\varepsilon^p}.$$

b) **Nierówność Czebyszewa-Bienaimé.** Dla dowolnego $\varepsilon > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}.$$

c) **Wykładnicza nierówność Czebyszewa.** Załóżmy, że $\mathbb{E}e^{pX} < \infty$ dla pewnego $p > 0$. Wówczas dla dowolnego $\lambda \in [0, p]$ oraz dowolnego ε ,

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda \varepsilon}}.$$

Przykłady

1. Przypuśćmy, że dokonujemy szeregu pomiarów jakiejś wielkości fizycznej, przy czym każdy pomiar obarczony jest pewnym błędem. Niech X_1, X_2, \dots, X_n oznaczają wyniki kolejnych pomiarów. W takiej sytuacji naturalnie jest zakładać, że X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o średniej równej prawdziwej wartości mierzonej wielkości fizycznej (oznaczymy ją przez m). Jeżeli wiemy, że wariancja zmiennych X_i jest nie większa od pewnej liczby A , z nierówności Czebyszewa-Bienaimé możemy wywnioskować, że dla każdego i ,

$$\mathbb{P}(|X_i - m| \geq \varepsilon) \leq \frac{A}{\varepsilon^2}.$$

Zauważmy teraz, że jeżeli ε jest małe w stosunku do A , powyższa nierówność może nie dawać nam żadnych informacji (np. dla $\varepsilon = \sqrt{A}$, prawa strona jest równa 1). Jeżeli jednak przybliżymy nieznaną nam liczbę m przez średnią arytmetyczną liczb X_i , dostaniemy

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - m\right| \geq \varepsilon\right) \leq \frac{\text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right)}{\varepsilon^2} = \frac{\sum_{i=1}^n \text{Var}(X_i)}{n^2 \varepsilon^2} \leq \frac{nA}{n^2 \varepsilon^2} = \frac{A}{n \varepsilon^2}.$$

Prawa strona powyższej nierówności zbiega do 0 dla $n \rightarrow \infty$, widzimy więc, że przy dużej liczbie pomiarów możemy z dużym prawdopodobieństwem uzyskać dobre przybliżenie nieznannej wartości m . Co więcej, znając A , wiemy ile pomiarów musimy wykonać, aby prawdopodobieństwo, że nasze przybliżenie będzie obciążone błędem większym niż ε nie przekraczało ustalonej liczby α .

2. Przypuśćmy, że mamy do czynienia z monetą o nieznanym nam prawdopodobieństwie wyrzucenia orła (oznaczymy je przez p). Aby to prawdopodobieństwo przybliżyć, możemy wykonać serię rzutów monetą i sprawdzić częstość wystąpienia orła. Jedno z teoretycznych uzasadnień tej metody wynika z nierówności Czebyszewa. Niech X_i będzie zmienną losową przyjmującą wartość 1 jeśli w i -tym rzucie wypadł orzeł i zero w przeciwnym przypadku. Wówczas zmienne X_i są niezależne, $\mathbb{E}X_i = p$ oraz

$$\text{Var}(X_i) = p - p^2 = p(1 - p).$$

Oznaczając $S_n = \sum_{i=1}^n X_i$, otrzymujemy $\text{Var}(S_n) = np(1 - p)$. Zatem z nierówności Czebyszewa otrzymujemy

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{p(1 - p)}{n \varepsilon^2}.$$

Ponieważ $p(1-p) \leq \frac{1}{4}$, ostatecznie dostajemy

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}.$$

Jeżeli więc zdecydujemy się przybliżyć p na podstawie serii 10000 rzutów, prawdopodobieństwo, że pomylimy się o więcej niż $1/10$ nie przekracza $\frac{1}{4 \cdot 10000 \cdot 10^{-2}} = \frac{1}{400}$.

Powyższy przykład ma jednak funkcje głównie ilustracyjne, okazuje się bowiem, że to prawdopodobieństwo jest dużo mniejsze, co może być wykazane przy użyciu wykładniczej nierówności Czebyszewa. Do tego zagadnienia wrócimy w dalszej części wykładu.

3. Powyższe przykłady dotyczyły sytuacji, w których parametry opisujące dane doświadczenie nie były znane, a nierówności pomagały je przybliżyć. Oczywiście nierówności można zastosować także wtedy, gdy znamy wszystkie parametry naszego modelu, ale chcemy oszacować prawdopodobieństwo pewnych konkretnych zdarzeń. Przypuśćmy, że rzucamy 1000 razy kostką i jesteśmy zainteresowani łączną liczbą wyrzuconych oczek. Wartość oczekiwana liczby oczek wyrzuconych w konkretnym rzucie wynosi 3.5, zatem spodziewamy się, że przy 1000 rzutach łączna liczba oczek powinna być bliska 3500. Oszacujmy prawdopodobieństwo, że będzie się ona różnić od 3500 o więcej niż 100. Jeśli przez X_i oznaczymy liczbę oczek wyrzuconą w i -tym rzucie, zaś $S_n = \sum_{i=1}^{1000} X_i$, mamy

$$\text{Var}(S_n) = \sum_{i=1}^{1000} \text{Var}(X_i) = 1000 \cdot \frac{35}{12} = \frac{35000}{12}.$$

Zatem

$$\mathbb{P}(|S_n - 3500| \geq 100) \leq \text{Var}(S_n)/100^2 = 35/120.$$

Również w tym przypadku istnieją lepsze oszacowania (także oparte o wykładniczą nierówność Czebyszewa), pokazujące, że w rzeczywistości prawdopodobieństwo to jest dużo mniejsze.

Jednym z wniosków z wykładniczej nierówności Czebyszewa, który okaże się przydatny w dokładniejszej analizie przykładu 2, jest tzw. nierówność Bernsteina.

Twierdzenie 46 (Nierówność Bernsteina). *Niech S_n będzie liczbą sukcesów w n -próbach Bernoulliego z prawdopodobieństwem sukcesu równym p . Wówczas, dla każdego $\varepsilon > 0$,*

$$\mathbb{P}\left(\frac{S_n}{n} \geq p + \varepsilon\right) \leq e^{-2\varepsilon^2 n}$$

oraz

$$\mathbb{P}\left(\frac{S_n}{n} \leq p - \varepsilon\right) \leq e^{-2\varepsilon^2 n}.$$

Uwaga: Łącząc obie nierówności z powyższego twierdzenia otrzymujemy, że

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq 2e^{-2\varepsilon^2 n}.$$

Do dowodu nierówności Bernsteina będzie nam potrzebna elementarna nierówność

$$(5) \quad pe^{q\lambda} + qe^{-p\lambda} \leq e^{\lambda^2/8}$$

dla $\lambda, p, q \geq 0, p + q = 1$. Aby ją udowodnić, rozpatrzmy funkcję $f(\lambda) = \ln(pe^{q\lambda} + qe^{-p\lambda})$. Chcemy wykazać, że dla $\lambda \geq 0, f(\lambda) \leq g(\lambda)$ dla $g(\lambda) = \lambda^2/8$. Ponieważ $f(0) = g(0)$, wystarczy wykazać, że $f'(\lambda) \leq g'(\lambda)$. Mamy

$$f'(\lambda) = \frac{pq(e^{q\lambda} - e^{-p\lambda})}{pe^{q\lambda} + qe^{-p\lambda}} = \frac{pq(e^\lambda - 1)}{pe^\lambda + q}$$

oraz $g'(\lambda) = \lambda/4$. Zatem $f'(0) = g'(0)$ i powtarzając powyższe rozumowanie, dochodzimy do wniosku, że wystarczy udowodnić, że $f''(\lambda) \leq g''(\lambda) = 1/4$. Korzystając z założenia $p + q = 1$, otrzymujemy

$$f''(\lambda) = \frac{pqe^\lambda(pe^\lambda + q) - pe^\lambda pq(e^\lambda - 1)}{(pe^\lambda + q)^2} = \frac{pqe^\lambda}{(pe^\lambda + q)^2}$$

Zatem $f''(\lambda) = t(1-t)$, gdzie $t = pe^\lambda/(pe^\lambda + q) \in [0, 1]$. Ponieważ dla $t \in [0, 1]$ mamy $t(1-t) \leq 1/4$, otrzymujemy stąd, że $f''(\lambda) \leq g''(\lambda)$, co pozwala zakończyć dowód.

Dowód nierówności Bernsteina. Zauważmy najpierw, że wystarczy jeśli udowodnimy pierwszą z nierówności, druga już z niej wynika. Rzeczywiście, $S_n/n \leq p - \varepsilon$ jest równoważne nierówności $(n - S_n)/n \geq q + \varepsilon$ dla $q = 1 - p$. Wystarczy więc zauważyć, że $n - S_n$ jest liczbą sukcesów w schemacie Bernoulliego o n próbach i prawdopodobieństwie sukcesu równym q (zmieniamy interpretację naszego oryginalnego doświadczenia, zamieniając znaczeniami słowa „sukces” i „porażka”).

Aby udowodnić pierwsze oszacowanie użyjemy wykładniczej nierówności Czebyszewa. Oznaczmy przez X_i zmienną przyjmującą wartość 1 jeśli i -ta próba zakończyła się sukcesem i 0 w przeciwnym przypadku. Wówczas zmienne X_i są niezależne, $S_n = \sum_{i=1}^n X_i$ oraz dla dowolnego $\lambda > 0$,

$$\mathbb{E}e^{\lambda(S_n - np)} = \mathbb{E} \prod_{i=1}^n e^{\lambda(X_i - p)} = \prod_{i=1}^n \mathbb{E}e^{\lambda(X_i - p)},$$

przy czym skorzystaliśmy z niezależności zmiennych $e^{\lambda(X_i - p)}$. Ponadto, korzystając z nierówności (5), otrzymujemy

$$\mathbb{E}e^{\lambda(X_i - p)} = pe^{\lambda q} + qe^{-\lambda p} \leq e^{\lambda^2/8},$$

Zatem

$$\mathbb{E}e^{\lambda(S_n - np)} \leq e^{\lambda^2 n/8}.$$

Z wykładniczej nierówności Czebyszewa otrzymujemy

$$\mathbb{P}\left(\frac{S_n}{n} \geq p + \varepsilon\right) = \mathbb{P}(S_n - np \geq n\varepsilon) \leq e^{\lambda^2 n/8 - \lambda n\varepsilon}$$

dla $\lambda \geq 0$. Dla każdej liczby nieujemnej λ , powyższa nierówność daje nam pewne oszacowanie na $\mathbb{P}\left(\frac{S_n}{n} \geq p + \varepsilon\right)$. Ponieważ zależy nam na jak najlepszym oszacowaniu należy teraz znaleźć wartość parametru λ , dla której prawa strona jest najmniejsza. Wartością tą jest $\lambda = 4\varepsilon$. Przy tej wartości λ prawa strona jest równa $e^{-2n\varepsilon^2}$, czyli daje oszacowanie, które chcieliśmy udowodnić. \square

Przykład: Wracając do przykładu 2 powyżej, możemy teraz porównać oszacowania uzyskane z nierówności Czebyszewa-Bienaymé i nierówności Bernsteina. Poprzednio uzyskaliśmy nierówność

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2},$$

co dla dużych n jest dużo gorszym oszacowaniem niż

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq 2e^{-2\varepsilon^2 n},$$

uzyskane z nierówności Bernsteina.

22. ZBIEŻNOŚĆ WEDŁUG PRAWDOPODOBIEŃSTWA I PRAWIE NA PEWNO

Często w praktyce mamy do czynienia z ciągiem zmiennych losowych X_1, X_2, \dots i interesuje nas zachowanie graniczne tego ciągu, tzn. np. rozkład zmiennych X_n dla dużych wartości n . Rozważmy następujący przykład. Przypuśćmy, iż chcemy odpowiedzieć na pytanie, czy dana moneta jest symetryczna czy nie. Rzucamy nią wiele razy (na potrzeby tego przykładu załóżmy, że nieskończenie wiele razy) i rozważamy ciąg zmiennych losowych $Y_n = 1_{\{\text{w } n\text{-tym rzucie reszka}\}}$, $n = 1, 2, \dots$. Jasne jest, że ciąg zmiennych

$$X_n = \frac{Y_1 + Y_2 + \dots + Y_n}{n}, \quad n = 1, 2, \dots,$$

powinien nam dać odpowiedź: jeśli, dla dużych n , X_n jest bliskie $1/2$, to mamy prawo przypuszczać, że moneta jest symetryczna; w przeciwnym razie mamy podstawę sądzić, że tak nie jest.

Od razu powstaje problem, w jakim sensie badać „graniczne” zachowanie ciągu $(X_n)_{n \geq 1}$. W rachunku prawdopodobieństwa rozważa się wiele różnych typów zbieżności. My zdefiniujemy tylko dwa z nich.

Definicja 36. *Mówimy, że ciąg $(X_n)_{n \geq 1}$ jest zbieżny prawie na pewno do X , jeśli*

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

Równoważnie, istnieje zdarzenie $\Omega' \subset \Omega$ pełnej miary (tzn. takie, że $\mathbb{P}(\Omega') = 1$) o tej własności, że dla każdego $\omega \in \Omega'$,

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Oznaczenie:

$$\lim_{n \rightarrow \infty} X_n = X \text{ p.n.} \quad \text{lub} \quad X_n \xrightarrow{\text{p.n.}} X.$$

Innym typem zbieżności jest zbieżność według prawdopodobieństwa.

Definicja 37. *Mówimy, że ciąg (X_n) jest zbieżny według prawdopodobieństwa do X , jeśli dla każdego $\varepsilon > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Równoważnie, dla każdego $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \leq \varepsilon) = 1.$$

Oznaczenie: $X_n \xrightarrow{\mathbb{P}} X$.

Nie będziemy tutaj głębiej wnikać we własności, związku i zależności pomiędzy powyższymi rodzajami zbieżności. Poprzestaniemy tylko na stwierdzeniu, iż zbieżność prawie na pewno jest silniejsza niż zbieżność według prawdopodobieństwa: jeśli $X_n \xrightarrow{\text{p.n.}} X$, to $X_n \xrightarrow{\mathbb{P}} X$. Ponadto, zachodzi następujący użyteczny fakt.

Twierdzenie 47. Załóżmy, że $(X_n)_{n \geq 1}$, $(Y_n)_{n \geq 1}$ są ciągami zmiennych losowych. Jeśli $(X_n)_{n \geq 1}$ zbiega do X i $(Y_n)_{n \geq 1}$ zbiega do Y prawie na pewno (odp. według prawdopodobieństwa), to $X_n \pm Y_n \rightarrow X \pm Y$ i $X_n \cdot Y_n \rightarrow XY$ prawie na pewno (odp., według prawdopodobieństwa).

23. PRAWA WIELKICH LICZB

Założmy, że X_1, X_2, \dots jest ciągiem zmiennych losowych. Prawa wielkich liczb mówią o zachowaniu ciągu sum tych zmiennych, tzn. ciągu

$$S_n = X_1 + X_2 + \dots + X_n, \quad n = 1, 2, \dots,$$

czy też raczej ciągu

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}, \quad n = 1, 2, \dots,$$

przy rozmaitych założeniach dotyczących struktury ciągu $(X_n)_{n \geq 1}$.

Rozpocniemy od słabych praw wielkich liczb. Termin „słabe” bierze się stąd, iż w tezie mamy zbieżność ciągu $(S_n/n)_{n \geq 1}$ według prawdopodobieństwa. Mocne prawa wielkich liczb mówią o zbieżności tego ciągu prawie na pewno.

Twierdzenie 48 (Słabe prawo wielkich liczb dla schematu Bernoulliego). *Założmy, że X_1, X_2, \dots są niezależne i mają rozkład*

$$\mathbb{P}(X_n = 1) = p = 1 - \mathbb{P}(X_n = 0), \quad n = 1, 2, \dots$$

Wówczas (S_n/n) zbiega według prawdopodobieństwa do p (tzn. zmiennej losowej stałej, równej p); innymi słowy, dla każdego $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{S_n}{n} - p \right| > \varepsilon \right) = 0.$$

Dowód. Przeprowadziliśmy go już wyżej, przy zastosowaniach nierówności Czebyszewa. \square

Co więcej, jeśli dokładniej przyjrzymy się dowodowi nierówności Czebyszewa, widać, że założenia powyższego twierdzenia można osłabić.

Twierdzenie 49 (Słabe prawo wielkich liczb dla zmiennych nieskorelowanych). *Założmy, że X_1, X_2, \dots jest ciągiem nieskorelowanych zmiennych losowych o wspólnie ograniczonej wariancji. Wówczas $(X_n)_{n \geq 1}$ spełnia słabe prawo wielkich liczb: $(S_n - \mathbb{E}S_n)/n \xrightarrow{\mathbb{P}} 0$, tzn. dla dowolnego $\varepsilon > 0$ mamy*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{S_n - \mathbb{E}S_n}{n} \right| > \varepsilon \right) = 0.$$

Uwagi:

1. Podkreślmy: w powyższym twierdzeniu zmienne X_1, X_2, \dots nie muszą mieć tego samego rozkładu.

2. Często wiadomo jednak, że zmienne X_1, X_2, \dots mają ten sam rozkład o średniej m ; wówczas $\mathbb{E}S_n/n = m$ dla każdego n i teza przybiera prostszą postać

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{S_n}{n} - m \right| > \varepsilon \right) = 0.$$

Przykłady:

1. Załóżmy, że A_1, A_2, \dots jest ciągiem parami niezależnych zdarzeń o prawdopodobieństwach p_1, p_2, \dots . Wówczas

$$\frac{1_{A_1} + 1_{A_2} + \dots + 1_{A_n}}{n} - \frac{p_1 + p_2 + \dots + p_n}{n} \xrightarrow{\mathbb{P}} 0.$$

Zatem, dla dużych n , częstość zachodzenia zdarzeń $(A_n)_{n \geq 1}$ jest w przybliżeniu równa ich teoretycznej częstości.

Aby udowodnić powyższą zbieżność, zauważmy, że zmienne $X_1 = 1_{A_1}, X_2 = 1_{A_2}, \dots$ spełniają założenia SPWL. Istotnie, są one nieskorelowane - są bowiem parami niezależne; ponadto,

$$\text{Var}X_n = \text{Var}1_{A_n} = \mathbb{E}(1_{A_n})^2 - (\mathbb{E}1_{A_n})^2 = \mathbb{E}1_{A_n} - (\mathbb{E}1_{A_n})^2 = p_n - p_n^2 \leq 1/4.$$

2. Jeśli założenie o wspólnej ograniczoności wariancji $(X_n)_{n \geq 1}$ nie jest spełnione, teza nie zachodzi. Przykładowo, załóżmy, że $(\varepsilon_n)_{n \geq 1}$ jest ciągiem niezależnych zmiennych losowych o wspólnym rozkładzie $\mathbb{P}(\varepsilon_i = -1) = \mathbb{P}(\varepsilon_i = 1) = 1/2$ i weźmy $X_n = 3^n \varepsilon_n$. Wówczas $\mathbb{E}X_n = 0, \mathbb{E}S_n = 0$, ponadto $\text{Var}X_n = 3^{2n} \rightarrow \infty$. Z drugiej strony,

$$\left| \frac{S_n - \mathbb{E}S_n}{n} \right| = \left| \frac{S_n}{n} \right| \geq \frac{|X_n| - |X_1 + X_2 + \dots + X_{n-1}|}{n} \geq \frac{3^n}{2n} \rightarrow \infty,$$

a więc zbieżność według prawdopodobieństwa nie ma miejsca.

3. Analogicznie, nie można pozbyć się założenia o nieskorelowaniu ciągu $(X_n)_{n \geq 1}$. Np. niech ε będzie zmienną o rozkładzie $\mathbb{P}(\varepsilon = -1) = \mathbb{P}(\varepsilon = 1) = 1/2$ i połóżmy $X_n = \varepsilon, n = 1, 2, \dots$. Zmienne $(X_n)_{n \geq 1}$ są skorelowane: $\text{Cov}(X_i, X_j) = 1 \neq 0$. Mamy $\mathbb{E}S_n = 0, \text{Var}X_n = 1$ (a zatem założenie o wspólnie ograniczonej wariancji jest spełnione!), ale

$$\left| \frac{S_n - \mathbb{E}S_n}{n} \right| = \left| \frac{S_n}{n} \right| = |X_n| = 1 \not\xrightarrow{\mathbb{P}} 0.$$

Przejdźmy teraz do mocnych praw wielkich liczb. Ponownie, zaczynamy od wersji dla rozkładu Bernoulliego.

Twierdzenie 50 (Mocne prawo wielkich liczb dla schematu Bernoulliego). *Załóżmy, że X_1, X_2, \dots są niezależne i mają rozkład*

$$\mathbb{P}(X_n = 1) = p = 1 - \mathbb{P}(X_n = 0), \quad n = 1, 2, \dots$$

Wówczas (S_n/n) zbiega prawie na pewno do p (tzn. zmiennej losowej stałej, równej p); innymi słowy, istnieje zdarzenie Ω' pełnej miary takie, że jeśli $\omega \in \Omega'$, to

$$\lim_{n \rightarrow \infty} \frac{S_n(\omega)}{n} = p.$$

W tym momencie otrzymujemy, iż formalna definicja prawdopodobieństwa, wprowadzona na pierwszym wykładzie, „pokrywa się” z definicją intuicyjną. Istotnie, dostajemy, iż w celu zdefiniowania prawdopodobieństwa sukcesu, wystarczy wziąć granicę częstości. Ogólniej, mamy następujący fakt.

Twierdzenie 51 (Mocne prawo wielkich liczb Kołmogorowa). *Załóżmy, że X_1, X_2, \dots jest ciągiem niezależnych całkowalnych zmiennych losowych o tym samym rozkładzie. Wówczas*

$$\frac{S_n}{n} \xrightarrow{p.n.} \mathbb{E}X_1.$$

Zatem intuicyjna definicja średniej zmiennej losowej pokrywa się z teoretyczną: wartością oczekiwaną.

Pewną niedogodnością związaną z prawami wielkich liczb (zwłaszcza mocnymi) jest to, iż nie wiemy nic o prędkości zbieżności do granicy - nie mamy oszacowania na błąd związany z przybliżeniem (S_n/n) za pomocą średniej $\mathbb{E}X_1$. Kłopot ten (częściowo) pozwala przezwyciężyć Centralne Twierdzenie Graniczne i nierówność Berry-Esséena, omawiane w dalszej części wykładu.

24. ZASTOSOWANIE PRAW WIELKICH LICZB: ZBIEŻNOŚĆ ŚREDNIEJ I WARIANCJI Z PRÓBKĄ, DYSTRYBUANTA EMPIRYCZNA I TWIERDZENIE GLIWENKI-CANTELLIEGO

Załóżmy, że $(X_n)_{n \geq 1}$ jest ciągiem całkowalnych niezależnych zmiennych losowych o tym samym rozkładzie. Wówczas MPWL mówi, iż

$$(*) \quad \bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{p.n.} \mathbb{E}X_1.$$

Analogicznie, jeśli $(X_n)_{n \geq 1}$ jest ciągiem zmiennych o tym samym rozkładzie, całkowalnych z kwadratem, to

$$(**) \quad \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2 \xrightarrow{p.n.} \text{Var} X_1.$$

Istotnie, mamy

$$\frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2 = \frac{1}{n} \sum_{k=1}^n (X_k^2 - 2X_k\bar{X} + \bar{X}^2) = I_1 - I_2 + I_3,$$

gdzie

$$I_1 = \frac{X_1^2 + X_2^2 + \dots + X_n^2}{n} \xrightarrow{p.n.} \mathbb{E}X_1^2,$$

na mocy MPWL zastosowanego do ciągu (X_n^2) całkowalnych niezależnych zmiennych o tym samym rozkładzie,

$$I_2 = \frac{2}{n} \sum_{k=1}^n X_k \bar{X} = 2\bar{X}^2 \xrightarrow{p.n.} 2(\mathbb{E}X_1)^2,$$

na mocy powyższego, oraz

$$I_3 = \bar{X}^2 \xrightarrow{p.n.} (\mathbb{E}X_1)^2.$$

Stąd żądana zbieżność.

Powyższe dwie zbieżności (*) oraz (**) mają istotne znaczenie w zastosowaniach. Istotnie, jeśli dysponujemy (liczną) próbką X_1, X_2, \dots, X_N danych pochodzących z tego samego (nieznanego) rozkładu, widzimy, iż dobrym przybliżeniem średniej oraz wariancji są średnia oraz wariancja empiryczna.

Zajmiemy się teraz kolejnym zastosowaniem praw wielkich liczb: dystrybuantą empiryczną. Jak wiemy, dysponując próbką danych możemy rozważać rozkład empiryczny μ_N związany z tą próbką; załóżmy, że X_1, X_2, \dots, X_N jest ciągiem niezależnych zmiennych losowych. Określamy

$$\mu_N(A) = \frac{1_A(X_1) + 1_A(X_2) + \dots + 1_A(X_N)}{N}.$$

Z MPWL widzimy, iż $\mu_N(A) \xrightarrow{p.n.} \mathbb{E}1_A(X_1) = \mathbb{P}(X_1 \in A)$, a zatem rozkład empiryczny zbiega do wspólnego rozkładu zmiennych (X_n) . Aby uniknąć pojawiających się tutaj licznych problemów technicznych (tempo zbieżności zależy od zbioru A), posłużymy się dystrybuantami rozkładów empirycznych. Przypomnijmy, jeśli X_1, X_2, \dots, X_N są zmiennymi losowymi, to funkcję $F_N : \mathbb{R} \rightarrow [0, 1]$, daną wzorem

$$F_N(t) = \frac{1_{\{X_1 \leq t\}} + 1_{\{X_2 \leq t\}} + \dots + 1_{\{X_N \leq t\}}}{N},$$

nazywamy N -tą dystrybuantą empiryczną.

Zauważmy, że F_N jest zmienną losową; co więcej, dla każdego $\omega \in \Omega$, $F_N(\omega)$ jest dystrybuantą. Mocne prawo wielkich liczb implikuje, iż jeśli zmienne X_n mają ten sam rozkład i są niezależne, to dystrybuanta ta przybliża dystrybuantę wspólnego rozkładu zmiennych X_n . Okazuje się, iż można udowodnić więcej: zbieżność jednostajną. Mamy następujący fakt.

Twierdzenie 52 (Twierdzenie Gliwienki-Cantelliego - podstawowe twierdzenie statystyki). *Załóżmy, że X_1, X_2, \dots są niezależne i mają ten sam rozkład o dystrybuancie F . Wówczas*

$$\sup_{t \in \mathbb{R}} |F_N(t) - F(t)| \xrightarrow[N \rightarrow \infty]{p.n.} 0.$$

25. CENTRALNE TWIERDZENIE GRANICZNE

Jak wiemy, jeśli X_1, X_2, \dots są niezależnymi zmiennymi losowymi o tym samym całkowalnym rozkładzie, to

$$\frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{p.n.} \mathbb{E}X_1.$$

Powstaje bardzo naturalne pytanie: jak dobre jest to przybliżenie dla dużych n ? Co można powiedzieć o błędzie tego przybliżenia?

Poniżej formułujemy najprostszą wersję Centralnego Twierdzenia Granicznego. Niech $\Phi : \mathbb{R} \rightarrow [0, 1]$,

$$\Phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx$$

oznacza dystrybuantę standardowego rozkładu normalnego. Odnotujmy ważną zależność: ponieważ rozkład $\mathcal{N}(0, 1)$ jest rozkładem symetrycznym, to zachodzi $\Phi(t) + \Phi(-t) = 1$ dla dowolnej liczby rzeczywistej t .

Twierdzenie 53. *Załóżmy, że X_1, X_2, \dots są niezależnymi zmiennymi losowymi o tym samym rozkładzie, takim, że $\mathbb{E}X_1^2 < \infty$. Oznaczmy $m = \mathbb{E}X_1$, $\sigma^2 = \text{Var} X_1$. Wówczas dla dowolnego $t \in \mathbb{R}$,*

$$\mathbb{P}\left(\frac{X_1 + X_2 + \dots + X_n - nm}{\sigma\sqrt{n}} \leq t\right) \xrightarrow{n \rightarrow \infty} \Phi(t).$$

Uwaga:

1. Łatwo udowodnić, iż powyższa zbieżność pociąga za sobą, iż dla dowolnego s ,

$$\mathbb{P}\left(\frac{X_1 + X_2 + \dots + X_n - nm}{\sigma\sqrt{n}} \geq s\right) \xrightarrow{n \rightarrow \infty} 1 - \Phi(s),$$

i ogólniej, dla dowolnych $s < t$,

$$\mathbb{P}\left(s \leq \frac{X_1 + X_2 + \dots + X_n - nm}{\sigma\sqrt{n}} \leq t\right) \xrightarrow{n \rightarrow \infty} \Phi(t) - \Phi(s).$$

Co więcej, dowolną z nierówności po lewej stronie możemy zmienić z nieostrej na ostrą, i zbieżności wciąż będą miały miejsce (z tymi samymi granicami).

2. CTG odnosi się do błędu związanego z MPWL. Istotnie, zauważmy np., że nierówność

$$\frac{X_1 + X_2 + \dots + X_n - nm}{\sigma\sqrt{n}} \leq t$$

jest równoważna

$$\frac{X_1 + X_2 + \dots + X_n}{n} - m \leq \frac{t\sigma}{\sqrt{n}},$$

a zatem odnosi się do zdarzenia, iż błąd („jednostronny”) związany z przybliżeniem średniej S_n/n przez jej teoretyczny odpowiednik m , nie przekracza progu $t\sigma/\sqrt{n}$.

Szczególnym przypadkiem CTG, mianowicie odnoszącym się do schematu Bernoulliego, jest twierdzenie de Moivre’a-Laplace’a.

Twierdzenie 54 (de Moivre’a-Laplace’a). *Załóżmy, że X_1, X_2, \dots jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie*

$$\mathbb{P}(X_n = 1) = p, \mathbb{P}(X_n = 0) = q, \quad p + q = 1.$$

Wówczas dla $s < t$,

$$\mathbb{P}\left(s \leq \frac{X_1 + X_2 + \dots + X_n - np}{\sqrt{npq}} \leq t\right) \xrightarrow{n \rightarrow \infty} \Phi(t) - \Phi(s)$$

(każdą z nierówności po lewej stronie możemy zamienić na ostrą).

Powyższe twierdzenia wciąż nie pozwalają precyzyjnie oszacować błędu. Możliwość tę daje następujące twierdzenie.

Twierdzenie 55 (Berry-Esséena). *Załóżmy, że (X_n) jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie takim, że $\mathbb{E}|X_1|^3 < \infty$. Oznaczmy $m = \mathbb{E}X_1$ i $\sigma = \sqrt{\text{Var}X_1}$. Wówczas*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\frac{X_1 + X_2 + \dots + X_n - nm}{\sigma\sqrt{n}} \leq t\right) - \Phi(t) \right| \leq C \frac{\mathbb{E}|X_1 - \mathbb{E}X_1|^3}{\sigma^3\sqrt{n}},$$

gdzie $C \in [1/\sqrt{2\pi}; 0, 77]$.

W praktyce będziemy jednak stosować tylko przybliżenie płynące bezpośrednio z Centralnego Twierdzenia Granicznego. Powyższe twierdzenie formułujemy wyłącznie w celach informacyjnych.

Przykłady:

1. Rzucono monetą 10 000 razy i okazało się, że orzeł wypadł 5200 razy. Czy są podstawy do przypuszczenia, że moneta jest niesymetryczna?

Załóżmy, że moneta była symetryczna i zobaczymy, jakie jest prawdopodobieństwo wypadnięcia nie mniej niż 5200 orłów. Rozważmy ciąg zmiennych określonych wzorem $X_i = 1_{\{\text{wypadł orzeł w } i\text{-tym rzucie}\}}$, $i = 1, 2, \dots, 10\,000$. Zmienne $X_1, X_2, \dots, X_{10\,000}$ są niezależne i mają ten sam rozkład $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = 1/2$. Obliczamy, iż $m = \mathbb{E}X_1 = 1/2$ oraz $\sigma = \sqrt{\text{Var}X_1} = 1/2$. Na mocy twierdzenia de Moivre’a-Laplace’a,

$$\begin{aligned} \mathbb{P}(X_1 + X_2 + \dots + X_{10\,000} \geq 5200) &= \mathbb{P}(X_1 + X_2 + \dots + X_{10\,000} - 10\,000 \cdot m \geq 5200 - 5000) \\ &= \mathbb{P}\left(\frac{X_1 + X_2 + \dots + X_{10\,000} - 10\,000 \cdot m}{\sigma\sqrt{10\,000}} \geq \frac{200}{50}\right) \approx 1 - \Phi(4). \end{aligned}$$

Sprawdzamy w tablicach, iż prawa strona jest w przybliżeniu równa 0,00003. Tak więc rozważane zdarzenie ma bardzo małe prawdopodobieństwo; są więc podstawy by sądzić, że moneta nie jest symetryczna.

2. Stwierdzono, że przeciętnie 30% spośród ogólnej liczby studentów przyjętych na studia kończy je w terminie. Ile trzeba przyjąć studentów na pierwszy rok, aby z prawdopodobieństwem w przybliżeniu 0,9, co najmniej 50 osób ukończyło studia w terminie?

Założmy, że przyjęto N osób na pierwszy rok. Wprowadźmy zmienne losowe $X_i = 1_{\{i\text{-ta osoba ukończy studia w terminie}\}}$, $i = 1, 2, \dots, N$. Przyjmujemy, że zmienne X_i są niezależne i zauważamy, że mają one ten sam dwupunktowy rozkład

$$\mathbb{P}(X_i = 1) = 0,3, \quad \mathbb{P}(X_i = 0) = 0,7.$$

Stąd obliczamy, iż $m = \mathbb{E}X_1 = 0,3$, $\sigma = \sqrt{\text{Var}X_1} = \sqrt{0,3 \cdot 0,7} \approx 0,46$. Interesuje nas zdarzenie

$$\{X_1 + X_2 + \dots + X_N \geq 50\},$$

czyli, równoważnie (sprowadzamy nierówność do postaci jak w twierdzeniu de Moivre'a-Laplace'a),

$$\left\{ \frac{X_1 + X_2 + \dots + X_N - Nm}{\sigma\sqrt{N}} \geq \frac{50 - 0,3N}{0,46\sqrt{N}} \right\}.$$

Korzystając z twierdzenia de Moivre'a-Laplace'a, prawdopodobieństwo powyższego zdarzenia wynosi w przybliżeniu

$$1 - \Phi\left(\frac{50 - 0,3N}{0,46\sqrt{N}}\right).$$

Powstaje więc pytanie, dla jakich N powyższa liczba jest w przybliżeniu równa 0,9. Z tablic rozkładu normalnego odczytujemy, iż $1 - \Phi(-1,29) \approx 0,90147$, zatem wystarczy wziąć N takie, by $50 - 0,3N/0,46\sqrt{N}$ było jak najbliższej $-1,29$; ma to miejsce dla $N = 194$. Tak więc należy przyjąć co najmniej 195 osób na pierwszy rok.

3. Sumujemy 400 liczb, każdą zaokrągloną z dokładnością do 10^{-2} . Założmy, że błędy spowodowane przez zaokrąglenia są niezależnymi zmiennymi losowymi o rozkładzie jednostajnym na $[-10^{-2}, 10^{-2}]$. Jakie jest prawdopodobieństwo, że błąd całkowity jest większy niż 0,1?

Niech dla $i = 1, 2, \dots, 400$, zmienna X_i będzie błędem powstałym przy zaokrągleniu i -tej liczby. Na mocy warunków zadania, zmienne X_i spełniają założenia CTG. Mamy $m = \mathbb{E}X_1 = 0$, $\sigma = \sqrt{\text{Var}X_1} \approx 0,006$, zatem

$$\begin{aligned} \mathbb{P}(X_1 + X_2 + \dots + X_{400} > 0,1) &= \mathbb{P}\left(\frac{X_1 + X_2 + \dots + X_{400} - 400 \cdot 0}{0,006\sqrt{400}} > \frac{0,1}{0,12}\right) \\ &\approx 1 - \Phi(0,1/0,12) \approx 0,202. \end{aligned}$$

4. Kolejnym przykładem, grającym ważną rolę w zastosowaniach, są tzw. *prze-działy ufności*. Przypuśćmy, iż dysponujemy (liczną) próbką X_1, X_2, \dots, X_n pochodzącą z rozkładu z pewnym nieznanym parametrem θ (np. wykonujemy ciąg 10 000 rzutów monetą o prawdopodobieństwie wypadnięcia orła wynoszącym θ

(którego nie znamy)). Przedziałem ufności (θ_1, θ_2) o współczynniku (poziomie) ufności $1 - \alpha$ nazywamy przedział taki, że

$$\mathbb{P}(\theta \in (\theta_1, \theta_2)) \geq 1 - \alpha;$$

θ_1 i θ_2 to pewne funkcje wyznaczone przez próbę X_1, X_2, \dots, X_n . Oczywiście, z punktu widzenia zastosowań, istotne jest, aby ów przedział był możliwie jak najkrótszy.

Weźmy pod uwagę konkretny przykład. Załóżmy, że próba X_1, X_2, \dots, X_n pochodzi z rozkładu jednostajnego o nieznannej średniej m i wariancji 1. Przypuśćmy, iż naszym zadaniem jest wyznaczyć, na podstawie tej próby, przedział (a, b) taki, że

$$(*) \quad \mathbb{P}(a < m < b) > 0,9.$$

Otóż, jak wiemy, dobrym kandydatem na przybliżenie średniej m jest jej średnia empiryczna \bar{X} . Naturalnym pomysłem jest więc, aby wziąć

$$a = \bar{X} - \varepsilon \quad \text{oraz} \quad b = \bar{X} + \varepsilon,$$

dla pewnego $\varepsilon > 0$, który wyznaczymy ze związku (*). Równoważnie, nierówność ta wygląda następująco:

$$\mathbb{P}(-\varepsilon < \bar{X} - m < \varepsilon) > 0,9.$$

Teraz przekształcamy ją tak, aby uzyskać postać jak w Centralnym Twierdzeniu Granicznym. Po pomnożeniu nierówności występującej pod prawdopodobieństwem przez \sqrt{n} i podzieleniu przez $\sqrt{\text{Var}X_1}$ dostajemy

$$\mathbb{P}\left(-\frac{\varepsilon\sqrt{n}}{\sqrt{\text{Var}X_1}} < \frac{X_1 + X_2 + \dots + X_n - nm}{\sqrt{n\text{Var}X_1}} < \frac{\varepsilon\sqrt{n}}{\sqrt{\text{Var}X_1}}\right) > 0,9.$$

Na mocy CTG (zmiennie X_i są niezależne i mają ten sam rozkład, całkowalny z kwadratem), powyższe prawdopodobieństwo jest w przybliżeniu równe $\Phi(\varepsilon\sqrt{n}) - \Phi(-\varepsilon\sqrt{n}) = 2\Phi(\varepsilon\sqrt{n}) - 1$ (tu korzystamy z równości $\text{Var}X_1 = 1$ oraz $\Phi(t) + \Phi(-t) = 1$ dla $t \in \mathbb{R}$). W tablicach sprawdzamy, że $2\Phi(1,64) - 1 \approx 0,9$; stąd bierzemy $\varepsilon = 1,64/\sqrt{n}$.

Tak więc, jeśli dysponujemy próbką o liczności 900, to przedziałem ufności dla m o poziome ufności 0,9 jest przedział

$$(\bar{X} - 0,055, \bar{X} + 0,055).$$

Zwróćmy uwagę: zgodnie z intuicją, im liczniejsza próbka, tym węższy przedział ufności.

26. ŁAŃCUCZY MARKOWA

Do tej pory koncentrowaliśmy się głównie na badaniu niezależnych zmiennych losowych. Choć założenie o niezależności badanych zmiennych jest często z powodzeniem stosowane w praktyce, z reguły stanowi ono dość duże uproszczenie rzeczywistości, gdyż charakterystyki większości realnych zjawisk są ze sobą nietrywialnie powiązane. Dla przykładu, wartości indeksów giełdowych mogą zależeć od nastrojów wśród inwestorów, które z kolei są uwarunkowane wydarzeniami ostatnich kilku miesięcy (w tym zachowaniem indeksów giełdowych w tym okresie). Modelowanie matematyczne zjawisk fizycznych czy społecznych polega często na znalezieniu kompromisu między wiernym opisem czynników wpływających na

przebieg danego zjawiska i prostotą budowanego modelu. Łańcuchy Markowa, jeden z klasycznych obiektów badań w rachunku prawdopodobieństwa, są wygodnym narzędziem do budowania modeli bardziej realistycznych niż te oparte o niezależne zmienne losowe, a równocześnie na tyle prostych, by poddały się analizie i pozwalały na przewidywanie przebiegu zjawisk. Są one szczególnie przydatne do modelowania procesów zachodzących w czasie. Ich prostota wynika z założenia, że przebieg procesu w przyszłości zależy jedynie od jego stanu w chwili obecnej.

Nie będziemy tutaj omawiać dokładnie teorii łańcuchów Markowa, skoncentrujemy się na tzw. jednorodnych łańcuchach o czasie dyskretnym i skończonej przestrzeni stanów. Aby zrozumieć podstawową ideę, rozważmy następujący, prosty przykład.

Pan X spędza wieczory w domu, w kinie lub w barze.

- (1) jeżeli konkretnego dnia pan X był w domu, następnego dnia idzie do kina lub baru (w każde z tych miejsc z prawdopodobieństwem $1/2$);
- (2) jeżeli był w barze, następnego dnia z prawdopodobieństwem $3/4$ boli go głowa i zostaje w domu, z prawdopodobieństwem $1/8$ idzie znów do baru, a z prawdopodobieństwem $1/8$ idzie do kina.
- (3) jeżeli był w kinie, następnego dnia z prawdopodobieństwem $1/2$ idzie do baru, z prawdopodobieństwem $1/4$ do kina, z prawdopodobieństwem $1/4$ zostaje w domu.

Widzimy więc, że pan X w każdy wieczór przebywa w jednym z trzech miejsc. Jeśli oznaczymy je symbolicznie przez 1, 2, 3 (odp. bar, kino, dom), zachowanie pana X w kolejnych dniach (oznaczymy je numerami $0, 1, 2, 3, \dots$) możemy opisać przez ciąg zmiennych losowych X_0, X_1, X_2, \dots , przyjmujących wartości w zbiorze $E = \{1, 2, 3\}$ – tzw. przestrzeni stanów modelu. Zmienne X_i nie są niezależne, natomiast cała informacja na temat zmiennej X_n zawarta w historii procesu (tzn. zmiennych X_0, X_1, \dots, X_{n-1}) zawarta jest w ostatniej zmiennej X_{n-1} ; ściślej,

$$\mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}),$$

dla dowolnego ciągu x_1, \dots, x_n dla którego prawdopodobieństwo warunkowe po lewej stronie jest dobrze określone (wówczas prawdopodobieństwo warunkowe po prawej stronie jest też dobrze określone).

Co więcej, prawa strona w naszej sytuacji nie zależy od n , odpowiednie prawdopodobieństwa warunkowe są takie same w kolejne dni (nasz proces jest *jednorodny w czasie*, por. definicja poniżej), np.

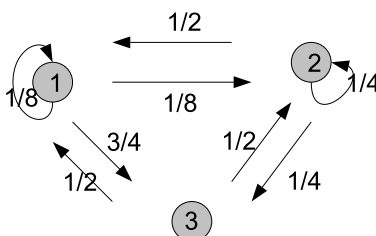
$$\begin{aligned} \mathbb{P}(X_n = 1 | X_{n-1} = 3) &= \mathbb{P}(n\text{-tego dnia pan } X \text{ był w barze} \mid \text{w dniu } (n-1) \text{ był w domu}) = 1/2, \\ \mathbb{P}(X_n = 2 | X_{n-1} = 2) &= \mathbb{P}(n\text{-tego dnia pan } X \text{ był w kinie} \mid \text{w dniu } (n-1) \text{ był w kinie}) = 1/4 \\ \mathbb{P}(X_n = 3 | X_{n-1} = 3) &= \mathbb{P}(n\text{-tego dnia pan } X \text{ był w domu} \mid \text{w dniu } (n-1) \text{ był w domu}) = 0. \end{aligned}$$

Oznaczając $p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$, możemy opisać dynamikę naszego procesu przez macierz $P = (p_{ij})_{i,j=1,2,3}$. Element p_{ij} , stojący na przecięciu i -tego wiersza i j -tej kolumny macierzy informuje nas jakie jest prawdopodobieństwo, że proces przejdzie ze stanu i do stanu j .

W naszym przypadku, odpowiednia macierz ma postać

$$P = \begin{bmatrix} \frac{1}{8} & \frac{1}{8} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

Alternatywnym i bardziej intuicyjnym opisem naszego procesu jest graf, pokazujący jakie są prawdopodobieństwa przejścia z jednych stanów do innych. W dalszej części wykładu nie będziemy posługiwać się grafami, wskazane jest jednak, aby przy każdym przykładzie zrobili Państwo odpowiedni rysunek: istotnie ułatwi on zrozumienie problemu. My ograniczymy się do graficznego zilustrowania tej jednej sytuacji, przedstawionej na poniższym rysunku:



Aby mieć pełną wiedzę na temat procesu, musimy jeszcze znać rozkład zmiennej X_0 . Jeśli wiemy z jakim prawdopodobieństwem pan X przebywał w barze, kinie i w domu zerowego dnia, korzystając z macierzy P możemy obliczyć jakie są odpowiednie prawdopodobieństwa w dniu n . Więcej na ten temat powiemy w dalszej części wykładu.

Wprowadźmy zatem następującą definicję

Definicja 38. Ciąg zmiennych losowych $(X_n)_{n=0}^{\infty}$ o wartościach w skończonym zbiorze E (tzw. przestrzeni stanów) nazwiemy łańcuchem Markowa, jeśli dla dowolnego $n = 1, 2, 3, \dots$ oraz dla dowolnego ciągu x_0, x_1, \dots, x_n elementów zbioru E zachodzi równość

$$\mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}),$$

o ile tylko $\mathbb{P}(X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) > 0$.

Jeśli dla dowolnych $i, j \in E$, $\mathbb{P}(X_n = j | X_{n-1} = i)$ nie zależy od n , łańcuch nazywamy jednorodnym lub jednorodnym w czasie. W takiej sytuacji definiujemy macierz przejścia łańcucha $P = (p_{ij})_{i,j \in E}$ wzorem

$$p_{ij} = \mathbb{P}(X_1 = j | X_0 = i)$$

Liczby p_{ij} nazywamy prawdopodobieństwami przejścia.

Uwagi:

1. Suma elementów w każdym wierszu macierzy P wynosi 1. Wynika to ze wzoru na prawdopodobieństwo całkowite.

2. Jeżeli $\mathbb{P}(X_n = j | X_{n-1} = i)$ zależy od n , możemy rozpatrywać tzw. macierz przejścia w n -tym kroku $P^{(n)} = (p_{ij}^{(n)})$, daną wzorem

$$p_{ij}^{(n)} = \mathbb{P}(X_n = j | X_{n-1} = i).$$

Teoria tego typu łańcuchów jest jednak bardziej skomplikowana.

3. Nie musimy też ograniczać się do skończonej przestrzeni stanów. Łańcuchy Markowa na przestrzeni przeliczalnej definiuje się w sposób analogiczny. Również definicja macierzy przejścia się przenosi, jest ona jednak w tym wypadku nieskończona. Łańcuchy tego typu pojawiają się w zastosowaniach dość naturalnie, mogą one np. opisywać cząstkę błądzącą po liczbach całkowitych i przechodzącą z i do $i+1$ z prawdopodobieństwem p , zaś do $i-1$ z prawdopodobieństwem $1-p$. Tego typu zagadnienia stanowią podstawę np. do analizy gier losowych, z powodu ograniczeń czasowych nie będziemy się jednak nimi zajmować i poprzestaniemy na uwadze, że zdecydowana większość podanych niżej faktów przenosi się bez zmian na łańcuchy o przeliczalnej liczbie stanów.

4. Jak już wspomnieliśmy, łańcuchy Markowa służą do opisu procesów, w których przyszłość zależy tylko od teraźniejszości (czyli dla prognozowania przyszłości ważny jest tylko obecny stan procesu, a nie to w jaki sposób proces rozwijał się w przeszłości). Łatwo zauważyć, że zmieniając przestrzeń E , można przy pomocy łańcuchów Markowa modelować również procesy, w których wartość X_{n+1} zależy np. tylko od ostatnich dwóch zmiennych, X_{n-1} i X_n . Należy w tym celu rozpatrywać łańcuch $Y_n = (X_n, X_{n+1}) \in \tilde{E} = E \times E$.

Zajmiemy się teraz wyznaczaniem rozkładu zmiennej X_n . Jak wiemy, aby określić rozkład prawdopodobieństwa na zbiorze E , o n elementach, należy podać n liczb. Możemy zatem utożsamiać rozkłady prawdopodobieństwa na E z wektorami, zapisując je w postaci $(p_i)_{i \in E}$. Aby to zilustrować, wróćmy do przykładu z panem X . Jeżeli w dniu 0 był on w barze z prawdopodobieństwem $1/4$, w kinie z prawdopodobieństwem $1/4$ i w domu z prawdopodobieństwem $1/2$, mamy $\mathbb{P}(X_0 = 1) = \mathbb{P}(X_0 = 2) = 1/4$, $\mathbb{P}(X_0 = 3) = 1/2$ i możemy powiedzieć, że rozkład zmiennej X_0 jest zadany przez wektor

$$Q = (Q_1, Q_2, Q_3) = (1/4, 1/4, 1/2).$$

Jeśli teraz chcemy np. wyznaczyć prawdopodobieństwo, że pierwszego dnia pan X był w domu (czyli $\mathbb{P}(X_1 = 3)$), możemy skorzystać ze wzoru na prawdopodobieństwo całkowite:

$$\begin{aligned} \mathbb{P}(X_1 = 3) &= \mathbb{P}(X_1 = 3|X_0 = 1)\mathbb{P}(X_0 = 1) + \mathbb{P}(X_1 = 3|X_0 = 2)\mathbb{P}(X_0 = 2) \\ &\quad + \mathbb{P}(X_1 = 3|X_0 = 3)\mathbb{P}(X_0 = 3) \\ &= \frac{3}{4} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} = \frac{1}{4}. \end{aligned}$$

Ogólniej

$$\begin{aligned} \mathbb{P}(X_1 = j) &= \mathbb{P}(X_1 = j|X_0 = 1)\mathbb{P}(X_0 = 1) + \mathbb{P}(X_1 = j|X_0 = 2)\mathbb{P}(X_0 = 2) \\ &\quad + \mathbb{P}(X_1 = j|X_0 = 3)\mathbb{P}(X_0 = 3) \\ &= Q_1 p_{1j} + Q_2 p_{2j} + Q_3 p_{3j}. \end{aligned}$$

W powyższym wzorze możemy rozpoznać mnożenie wektora przez macierz. Jest to jedna z przyczyn, dla których rozkłady prawdopodobieństwa i prawdopodobieństwa przejścia zapisujemy jako wektory i macierze. W ogólnej sytuacji ilustrują to poniższa definicja i twierdzenie.

Definicja 39. Jeżeli X_0, X_1, X_2, \dots jest łańcuchem Markowa na E , to rozkład zmiennej X_0 (o którym myślimy jako o wektorze $Q = (Q_i)_{i \in E}$) nazywamy rozkładem początkowym łańcucha.

Twierdzenie 56. Jeśli X_0, X_1, X_2, \dots jest łańcuchem Markowa o rozkładzie początkowym Q i macierzy przejścia P , to zmienna X_n ma rozkład

$$QP^n.$$

Innymi słowy dla dowolnego $j \in E$ zachodzi

$$\mathbb{P}(X_n = j) = \sum_{i_0 \in E} \sum_{i_1 \in E} \cdots \sum_{i_{n-1} \in E} Q_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} j}.$$

Podobnie

$$\mathbb{P}(X_n = j | X_0 = i) = \sum_{i_1 \in E} \cdots \sum_{i_{n-1} \in E} p_{i i_1} p_{i_1 i_2} \cdots p_{i_{n-1} j},$$

co oznacza, że tzw. macierz przejścia w n krokach (nie będziemy jej formalnie definiować, jej znaczenie powinno być jasne z kontekstu) jest równa P^n .

Uwaga: Powyższe wzory mogą się wydawać dość skomplikowane, nie należy jednak się ich obawiać. Wynikają one ze wzorów na prawdopodobieństwo całkowite i ze wzoru „łańcuchowego” na prawdopodobieństwo iloczynu zdarzeń. Dokładniej, prawdopodobieństwo zdarzenia $\{X_n = j\}$ obliczamy sumując prawdopodobieństwa wszystkich możliwych sytuacji, które mogły wydarzyć się pomiędzy czasem 0 i czasem n , czyli zdarzeń postaci

$$\{X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = j\}.$$

Przykładowo, jeśli zerowego dnia pan X był w barze, mamy

$$\begin{aligned} \mathbb{P}(X_2 = 1, X_1 = 1, X_0 = 1) &= \mathbb{P}(X_2 = 1 | X_0 = 1, X_1 = 1) \mathbb{P}(X_0 = 1, X_1 = 1) \\ &= \mathbb{P}(X_2 = 1 | X_0 = 1, X_1 = 1) \mathbb{P}(X_1 = 1 | X_0 = 1) \mathbb{P}(X_0 = 1) \\ &= \mathbb{P}(X_2 = 1 | X_1 = 1) \mathbb{P}(X_1 = 1 | X_0 = 1) \mathbb{P}(X_0 = 1) \\ &= p_{11} p_{11} \mathbb{P}(X_0 = 1) \\ &= \frac{1}{8} \cdot \frac{1}{8} \cdot 1, \end{aligned}$$

przy czym w trzeciej równości skorzystaliśmy z faktu, że zmienne X_i tworzą łańcuch Markowa.

Interpretując wzory w powyższym twierdzeniu w graficznej reprezentacji łańcucha Markowa, rozpatrujemy wszystkie ścieżki długości n , którymi możemy dojść od punktu i do j , dla każdej takiej ścieżki przemnażamy odpowiadające jej krawędziom wagi i dodajemy otrzymane iloczyny.

Zajmiemy się teraz dwoma zagadnieniami, związanymi z analizą łańcuchów Markowa. Pierwsze dotyczy przybliżania rozkładu zmiennej X_n dla dużych n . Choć rozkład ten może zależeć od n , okazuje się, że przy dość ogólnych założeniach, rozkład X_n zbiega do rozkładu granicznego. Przed wyjaśnieniem szczegółów, wprowadźmy kilka definicji. Poniżej będziemy używać oznaczenia

$$p_{ij}(n) = \mathbb{P}(X_n = j | X_0 = i)$$

na prawdopodobieństwo przejścia w n krokach ze stanu i do stanu j . Tak więc $p_{ij}(1) = p_{ij}$, zaś $p_{ij}(n)$ jest odpowiednim wyrazem macierzy P^n (zgodnie z Twierdzeniem 56).

Definicja 40. *Łańcuch Markowa nazwiemy nieprzywiedlnym, jeśli dla dowolnych $i, j \in E$ istnieje $n > 0$, takie że $p_{ij}(n) > 0$.*

Oznacza to, że z każdego stanu możemy po pewnym czasie dojść do dowolnego innego (każde dwa stany się *komunikują*). W terminach grafów, oznacza to, że z dowolnego wierzchołka można dotrzeć do dowolnego innego idąc wzdłuż strzałek, zgodnie z ich kierunkiem (przyjmujemy, że w grafie nie ma strzałek z wagą równą 0, proszę zauważyć, że w przykładzie dot. pana X nie ma strzałki łączącej wierzchołek nr 3 z nim samym).

Przykłady:

1. Łańcuch opisujący wieczorne wędrówki pana X jest nieprzywiedlny.
2. Łańcuch na przestrzeni stanów $E = \{1, 2\}$, zadany przez prawdopodobieństwa przejścia $p_{11} = p_{12} = 1/2$, $p_{22} = 1 = 1 - p_{21}$ nie jest nieprzywiedlny, gdyż ze stanu 2 nie można przejść do stanu 1.

Definicja 41. *Okresem $o(j)$ stanu j nazywamy największy wspólny dzielnik liczb ze zbioru $\{n: p_{ii}(n) > 0\}$. Stan j nazywamy okresowym jeśli $o(j) > 1$, i nieokresowym jeśli $o(j) = 1$.*

Przykłady:

1. W przykładzie dot. pana X , wszystkie stany są nieokresowe, do stanów 1,2 możemy powrócić już w jednym kroku, zaś do stanu 3 możemy powrócić po dwóch ($3 \rightarrow 1 \rightarrow 3$) i po trzech ($3 \rightarrow 1 \rightarrow 1 \rightarrow 3$) krokach (a największy wspólny dzielnik liczb 2 i 3 to 1).

2. W łańcuchu na przestrzeni stanów $E = \{1, 2, 3\}$, w którym ze stanów 1,3 zawsze przechodzimy do 2 a ze stanu 2 przechodzimy z prawdopodobieństwem $1/2$ do 1 i z takim samym prawdopodobieństwem do 3, wszystkie stany mają okres równy 2. Rzeczywiście, analiza odpowiedniego rysunku pokazuje, że powrót do każdego stanu możliwy jest tylko po parzystej liczbie kroków, co więcej zawsze można wrócić już po dwóch krokach.

W powyższych przykładach wszystkie stany miały ten sam okres. Okazuje się że nie jest to przypadek, prawdziwe jest następujące twierdzenie.

Twierdzenie 57. *W nieprzywiedlnym łańcuchu Markowa wszystkie stany mają ten sam okres.*

Definicja 42. *Nieprzywiedlny łańcuch Markowa, w którym wszystkie stany mają okres 1 nazywamy łańcuchem nieokresowym, w przeciwnym wypadku łańcuch nazywamy okresowym.*

Aby przedstawić twierdzenie opisujące zachowanie graniczne nieprzywiedlnego, nieokresowego łańcucha Markowa, potrzebujemy jeszcze jednej definicji.

Definicja 43. *Rozkład prawdopodobieństwa π na E (reprezentowany jako wektor $(\pi_i)_{i \in E}$) nazwiemy rozkładem stacjonarnym łańcucha Markowa, jeśli*

$$\pi P = \pi,$$

gdzie P oznacza macierz przejścia łańcucha.

Rozkład stacjonarny ma więc następującą własność, tłumaczącą nazwę „stacjonarny”: mianowicie, jeżeli X_0 ma rozkład π , to wszystkie zmienne X_i będą także miały rozkład π .

Przykład: Aby znaleźć rozkład stacjonarny łańcucha, rozwiązujemy układ równań liniowych. Dla łańcucha z naszego głównego przykładu jest to układ

$$(\pi_1, \pi_2, \pi_3) \begin{bmatrix} \frac{1}{8} & \frac{1}{8} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} = (\pi_1, \pi_2, \pi_3),$$

uzupełniony o równanie $\pi_1 + \pi_2 + \pi_3 = 1$, czyli

$$\begin{aligned} \frac{1}{8}\pi_1 + \frac{1}{2}\pi_2 + \frac{1}{2}\pi_3 &= \pi_1 \\ \frac{1}{8}\pi_1 + \frac{1}{4}\pi_2 + \frac{1}{2}\pi_3 &= \pi_2 \\ \frac{3}{4}\pi_1 + \frac{1}{4}\pi_2 + 0 \cdot \pi_3 &= \pi_3 \\ \pi_1 + \pi_2 + \pi_3 &= 1. \end{aligned}$$

Rozwiązaniem jest $\pi_1 = 4/11$, $\pi_2 = 16/55$, $\pi_3 = 19/55$.

Związek rozkładu stacjonarnego z zachowaniem łańcucha Markowa podaje następujący fakt, zwany Twierdzeniem Ergodycznym.

Twierdzenie 58. *Rozważmy nieprzywiedlny, nieokresowy łańcuch Markowa na skończonej przestrzeni stanów. Wówczas łańcuch ten ma dokładnie jeden rozkład stacjonarny π . Co więcej, dla dowolnych $i, j \in E$,*

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \pi_j.$$

Uwagi:

1. Intuicyjnie, powyższe twierdzenie mówi, że stan łańcucha w bardzo odległej przyszłości jest „prawie niezależny” od stanu w chwili obecnej, gdyż prawdopodobieństwa warunkowe zbiegają do granicy niezależnej od i .

2. Jeśli Q jest rozkładem początkowym, to ze wzoru na prawdopodobieństwo całkowite mamy

$$\mathbb{P}(X_n = j) = \sum_{i \in E} Q_i p_{ij}(n) \rightarrow \pi_j$$

przy $n \rightarrow \infty$. Pozwala nam to przybliżyć rozkład zmiennej X_n . Na przykład mówi to nam, że prawdopodobieństwo spotkania pana X w domu w konkretnym dniu w odległej przyszłości wynosi w przybliżeniu $19/55$. Można je też zinterpretować inaczej: jeśli wiemy, że pan X zachowuje się w sposób opisany w przykładzie od bardzo dawna, a my nie mamy informacji na temat tego, gdzie był ostatnio, $19/55$ to dla nas przybliżenie prawdopodobieństwa, że dziś wieczorem będzie w domu.

Na zakończenie rozpatrzmy pokrótce (tylko na prostym przykładzie) jeszcze jedno zagadnienie związane z łańcuchami Markowa. Rozważmy w związku z tym

łańcuch o przestrzeni stanów $E = \{1, 2, 3, 4\}$ i macierzy przejścia

$$P = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 & 0 \\ 1/4 & 1/4 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Widzimy, że $p_{22} = p_{44} = 1$, co oznacza że jeżeli łańcuch znajdzie się w stanie 2 lub w stanie 4, to już tego stanu nie opuści. Stany tego typu nazywamy pochlaniającymi. Są one przydatne do modelowania procesów, które mogą się nagle zakończyć. Na przykład, gdy chcemy opisać grę losową, w której jeden z graczy startuje z kapitałem a , drugi b i rzucają monetą zakładając się o 1 złotówkę tak długo, aż jeden z nich nie zbankrutuje, rozpatrujemy łańcuch na przestrzeni stanów $\{0, 1, 2, \dots, a+b\}$ (gdzie stan i interpretowany jest jako chwilowy kapitał pierwszego gracza), z prawdopodobieństwami przejścia $p_{i,i+1} = p_{i,i-1} = 1/2$ dla $i \notin \{0, a+b\}$, $p_{00} = p_{a+b,a+b} = 1$. Z reguły w takich sytuacjach interesuje nas prawdopodobieństwo pochłonięcia przez konkretne stany. Można je prosto obliczyć korzystając ze wzoru na prawdopodobieństwo całkowite. Zilustrujemy to, obliczając prawdopodobieństwo pochłonięcia przez stan 2 dla naszego łańcucha. Oznaczmy przez x_i ($i = 1, 2, 3, 4$) prawdopodobieństwo, że startując ze stanu i , w pewnym momencie osiągniemy stan 2. Analizując pierwszy krok łańcucha, ze wzoru na prawdopodobieństwo całkowite łatwo wynioskować, że liczby x_i spełniają równania

$$\begin{aligned} x_1 &= \frac{1}{3}x_2 + \frac{1}{3}x_3 + \frac{1}{3}x_4, \\ x_3 &= \frac{1}{4}x_1 + \frac{1}{4}x_2 + \frac{1}{2}x_3. \end{aligned}$$

Ponadto oczywiście $x_2 = 1$ i $x_4 = 0$. Rozwiązując ten układ, dostajemy $x_1 = 3/5$ oraz $x_3 = 4/5$.