

DU | **DOSKONAŁY
UNIWERSYTET**

Rachunek prawdopodobieństwa 1, 2

Jerzy Ombach

Kraków, 2022

Spis treści

| | | |
|----------|---|-----------|
| 1 | Wstęp | 4 |
| 1.1 | Prezentacja | 4 |
| 1.2 | Ćwiczenia komputerowe | 4 |
| 1.3 | Pomocne podręczniki | 5 |
| | Rachunek prawdopodobieństwa 1 | 6 |
| 2 | Podstawowe schematy probabilistyczne | 6 |
| 2.1 | Aksjomaty przestrzeni probabilistycznej | 6 |
| 2.2 | Przykłady przestrzeni probabilistycznych | 8 |
| 2.3 | Losowania | 11 |
| 2.4 | Pytania | 12 |
| 3 | Prawdopodobieństwo warunkowe i niezależność | 13 |
| 3.1 | Prawdopodobieństwo warunkowe | 13 |
| 3.2 | Zdarzenia niezależne | 15 |
| 3.3 | Iloczyn kartezjański przestrzeni probabilistycznych | 15 |
| 3.4 | Schemat Bernoulliego | 16 |
| 3.5 | Pytania | 17 |
| 4 | Rozkład prawdopodobieństwa w \mathbb{R}^n | 18 |
| 4.1 | Rozkład jednowymiarowy. Dystrybuanta | 19 |
| 4.2 | Rozkłady dyskretne i rozkłady ciągłe | 20 |
| 4.3 | Pytania | 24 |
| 5 | Zmienne i wektory losowe | 26 |
| 5.1 | Definicje i własności | 26 |
| 5.2 | Rozkłady brzegowe i warunkowe | 29 |
| 5.3 | Niezależność zmiennych/wektorów losowych | 34 |
| 5.4 | Funkcje zmiennych/wektorów losowych | 35 |
| 5.5 | Pytania | 36 |
| 6 | Nadzieja matematyczna i wariancja | 37 |
| 6.1 | Nadzieja matematyczna – definicja i własności | 37 |
| 6.2 | Wariancja i odchylenie standardowe | 43 |
| 6.3 | Kowariancja i korelacja | 44 |
| 6.4 | Pytania | 45 |
| 7 | Nierówność Czebyszewa i prawa wielkich liczb | 46 |
| 7.1 | Nierówność Czebyszewa | 46 |
| 7.2 | Słabe prawo wielkich liczb | 47 |
| 7.3 | Pytania | 48 |

| | |
|---|------------|
| 8 Wybrane rozkłady prawdopodobieństwa | 49 |
| 8.1 Rozkład dwumianowy – $B(n, p)$ | 50 |
| 8.2 Rozkład Poissona – P_λ | 51 |
| 8.3 Rozkład hipergeometryczny | 54 |
| 8.4 Rozkład geometryczny, G_p | 54 |
| 8.5 Rozkład Pascala, ujemny rozkład dwumianowy | 55 |
| 8.6 Rozkład wykładniczy, E_λ | 56 |
| 8.7 Proces Poissona | 58 |
| 8.8 Pytania | 59 |
| 9 Rozkład normalny | 60 |
| 9.1 Standardowy rozkład normalny, $N(0, 1)$ | 61 |
| 9.2 Centralne twierdzenie graniczne | 61 |
| 9.3 Pytania | 67 |
| 10 Zbieżność zmiennych losowych | 68 |
| 10.1 Rodzaje zbieżności | 68 |
| 10.2 Mocne Prawa Wielkich Liczb | 70 |
| 10.3 Pytania | 77 |
| Rachunek prawdopodobieństwa 2 | 78 |
| 11 Zbieżność rozkładów i funkcje charakterystyczne | 78 |
| 11.1 Charakteryzacja zbieżności ciągu rozkładów | 78 |
| 11.2 Funkcje charakterystyczne | 81 |
| 11.3 Dowód centralnego twierdzenia granicznego | 85 |
| 11.4 Funkcje tworzące | 87 |
| 11.5 Pytania | 90 |
| 12 Metody Monte Carlo | 92 |
| 12.1 Liczby pseudolosowe | 92 |
| 12.2 Estymatory i przedziały ufności | 93 |
| 12.3 Całkowanie metodami Monte Carlo | 96 |
| 12.4 Optymalizacja | 98 |
| 12.5 Pytania | 100 |
| 13 Warunkowa wartość oczekiwana | 101 |
| 13.1 Wartości oczekiwana rozkładów warunkowych | 101 |
| 13.2 Twierdzenie Radona–Nikodyma | 104 |
| 13.3 Warunkowa wartość oczekiwana – sytuacja ogólna | 104 |
| 13.4 Rozkład nadziei warunkowej | 108 |
| 13.5 Pytania | 109 |
| 14 Warunkowania | 110 |
| 14.1 Przykłady | 110 |
| 14.2 Nierówność Jensena i obniżanie wariancji | 113 |
| 14.3 Pytania | 115 |
| 15 Martyngały | 117 |
| 15.1 Definicje i przykłady | 117 |
| 15.2 Wybór strategii w grze | 118 |
| 15.3 Momenty stopu | 120 |
| 15.4 Twierdzenie o zbieżności | 122 |
| 15.5 Pytania | 122 |

| | |
|--|------------|
| 16 Definicja i przykłady łańcuchów Markowa | 123 |
| 16.1 Jednorodny łańcuch Markowa | 123 |
| 16.2 Macierz przejścia i jej potęgi | 126 |
| 16.3 Pytania | 129 |
| 17 Nieredukowalne łańcuchy Markowa | 130 |
| 17.1 Powracanie | 130 |
| 17.2 Okresowość i ergodyczność | 131 |
| 17.3 Łańcuch Markowa jako graf | 134 |
| 17.4 Spacerzy losowe po grafie (nieskierowanym) | 136 |
| 17.5 Markowskie metody Monte Carlo (MCMC) | 136 |
| 17.6 Pytania | 137 |
| 18 Rozkłady wielowymiarowe | 139 |
| 18.1 Macierze symetryczne dodatnio określone | 139 |
| 18.2 Nadzieja matematyczna i macierz kowariancji | 140 |
| 18.3 Funkcje generujące momenty. | 140 |
| 18.4 Pytania | 142 |
| 19 Rozkład normalny wielowymiarowy | 143 |
| 19.1 Definicja i własności | 143 |
| 19.2 Rozkład normalny na płaszczyźnie | 144 |
| 19.3 Warunkowanie rozkładów normalnych | 145 |
| 19.4 Wnioskowanie bayesowskie. | 147 |
| 19.5 Liczby pseudolosowe z rozkładu normalnego | 148 |
| 19.6 Pytania | 149 |
| 20 Regresja | 150 |
| 20.1 Ogólny problem regresji | 150 |
| 20.2 Nadzieja warunkowa jako rzutowanie | 152 |
| 20.3 Regresja liniowa | 153 |
| 20.4 Regresja jako narzędzie statystyczne | 155 |
| 20.5 Pytania | 157 |
| 21 Odpowiedzi do pytań | 158 |
| Bibliografia | 178 |
| Indeks | 179 |

Rozdział 1

Wstęp

Prezentowany podręcznik odpowiada przedmiotom: Rachunek prawdopodobieństwa 1 (rozdziały 2 – 10) oraz Rachunek prawdopodobieństwa 2 (rozdziały 11 – 20), prowadzonym od kilku lat dla studentów II roku na specjalnościach matematyka stosowana, matematyka w ekonomii i matematyka ogólna na kierunku matematyka, studia I stopnia, na Uniwersytecie Jagiellońskim. Składa się on z dwóch różnych części; prezentacji oraz ćwiczeń komputerowych.

1.1 Prezentacja

Prezentacja w sposób dość zwięzły wprowadza podstawowe narzędzia i metody rachunku prawdopodobieństwa niezbędne do studiowania przedmiotów takich jak statystyka, procesy stochastyczne, ekonometria i matematyka finansowa. Stosunkowo dużo miejsca zajmują przykłady, które mają na celu wyrobienie u studentów intuicji probabilistycznych. Podobną rolę spełnia dość duża liczba zalecanych na bieżąco ćwiczeń.

Zakładamy znajomość podstaw teorii mnogości, analizy matematycznej funkcji jednej zmiennej, algebry liniowej z geometrią i topologii. W trakcie II roku studenci uczestniczą w równoległych zajęciach z analizy matematycznej funkcji wielu zmiennych oraz z teorii miary i całki. Dzięki temu w miarę upływu czasu będziemy w naszym kursie korzystać z powyższego faktu używając poznane na tych kursach pojęcia i metody. Także to jest powodem, że w Prezentacji zdecydowaliśmy się pominąć dowody pewnych twierdzeń stanowiących podstawę rachunku prawdopodobieństwa, gdyż są one omawiane w trakcie wymienionych kursów.

1.2 Ćwiczenia komputerowe

W trakcie kursu warto samodzielnie przeprowadzić odpowiednio dużą liczbę eksperymentów ilustrujących przerabiany materiał oraz rozwiązać szereg zagadnień wykorzystujących możliwości stwarzane obecnie przez komputer. Mogą do tego służyć przygotowane ćwiczenia komputerowe dające możliwość prowadzenia symulacji, obliczeń oraz prezentacji nawiązujących i rozszerzających program kursu.

Ćwiczenia komputerowe są formalnie niezależne od prezentacji, chociaż tematycznie odpowiadają poszczególnym rozdziałom, a niektóre przykłady z Prezentacji są rozwijane w poszczególnych punktach. Każde z ćwiczeń zawiera wprowadzenie, wskazówki i komentarze, więc osoby orientujące się w tematyce nie muszą przeglądać prezentacji przed przystąpieniem do poszczególnych ćwiczeń. Ich zaletą jest możliwość samodzielnej modyfikacji parametrów, a nawet całych procedur, co powinno inspirować użytkownika do samodzielnych eksperymentów. Ćwiczenia były przygotowywane przy użyciu Maple 18. Zdecydowana większość działa sprawnie w kilku poprzednich wersjach Maple oraz w Maple 2019, 2020 i 2021¹. Do ich przerabiania zalecany jest interface Classic, który wyraźniej niż domyślny interface Standard pozwala na odróżnieniu kodu od tekstu komentarzy, co pozwala myślącemu użytkownikowi na bardziej samodzielną pracę.

Należy podkreślić, że w sieci można znaleźć wiele innych bardzo ciekawych ćwiczeń/demonstracji wspomagających kurs rachunku prawdopodobieństwa. Zostały przygotowane przy użyciu różnorodnego oprogramowania. Także kolejne wersje Maple oferują coraz więcej gotowych aplikacji na ten temat.

¹Maple 2019 ma problem z symbolicznym obliczeniem sum nieskończonych!

1.3 Pomocne podręczniki

W zakresie naszych kursów rachunku prawdopodobieństwa niniejszy podręcznik jest w zasadzie samowystarczalny. Niemniej, warto nieraz porównać lub skonfrontować przerabiany materiał z innymi źródłami. Warto polecić:

- **Jakubowski J., Sztencel R., *Wstęp do teorii prawdopodobieństwa*, Script, Warszawa 2010.**
Jest to poważny i kompletny podręcznik o bardziej teoretycznym charakterze, który w niektórych miejscach istotnie poszerza niniejszy kurs.
- **Jakubowski J., Sztencel R., *Rachunek prawdopodobieństwa dla prawie każdego*, Script, Warszawa 2006.** Zawiera mniej treści i szczegółów niż poprzedni, ale osobom mniej zaawansowanym będzie się go dobrze czytać.
- **Ash R.B., Doleans-Dade C., *Probability and measure theory 2nd Edition*, Academic Press, New York 2000.**
Jest to dość formalny wykład rachunku prawdopodobieństwa z uwypukleniem powiązań z teorią miary, topologią i analizą funkcjonalną.
- **Ombach J., *Rachunek prawdopodobieństwa wspomagany komputerowo dla studentów matematyki stosowanej*, Wydawnictwo UJ, Kraków 2018.**
Najbardziej zbliżony do obecnego podręcznika: jest jednak inaczej zorganizowany, zawiera zestawy zadań do samodzielnego rozwiązania oraz dodatki. Jest silniej niż obecna prezentacja zależny od wyników ćwiczeń komputerowych. Nie zawiera jednak szeregu ważnych tematów (funkcje tworzące, momenty stopu, i inne), które pojawiają się obecnie. Obecny podręcznik zawiera ponad 100 pytań sprawdzających stopień zrozumienia poszczególnych rozdziałów, przy czym po każdym pytaniu Czytelnik może ewentualnie skorzystać z podpowiedzi lub nawet pełnych rozwiązań.

Istnieje wiele innych dobrych podręczników do rachunku prawdopodobieństwa. Część z nich została wskazana na stronie 178 oraz w podręczniku [20]. Są tam także wymienione pozycje, które w sposób istotny poszerzają omawiany w tym kursie materiał.

Do korzystania z Ćwiczeń komputerowych wystarczy niewielka znajomość Maple. Program ten zapewnia wszechstronny system Pomocy, co daje możliwość pracy nawet mało doświadczonym użytkownikom.

Rozdział 2

Podstawowe schematy probabilistyczne

2.1 Aksjomaty przestrzeni probabilistycznej

Zaczynamy od podania definicji przestrzeni probabilistycznej zaproponowanej prawie sto lat temu i obecnie najczęściej używanej.

Definicja – 2.1 (Kołmogorow). Niech będą dane: niepusty zbiór Ω , pewna rodzina Σ podzbiorów zbioru Ω i funkcja $P: \Sigma \rightarrow \mathbb{R}$. Trójkę (Ω, Σ, P) nazywamy przestrzenią probabilistyczną, gdy zachodzą następujące warunki:

1. $\Omega \in \Sigma$,
2. jeżeli zbiory $A_1, A_2, A_3, \dots \in \Sigma$, to $\bigcup_{i=1}^{\infty} A_i \in \Sigma$,
3. jeżeli $A, B \in \Sigma$, to $A \setminus B \in \Sigma$,
4. jeżeli $A \in \Sigma$, to $P(A) \geq 0$,
5. jeżeli zbiory $A_1, A_2, A_3, \dots \in \Sigma$ są parami rozłączne, to:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i),$$

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

Terminologia

Rodzina Σ spełniająca warunki 1 – 3 – *σ -algebra*,

Funkcja P spełniająca warunki 4, 5 – *miara*,

Funkcja P spełniająca warunki 4–6 – *miara probabilistyczna*,

$\omega \in \Omega$ – *zdarzenie elementarne*,

$A \in \Sigma$ – *zdarzenie*,

$\Omega \setminus A$ – *zdarzenie przeciwne do A* ,

$P(A)$ – *prawdopodobieństwo zdarzenia A* ,

\emptyset – *zdarzenie niemożliwe*,

Ω – *zdarzenie pewne*.

Uwaga. Nie zawsze $\{\omega\} \in \Sigma$, czyli zdarzenia elementarne nie muszą być zdarzeniami!

Przykład – 2.2. Rzut kostką. Informacja: czy wypadnie „6” (TAK/NIE).

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \quad \Sigma = \{\emptyset, \{1, 2, 3, 4, 5\}, \{6\}, \Omega\}, \quad P(A) = \frac{\#A}{6}.$$

Podstawowe własności:

1. $\emptyset \in \Sigma$,
bo $\emptyset = \Omega \setminus \Omega$.

2. Jeżeli zbiory $A_1, A_2, A_3, \dots \in \Sigma$, to $\bigcap_{i=1}^{\infty} A_i \in \Sigma$,
bo prawa de Morgana.

3. $P(\emptyset) = 0$.

$$\text{bo } P(\emptyset) = P\left(\bigcup_{i=1}^{\infty} \emptyset\right) = \sum_{i=1}^{\infty} P(\emptyset).$$

4. Jeżeli zbiory $A_1, A_2, A_3, \dots, A_n \in \Sigma$ oraz $A_i \cap A_j = \emptyset$ dla $i \neq j$, to: $P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$.

$$\text{bo } P\left(\bigcup_{i=1}^n A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right), \text{ gdzie } A_i = \emptyset \text{ dla } i > n.$$

5. Jeżeli A i B są takimi zdarzeniami, że $A \subset B$, to:

$$P(B) = P(A) + P(B \setminus A),$$

$$\text{bo } B = A \cup (B \setminus A).$$

6. Dla każdego zdarzenia A :

$$P(\Omega \setminus A) = 1 - P(A),$$

7. Jeżeli A i B są takimi zdarzeniami, że $A \subset B$, to:

$$P(A) \leq P(B),$$

8. Dla dowolnych zdarzeń A i B :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

bo $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$ – suma zbiorów rozłącznych.

Własności ciągów zdarzeń:

1. Dla dowolnych zdarzeń A_1, A_2, A_3, \dots :

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

bo $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$ – suma zbiorów rozłącznych, gdzie
 $B_1 = A_1, B_2 = A_2 \setminus A_1, B_3 = A_3 \setminus (A_1 \cup A_2), \dots$

2. Jeżeli $P(A_i) = 0, i = 1, \dots, n, n \leq \infty$, to $P\left(\bigcup_{i=1}^n A_i\right) = 0$.

3. Jeżeli $P(A_i) = 1, i = 1, \dots, n, n \leq \infty$, to $P\left(\bigcap_{i=1}^n A_i\right) = 1$, bo prawo de Morgana.

Twierdzenie – 2.3 (o ciągu zdarzeń wstępujących). Jeżeli $A_1 \subset A_2 \subset A_3 \subset \dots$, to:

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

Dowód. $P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} (A_n \setminus A_{n-1})\right) = \sum_{n=1}^{\infty} P(A_n \setminus A_{n-1}) = \lim_{N \rightarrow \infty} \sum_{n=1}^N P(A_n \setminus A_{n-1}) = \lim_{N \rightarrow \infty} P(A_N)$. Tutaj $A_0 = \emptyset$. □

Twierdzenie – 2.4 (o ciągu zdarzeń zstępujących). *Jeżeli $A_1 \supset A_2 \supset A_3 \supset \dots$, to:*

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

Dowód. Prawa de Morgana. □

2.2 Przykłady przestrzeni probabilistycznych

1. Schemat klasyczny.

Niech: $\Omega = \{\omega_1, \dots, \omega_n\}$ oraz niech Σ składa się ze wszystkich podzbiorów zbioru Ω , czyli: $\Sigma = \mathcal{P}(\Omega)$. Jeżeli $A \in \Sigma$, to przyjmijmy: $P(A) = \frac{\#A}{n}$.

Trójka (Ω, Σ, P) stanowi przestrzeń probabilistyczną.

Schemat klasyczny jest modelem wielu zjawisk w których liczba zdarzeń elementarnych jest skończona, są one jednakowo prawdopodobne i mamy pełną informację o eksperymencie.

Rzut kostką symetryczną: $\Omega = \{1, 2, 3, 4, 5, 6\}$ – wtedy prawdopodobieństwo każdego zdarzenia elementarnego wynosi $\frac{1}{6}$.

Rzut dwiema kostkami symetrycznymi: $\Omega = \{(i, j) : i, j = 1, \dots, 6\}$ – zbiór 36 par utworzonych z liczb 1, 2, 3, 4, 5, 6. Wtedy prawdopodobieństwo każdego zdarzenia elementarnego wynosi $\frac{1}{36}$.

Gdy startując w konkursie wybieramy jedno z 20 pytań, nasz zbiór Ω ma 20 elementów i prawdopodobieństwo zdarzenia elementarnego jest równe $\frac{1}{20}$.

2. Schemat klasyczny z niepełną informacją.

Przykład – 2.5. Rzut dwiema kostkami. Informacja: suma oczek.

Na przykład, przy 10 rzutach można dostać taką informację:

$$4, 7, 5, 2, 6, 11, 7, 9, 9, 6.$$

Zdarzeniem elementarnym jest para liczb, czyli $\Omega = \{(i, j) : i, j = 1, \dots, 6\} = \{1, \dots, 6\}^2$.

Zdarzeniami są zbiory F_2, \dots, F_{12} , które składają się ze zdarzeń elementarnych. Na przykład $F_4 = \{(1, 3), (2, 2), (3, 1)\}$.

Ogólnie: $F_k = \{(i, j) : i + j = k\}$, $k = 2, \dots, 12$.

Na podstawie informacji, którą mamy można jednak powiedzieć, że na przykład zaszło zdarzenie: suma oczek jest większa od 10, gdyż jest to zdarzenie $F_{11} \cup F_{12}$, lub, że suma oczek jest liczbą pierwszą: $F_2 \cup F_3 \cup F_5 \cup F_7 \cup F_{11}$. σ -algebrą Σ jest więc rodzina składająca się ze wszystkich sum zbiorów F_2, \dots, F_{12} .

Powyższy przykład można uogólnić.

Niech $\Omega = \{\omega_1, \dots, \omega_n\}$.

Niech F_1, \dots, F_k będą podzbiórmi Ω spełniającymi warunki:

$$\bigcup_{i=1}^k F_i = \Omega, \quad \text{oraz} \quad F_i \cap F_j = \emptyset \text{ dla } i \neq j.$$

Określamy:

$$\Sigma = \left\{ \bigcup_{j \in J} F_j : J \subset \{1, 2, \dots, k\} \right\}.$$

Określamy $P : \Sigma \rightarrow \mathbb{R}$ jako $P(A) = \frac{\#A}{n}$.

Oczywiście Σ jest σ -algebrą, a (Ω, Σ, P) przestrzenią probabilistyczną.

Przykład – 2.6. Doświadczenie: trzykrotny rzut monetą symetryczną. Informacja: ile wypadło orłów.

$\Omega = \{O, R\}^3$.

$F_0 = \{(R, R, R)\}$, $F_1 = \{(O, R, R), (R, O, R), (R, R, O)\}$, $F_2 = \{(O, O, R), (O, R, O), (R, O, O)\}$, $F_3 = \{(O, O, O)\}$.

3. Schemat dyskretny skończony.

Niech: $\Omega = \{\omega_1, \dots, \omega_n\}$, $\Sigma = \mathcal{P}(\Omega)$. Niech p_1, p_2, \dots, p_n będzie ciągiem liczb dodatnich (lub nieujemnych) takich, że $\sum_{i=1}^n p_i = 1$. Określamy:

$$P(A) = \sum_{i: \omega_i \in A} p_i, \text{ dla } A \in \Sigma.$$

Można łatwo sprawdzić (ćwiczenie), że (Ω, Σ, P) jest przestrzenią probabilistyczną.

Gdy $p_i = \frac{1}{n}$, to otrzymujemy schemat klasyczny.

Gdy rzucamy fałszywą kostką, to p_i nie są sobie równe. Na przykład mogą wynosić: 0.3, 0.15, 0.15, 0.15, 0.15, 0.1.

Oczywiście $P(\{\omega_i\}) = p_i$, dla każdego i .

4. Schemat dyskretny nieskończony.

Niech: $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$, $\Sigma = \mathcal{P}(\Omega)$. Niech p_1, p_2, p_3, \dots będzie ciągiem liczb dodatnich (lub nieujemnych) takich, że $\sum_{i=1}^{\infty} p_i = 1$. Określamy:

$$P(A) = \sum_{i: \omega_i \in A} p_i, \text{ dla } A \in \Sigma.$$

Można sprawdzić, że (Ω, Σ, P) jest przestrzenią probabilistyczną.

Przykład – 2.7. Powyższy schemat stosowany jest do opisu następującej sytuacji: Gracz rzuca tak długo kostką, aż uzyska „6”. Interesuje nas, w którym rzucie to nastąpi. Inaczej, warto tak określić przestrzeń probabilistyczną, żeby zdarzeniami elementarnymi były liczby 1, 2, 3, ... Zakładając, że mamy pełną informację o przebiegu eksperymentu, określamy więc:

$$\Omega = \{1, 2, 3, \dots\}, \Sigma = \mathcal{P}(\Omega).$$

Pamiętając, że $p_i = P(\{i\})$ możemy kolejno określić: p_1 – prawdopodobieństwo tego, że w pierwszym rzucie pojawi się „6” = $\frac{1}{6}$. p_2 – prawdopodobieństwo tego, że w pierwszym rzucie pojawi się coś innego niż „6”, a w drugim rzucie „6” = $\frac{5}{36}$. Stosujemy tutaj schemat klasyczny:

$$\Omega = \{1, \dots, 6\}^2, A = \{(i, j) \in \Omega : 1 \leq i \leq 5, j = 6\}, p_2 = P(A).$$

Podobnie, stosując kolejno schematy klasyczne możemy obliczyć, że:

$$p_i = \frac{5^{i-1}}{6^i}, \text{ dla } i = 3, 4, 5, \dots$$

Zauważmy, że $\sum_{i=1}^{\infty} p_i = \frac{1}{6} \frac{1}{1 - \frac{5}{6}} = 1$.

5. Przestrzeń probabilistyczna o nośniku \mathbb{R}^n .

Niech X będzie niepustym zbiorem. Niech $\mathcal{F} \subset \mathcal{P}(X)$ będzie rodziną zbiorów. Przez $\sigma(\mathcal{F})$ oznaczamy *najmniejszą σ -algebrę zawierającą rodzinę \mathcal{F}* . Zauważmy, że $\sigma(\mathcal{F})$ jest przecięciem wszystkich σ -algebr zawierających \mathcal{F} .

Jeżeli \mathcal{F} jest rodziną wszystkich zbiorów otwartych w X , to $\sigma(\mathcal{F})$ nazywamy *σ -algebrą zbiorów borelowskich* i oznaczamy $\mathcal{B}(X)$.

Gdy $X = \mathbb{R}$, $\mathcal{B}(\mathbb{R})$ może być scharakteryzowana również na inne równoważne sposoby. Można udowodnić, że: $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{G})$, gdzie \mathcal{G} jest:

rodziną wszystkich zbiorów domkniętych, lub

rodziną wszystkich przedziałów (a, b) , lub

rodziną wszystkich przedziałów $(a, b]$, lub

rodziną wszystkich przedziałów $(-\infty, b]$, itd.

Podobnie można na różne sposoby scharakteryzować $\mathcal{B}(\mathbb{R}^n)$, dla $n > 1$.

Jakkolwiek $\mathcal{B}(\mathbb{R}^n)$ zawiera większość podzbiorów zawartych w \mathbb{R}^n rozważanych w praktyce i teorii, to istnieją podzbiory \mathbb{R}^n , które nie są zbiorami borelowskimi:

$$\mathcal{P}(\mathbb{R}^n) \setminus \mathcal{B}(\mathbb{R}^n) \neq \emptyset.$$

Okazuje się, że na zbiorze $\mathcal{B}(\mathbb{R}^n)$ można określić wiele różnych miar. Najważniejszą z nich jest *miara Lebesgue’a*, która w naturalny sposób uogólnia pojęcie długości w \mathbb{R} , pola w \mathbb{R}^2 i objętości w \mathbb{R}^3 . Będziemy tę miarę oznaczać przez μ_{L_n} .

Można pokazać, że miary Lebesgue’a nie można rozszerzyć na σ -algebrę wszystkich zbiorów, $\mathcal{P}(\mathbb{R}^n)$.

Za pomocą miary Lebesgue’a można w różny sposób określać miary probabilistyczne. Najprostszy sposób to:

Prawdopodobieństwo geometryczne.

Niech $\Omega \in \mathcal{B}(\mathbb{R}^n)$ będzie takim zbiorem, że $0 < \mu_{L_n}(\Omega) < \infty$.

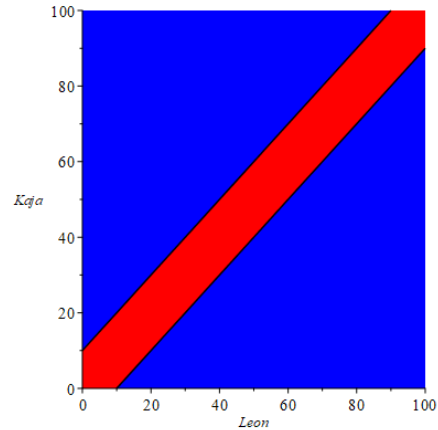
Niech $\Sigma = \{A \subset \Omega : A \in \mathcal{B}(\mathbb{R}^n)\}$.

Niech $P(A) = \frac{\mu_{L_n}(A)}{\mu_{L_n}(\Omega)}$, dla $A \in \Sigma$.

Można łatwo sprawdzić (ćwiczenie), że (Ω, Σ, P) jest przestrzenią probabilistyczną.

Przykład – 2.8. Kaja i Leon losują na chybił-trafił i niezależnie od siebie po jednej liczbie od 0 do 100. Jeżeli liczby różnią się nie więcej niż o 10, uważają, że warto pójść razem do kina. Jakie jest prawdopodobieństwo, że Kaja i Leon pójdą razem do kina?

Przestrzenią zdarzeń elementarnych Ω może być kwadrat $[0, 100] \times [0, 100]$, a miarą probabilistyczną – miara P określona jako $P(A) = \frac{\mu_{L_2}(A)}{100^2}$ dla $A \in \mathcal{B}(\Omega)$. Zdarzenie, które nas interesuje, $K = \left\{ (x, y) \in \Omega : |x - y| \leq \frac{1}{10} \right\}$ ma, jak łatwo się przekonać, prawdopodobieństwo $P(K) = 0.19$.



Wybierając model matematyczny czynimy zawsze pewne założenia, które wydają się odpowiadać rzeczywistej sytuacji. W powyższym przykładzie założyliśmy, że niezależny od siebie wybór dwóch liczb od 0 do 100 jest identyczny z wyborem punktu w kwadracie o boku mającym długość 100. Gdybyśmy jednak wiedzieli, że, na przykład, Kaja będzie wybierać chętniej małe liczby, ten model byłby bezużyteczny.

Zawsze upewnij się, co w danej sytuacji należy rozumieć przez słowa „wybór losowy”.

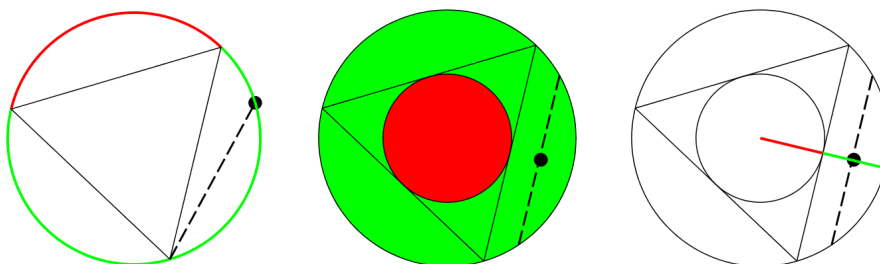
Przykład – 2.9. Dany jest okrąg o promieniu $r = 1$. Jakie jest prawdopodobieństwo tego, że losowo wybrana cięciwa tego okręgu jest krótsza niż bok trójkąta równobocznego wpisanego w ten okrąg?

Rozwiązanie:

(1) Mając ustalony jeden z końców cięciwy, wybieramy taki trójkąt równoboczny wpisany, że jego wierzchołek jest w tym punkcie. Losowanie cięciwy oznacza ustalenie drugiego jej końca. Tak więc Ω jest całym okręgiem, za Σ naturalnie jest przyjąć rodzinę zbiorów mierzalnych na okręgu, a P jest miarą Lebesgue’a (długością) podzieloną przez 2π . Widać, że prawdopodobieństwo interesującego nas zdarzenia jest równe $\frac{2}{3}$.

(2) Aby wylosować cięciwę, wystarczy podać jej środek. Tutaj Ω jest kołem, Σ jest rodziną zbiorów borelowskich zawartych w kole, a P jest miarą Lebesgue’a podzieloną przez pole koła π . Łatwo policzyć, że teraz prawdopodobieństwo naszego zdarzenia jest równe polu pierścienia podzielonemu przez $\pi = \frac{\pi - (\frac{1}{2})^2\pi}{\pi} = \frac{3}{4}$.

(3) Losujemy liczbę d od 0 do 1 i przez dowolnie wybrany punkt odległy od środka okręgu o d prowadzimy cięciwę prostopadłą do promienia okręgu wyznaczonego przez ten punkt. Teraz Ω jest odcinkiem $[0, 1]$, a prawdopodobieństwem miara Lebesgue’a. Prawdopodobieństwo, że cięciwa jest krótsza od boku trójkąta wynosi teraz $\frac{1}{2}$.



Gdzie tkwi błąd? Użyto frazę „losowo wybrana cięciwa” nie precyzując co to znaczy.

2.3 Losowania

Wiele problemów można sprowadzić do kwestii losowania. Omówimy dwa najprostsze schematy losowań.

Wyobraźmy sobie, że w urnie jest $N = 10$ ponumerowanych kul i że wyciągamy z tej urny po kolei $n = 5$ kul. Zauważmy, że mogą być tu zastosowane dwie metody losowania:

- Po wyciągnięciu kuli zapisujemy jej numer i wrzucamy ją z powrotem do urny. Zdarzeniami są na przykład **ciągi**:

$$(4, 2, 10, 5, 7), \quad (3, 3, 6, 1, 2), \quad (1, 2, 3, 4, 5) \quad \text{lub} \quad (7, 7, 7, 7, 5).$$

- Kolejne wyciągnięte kule ustawiamy obok urny. Zdarzeniami są na przykład **zbiory**:

$$\{3, 5, 6, 7, 8\}, \quad \{1, 2, 4, 8, 10\} \quad \text{lub} \quad \{2, 3, 4, 5, 6\}.$$

Zauważmy, że:

W pierwszym przypadku kule mogą się powtarzać i że musimy uwzględniać kolejność, w jakiej się pojawiają. Gdybyśmy nie uwzględnili kolejności, wynik $(1, 2, 3, 4, 5)$ odpowiadałby wielu losowaniom, a wynik $(4, 4, 4, 4, 4)$ tylko jednemu, tak więc prawdopodobieństwo drugiego wyniku musiałoby być mniejsze niż prawdopodobieństwo wyniku pierwszego, co w schemacie klasycznym nie może zachodzić – a właśnie ten schemat chcemy wykorzystać.

W drugim przypadku każdy taki zbiór jest dla nas zdarzeniem elementarnym. Zauważmy, że kule nie mogą się oczywiście powtarzać i że nie uwzględniamy kolejności – wyobraźmy sobie, że zamiast wyciągać kule po kolei bierzemy pięć kul jednocześnie (i dopiero wtedy odczytujemy numery).

Sytuacja ogólna:

Losowanie n elementów ze zwracaniem z N elementowego zbioru X

$$\Omega = X^n = \{(x_1, \dots, x_n) : x_i \in X \text{ dla } i = 1, \dots, n\}, \quad \#\Omega = N^n.$$

Losowanie n elementów bez zwracania z N elementowego zbioru X

$$\Omega = \{A \subset X : \#A = n\}, \quad \#\Omega = \binom{N}{n}.$$

Przykład – 2.10. Dany jest N elementowy zbiór X oraz N_0 elementowy podzbiór $W \subset X$. Losujemy n elementów ze zbioru X . Jakie jest prawdopodobieństwo, że wśród nich dokładnie k elementów pochodzi ze zbioru W ?

Rozwiązanie (losowanie ze zwracaniem). Zdarzeniem sprzyjającym A jest tutaj zbiór ciągów n elementowych, z których dokładnie k elementów jest ze zbioru W . Liczność takiego zbioru jest $\#A = \binom{n}{k} N_0^k (N - N_0)^{n-k}$. A więc:

$$P(A) = \frac{\binom{n}{k} N_0^k (N - N_0)^{n-k}}{N^n} = \binom{n}{k} \left(\frac{N_0}{N}\right)^k \left(1 - \frac{N_0}{N}\right)^{n-k}. \quad (2.1)$$

Rozwiązanie (losowanie bez zwracania). Zdarzeniem sprzyjającym A jest tutaj zbiór podzbiorów n elementowych, z których dokładnie k elementów jest ze zbioru W . Liczność takiego zbioru jest $\#A = \binom{N_0}{k} \binom{N - N_0}{n - k}$. A więc:

$$P(A) = \frac{\binom{N_0}{k} \binom{N - N_0}{n - k}}{\binom{N}{n}}.$$

Przykład – 2.11. Jakie jest prawdopodobieństwo tego, że na roku liczącym 75 studentów znajdziemy dwie osoby obchodzące urodziny tego samego dnia?

Tutaj zdarzeniem elementarnym jest 75-elementowy ciąg o elementach będących kolejnymi dniami roku. Przyjmując upraszczające założenia, że rok ma 365 dni i że we wszystkich dniach roku rodzi się mniej więcej tyle samo ludzi mamy do czynienia z zagadnieniem równoważnym losowaniu ze zwracaniem – 75 razy losujemy datę. Mamy więc $\#\Omega = 365^{75}$ (jest to naprawdę duża liczba). W tym zadaniu wygodniej będzie wyznaczyć zdarzenie przeciwne do interesującego nas zdarzenia, oznaczymy je przez A . Zdarzenie A składa się ze zdarzeń elementarnych odpowiadających ciągom, których wszystkie wyrazy są różne. Ponieważ mając wybrany 75-elementowy zbiór, możemy z niego utworzyć 75! różnych ciągów, a z 365 elementów możemy utworzyć $\binom{365}{75}$ różnych zbiorów, więc $\#A$ jest równa iloczynowi tych dwóch liczb. W takim razie:

$$P(A) = \frac{75! \binom{365}{75}}{365^{75}}.$$

Wyliczamy (np. Maple):

$$P(A) \approx 0.0002801218262 \quad \text{oraz} \quad 1 - P(A) \approx 0.9997198782.$$

2.4 Pytania

Pytanie 2.1. Niech $A, B \in \Sigma$. Wykazać, że jeżeli $P(A) = 1$, to $P(A \cap B) = P(B)$. ROZWIĄZANIE

Pytanie 2.2. Wyprowadzić wzór na $P(A \cup B \cup C)$. Uogólnić na dowolną sumę zbiorów. ROZWIĄZANIE

Pytanie 2.3. Gracz otrzymuje 13 kart wylosowanych z talii 52 kart i podaje liczbę otrzymanych asów oraz sumaryczną liczbę pozostałych figur. Wskazać zdarzenia, zdarzenia elementarne. Czy za pomocą prawdopodobieństw zdarzeń można obliczyć prawdopodobieństwo tego, że: (a) gracz otrzymał same piki, (b) gracz nie dostał asa?

ROZWIĄZANIE

Pytanie 2.4. Gracz rzuca trzema monetami jednocześnie tak długo, aż na wszystkich trzech pojawi się orzeł. Opisać przestrzeń probabilistyczną odpowiadającą temu doświadczeniu. ROZWIĄZANIE

Pytanie 2.5. Niech \mathcal{F} oznacza rodzinę wszystkich przedziałów (a, b) , gdzie $a < b$, $a, b \in \mathbb{R}$. Wykazać, że $\sigma(\mathcal{F}) = \mathcal{B}(\mathbb{R})$. ROZWIĄZANIE

Pytanie 2.6. Do gry w pliszki potrzebne są dwa patyki: jeden o długości co najmniej 70 cm, drugi o długości 10 do 20 cm. Aby je uzyskać gracze łamią znaleziony przez siebie patyk o długości 1 m w losowo wybranym punkcie. Jakie jest prawdopodobieństwo, że mogą grać? ROZWIĄZANIE

Rozdział 3

Prawdopodobieństwo warunkowe i niezależność

3.1 Prawdopodobieństwo warunkowe

Dana jest przestrzeń probabilistyczna (Ω, Σ, P) oraz zdarzenie $W \in \Sigma$, przy czym $P(W) > 0$.

Dla dowolnego zdarzenia $A \in \Sigma$ określamy jego prawdopodobieństwo warunkowe $P(A | W)$ wzorem:

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

Funkcja $P(\cdot | W)$ jest miarą probabilistyczną na Σ posiadającą tę właściwość, że dwa zbiory mające jednakowe przecięcia ze zbiorem W , mają także taką samą miarę (ćwiczenie).

Często znamy prawdopodobieństwo warunkowe $P(A | W)$ oraz prawdopodobieństwo $P(W)$ i na tej podstawie obliczamy prawdopodobieństwo

$$P(A \cap W) = P(A | W)P(W)$$

oraz inne prawdopodobieństwa.

Twierdzenie – 3.1 (Wzór na prawdopodobieństwo całkowite). *Dana jest przestrzeń probabilistyczna (Ω, Σ, P) oraz zdarzenia $W_1, \dots, W_n \in \Sigma$ spełniające warunki:*

- (i) $P(W_i) > 0$ dla każdego $i = 1, \dots, n$,
- (ii) $W_i \cap W_j = \emptyset$, dla wszystkich $i \neq j$,
- (iii) $W_1 \cup \dots \cup W_n = \Omega$.

Wtedy dla każdego zdarzenia $A \in \Sigma$ zachodzi wzór:

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

Dowód. Ponieważ $A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^n W_i \right) = \bigcup_{i=1}^n (A \cap W_i)$, mamy $P(A) = \sum_{i=1}^n P(A \cap W_i) = \sum_{i=1}^n P(A | W_i)P(W_i)$. \square

Powyższe zdarzenia W_1, \dots, W_n nazywamy warunkami.

Przykład – 3.2. Kaja i Leon umówili się w sprawie sprzątanania, a ponieważ Kaja sprząta dokładniej niż Leon, ustalili następujące zasady. Jeżeli w pewnym dniu sprząta Leon, to rzuca kostką i jeżeli nie wyrzuci „6”, to sprząta także w następnym dniu, gdy wypadnie „6” – sprząta Kaja. Jeżeli sprząta Kaja, to w następnym dniu nie sprząta nikt. Jeżeli w jakimś dniu nikt nie sprząta, to o sprzątananiu w następnym dniu decyduje rzut monetą. O sprzątananiu w pierwszym dniu umowy decyduje rzut monetą. Wyznaczyć prawdopodobieństwa sprzątanania Kai i Leona w drugim i trzecim dniu umowy.

Niech $K_i, L_i, N_i, i = 1, 2, 3$ oznaczają zdarzenia, że w i -tym dniu umowy sprząta Kaja, Leon oraz nikt nie sprząta. Uważamy, że są to zdarzenia w pewnej przestrzeni probabilistycznej (Ω, Σ, P) i chociaż jej formalnie nie konstruujemy (można to oczywiście zrobić), uważamy, że informacje podane w naszym zadaniu mogą być dzięki niej poprawnie zinterpretowane.

Z umowy wynika, że $P(K_1) = P(L_1) = \frac{1}{2}$, $P(N_1) = 0$. Aby obliczyć $P(K_2)$ oraz $P(L_2)$ skorzystamy dwukrotnie z wzoru na prawdopodobieństwo całkowite. Warunkami są tutaj K_1, L_1 . Widać, że spełniają one wymagane założenia.

Z umowy wynika też, że: $P(K_2 | K_1) = 0$, $P(K_2 | L_1) = \frac{1}{6}$, więc

$$P(K_2) = P(K_2 | K_1)P(K_1) + P(K_2 | L_1)P(L_1) = \frac{1}{12}.$$

$$P(L_2) = P(L_2 | K_1)P(K_1) + P(L_2 | L_1)P(L_1) = \frac{5}{12}.$$

Widzimy też, że $P(N_2) = P(N_2 | K_1)P(K_1) + P(N_2 | L_1)P(L_1) = \frac{1}{2}$, gdyż $P(N_2 | K_1) = 1$, $P(N_2 | L_1) = 0$.

Teraz warunkami są K_2, L_2, N_2 . Mamy:

$$P(K_3) = P(K_3 | K_2)P(K_2) + P(K_3 | L_2)P(L_2) + P(K_3 | N_2)P(N_2) = 0 + \frac{1}{6} \cdot \frac{5}{12} + \frac{1}{2} \cdot \frac{1}{2} = \frac{23}{72}.$$

$$P(L_3) = P(L_3 | K_2)P(K_2) + P(L_3 | L_2)P(L_2) + P(L_3 | N_2)P(N_2) = 0 + \frac{5}{6} \cdot \frac{5}{12} + \frac{1}{2} \cdot \frac{1}{2} = \frac{43}{72}.$$

Ponieważ $P(K_3) + P(L_3) + P(N_3) = 1$, to $P(N_3) = \frac{6}{72}$, co można także sprawdzić stosując jeszcze raz wzór na prawdopodobieństwo całkowite.

Jak wygląda sytuacja w kolejnych dniach zobaczymy, gdy poznamy teorię łańcuchów Markowa, przykład 17.12. Wcześniej warto zobaczyć M.3.1.

Przykład – 3.3. Kontynuując przykład poprzedni założmy, że wiemy iż w trzecim dniu umowy sprzątał Leon. Jakie jest prawdopodobieństwo tego, że sprzątał:

- (a) w drugim dniu umowy?
- (b) w pierwszym dniu umowy?

Ad (a) Chcemy obliczyć $P(L_2 | L_3)$.

$$P(L_2 | L_3) = \frac{P(L_2 \cap L_3)}{P(L_3)} = \frac{P(L_3 | L_2)P(L_2)}{P(L_3)} = \frac{\frac{5}{6} \cdot \frac{5}{12}}{\frac{43}{72}} = \frac{25}{43}.$$

Ad (b) Chcemy obliczyć $P(L_1 | L_3)$. Zauważmy, że ponieważ L_1 jest sumą rozłącznych zbiorów $L_1 \cap K_2, L_1 \cap L_2, L_1 \cap N_2$, to

$$P(L_1 | L_3) = P(L_1 \cap K_2 | L_3) + P(L_1 \cap L_2 | L_3) + P(L_1 \cap N_2 | L_3).$$

Teraz wyliczamy kolejno:

$$P(L_1 \cap K_2 | L_3) = \frac{P(L_1 \cap K_2 \cap L_3)}{P(L_3)} = \frac{P(L_3 | L_1 \cap K_2)P(L_1 \cap K_2)}{P(L_3)} = \frac{P(L_3 | K_2)P(L_1 \cap K_2)}{P(L_3)} = 0.$$

$$P(L_1 \cap L_2 | L_3) = \frac{P(L_1 \cap L_2 \cap L_3)}{P(L_3)} = \frac{P(L_3 | L_1 \cap L_2)P(L_1 \cap L_2)}{P(L_3)} = \frac{P(L_3 | L_2)P(L_2 | L_1)P(L_1)}{P(L_3)} = \frac{\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{2}}{\frac{43}{72}} = \frac{25}{43}$$

$$P(L_1 \cap N_2 | L_3) = \frac{P(L_1 \cap N_2 \cap L_3)}{P(L_3)} = \frac{P(L_3 | L_1 \cap N_2)P(L_1 \cap N_2)}{P(L_3)} = \frac{P(L_3 | N_2)P(N_2 | L_1)P(L_1)}{P(L_3)} = \frac{\frac{1}{2} \cdot 0 \cdot \frac{1}{2}}{\frac{43}{72}} = 0.$$

$$\text{Tak więc } P(L_1 | L_3) = \frac{25}{43}.$$

Poprzedni przykład stanowi ilustrację rozumowania wprowadzonego przez Bayesa. Formalizuje to następujące twierdzenie, a jego dowód jest oczywisty.

Twierdzenie – 3.4 (twierdzenie Bayesa). *Przy założeniach twierdzenia 3.1 o prawdopodobieństwie całkowitym oraz przy warunku, że $P(A) > 0$, zachodzi następująca równość:*

$$P(W_k | A) = \frac{P(A | W_k)P(W_k)}{\sum_{i=1}^n P(A | W_i)P(W_i)}$$

dla każdego $k = 1, \dots, n$.

Terminologia

$P(W_i)$ – prawdopodobieństwa a priori,

$P(W_i | A)$ – prawdopodobieństwa a posteriori.

3.2 Zdarzenia niezależne

Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną. Niech $A, B \in \sigma$.

Definicja – 3.5. A, B są *niezależne* \iff

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

Zauważmy, że gdy $P(A) > 0$ mamy natychmiastową równoważność:

$$A, B \text{ są niezależne} \iff P(B | A) = P(B).$$

Przykład – 3.6. Rzucając dwiema kostkami łatwo sprawdzić, że:

Niezależnymi zdarzeniami są A, B określone następująco: A – na pierwszej kostce wypadła „6”, B – na drugiej kostce wypadła liczba pierwsza.

Zależnymi zdarzeniami są A, B określone następująco: A – suma oczek na kostkach jest ≥ 10 , B – na drugiej kostce wypadła „5”.

Zależnymi zdarzeniami są każde dwa zdarzenia rozłączne A, B , o ile $P(A)P(B) > 0$.

Definicja – 3.7. Zdarzenia A_1, \dots, A_n są *niezależne* \iff

dla każdego podciągu A_{k_1}, \dots, A_{k_r} zachodzi:

$$P(A_{k_1} \cap \dots \cap A_{k_r}) = P(A_{k_1}) \cdot \dots \cdot P(A_{k_r}).$$

Zdarzenia A_1, A_2, A_3, \dots są *niezależne* \iff

dla każdego $n \geq 2$ zdarzenia A_1, \dots, A_n są niezależne.

3.3 Iloczyn kartezjański przestrzeni probabilistycznych

Niech będą dane dwie przestrzenie probabilistyczne $(\Omega_1, \Sigma_1, P_1)$ oraz $(\Omega_2, \Sigma_2, P_2)$. Niech

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}.$$

Można teraz zbudować σ -algebrę Σ na zbiorze Ω oraz miarę probabilistyczną $P: \Sigma \rightarrow \mathbb{R}$. Jako Σ bierze się najmniejszą σ -algebrę zawierającą wszystkie iloczyny kartezjańskie $A_1 \times A_2$, gdzie $A_1 \in \Sigma_1$ i $A_2 \in \Sigma_2$:

$$\Sigma = \sigma(\{A_1 \times A_2 : A_1 \in \Sigma_1, A_2 \in \Sigma_2\}).$$

Dowodzi się: Istnieje dokładnie jedna miara P spełniająca warunek: dla każdych dwóch zdarzeń $A_1 \in \Sigma_1$ i $A_2 \in \Sigma_2$

$$P(A_1 \times A_2) = P_1(A_1)P_2(A_2).$$

Stosujemy często następujące oznaczenia:

$$\Sigma = \Sigma_1 \times \Sigma_2 = \Sigma_1 \otimes \Sigma_2 \quad P = P_1 \times P_2 = P_1 \otimes P_2.$$

Oznaczenie \times stanowi kolizję z podobnymi oznaczeniami stosowanymi w teorii mnogości, lecz jest często stosowane.

Definicja – 3.8. Trójkę (Ω, Σ, P) skonstruowaną powyżej nazywamy iloczynem kartezjańskim przestrzeni $(\Omega_1, \Sigma_1, P_1)$ oraz $(\Omega_2, \Sigma_2, P_2)$.

Uwaga – 3.9. Niech (Ω, Σ, P) będzie iloczynem kartezjańskim przestrzeni $(\Omega_1, \Sigma_1, P_1)$ oraz $(\Omega_2, \Sigma_2, P_2)$. Niech $A_1 \in \Sigma_1, A_2 \in \Sigma_2$. Zdefiniujmy:

$$Z_1 = A_1 \times \Omega_2, \quad Z_2 = \Omega_1 \times A_2.$$

Wtedy Z_1, Z_2 są niezależne.

Dowód. $P(Z_1 \cap Z_2) = P((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)) = P(A_1 \times A_2) = P_1(A_1)P_2(A_2) = P_1(A_1)P_2(\Omega_2)P_1(\Omega_1)P_2(A_2) = P(A_1 \times \Omega_2)P(\Omega_1 \times A_2) = P(Z_1)P(Z_2)$. \square

Interpretacja. Przypuśćmy, że prowadzimy dwuetapowy eksperyment, przy czym etapy te są niezależne od siebie (np. wykonujemy dwa kolejne rzuty kostką). Załóżmy, że etapy te są opisywane dwiema przestrzeniami probabilistycznymi $(\Omega_1, \Sigma_1, P_1)$ oraz $(\Omega_2, \Sigma_2, P_2)$. Wtedy ich iloczyn kartezjański odpowiada łącznemu opisowi obydwu etapów, przy czym odpowiednikiem zdarzenia A_1 jest w nowej przestrzeni zdarzenie Z_1 , a zdarzenia A_2 zdarzenie Z_2 .

Można zdefiniować iloczyn kartezjański skończonej liczby przestrzeni probabilistycznych.

Sposób 1. Wykorzystać zwykłą procedurę indukcyjną (ćwiczenie).

Sposób 2. Powtórzyć poprzednią definicję dla ustalonej liczby przestrzeni n (ćwiczenie). Można udowodnić, że obydwie procedury dają faktycznie tę samą przestrzeń.

Uwaga – 3.10. Niech (Ω, Σ, P) będzie iloczynem kartezjańskim przestrzeni $(\Omega_1, \Sigma_1, P_1), \dots, (\Omega_n, \Sigma_n, P_n)$. Niech $A_i \in \Sigma_i, i = 1, \dots, n$. Zdefiniujmy zbiory:

$$Z_i = \{(\omega_1, \dots, \omega_n) \in \Omega : \omega_i \in A_i\}, \text{ dla } i = 1, \dots, n.$$

Wtedy Z_1, \dots, Z_n są niezależne.

Dowód. (ćwiczenie). □

Gdy $(\Omega_1, \Sigma_1, P_1) = \dots = (\Omega_n, \Sigma_n, P_n) = (\Omega, \Sigma, P)$, często oznaczmy iloczyn kartezjański tych przestrzeni symbolem: $(\Omega^n, \Sigma^n, P^n)$.

3.4 Schemat Bernoulliego

Niech $\Omega = \{0, 1\}$, $\Sigma = \mathcal{P}(\Omega)$, a miara P zadana jest przez warunki:

$$P(\{0\}) = 1 - p, \quad P(\{1\}) = p,$$

gdzie $0 < p < 1$ jest ustaloną liczbą. Taka przestrzeń, **próba Bernoulliego**, może być matematycznym modelem doświadczenia, które:

1. kończy się dokładnie dwoma wynikami,
2. znane są prawdopodobieństwa ich uzyskania,
3. prawdopodobieństwa te są takie same w każdej próbie.

Używamy często nazw „sukces” oraz „porażka” i w modelu identyfikujemy je jako 1 oraz 0.

Definicja – 3.11. Przestrzeń $(\Omega^n, \Sigma^n, P^n)$ nazywa się *schematem Bernoulliego*.

Jest to model doświadczenia składającego się z n niezależnych prób Bernoulliego.

Przykład – 3.12. Typowym zagadnieniem związanym ze schematem Bernoulliego jest następujący problem. Ustalmy $k, 0 \leq k \leq n$:

Obliczyć $P^n(A)$, gdzie $A = \{\omega = (\omega_1, \dots, \omega_n) \in \Omega^n : \sum_{i=1}^n \omega_i = k\}$.

Zauważmy, że dla każdego $\omega = (\omega_1, \dots, \omega_n) \in \Omega$

$$P^n(\{\omega\}) = P(\{\omega_1\}) \cdot \dots \cdot P(\{\omega_n\}) = p^{\sum_{i=1}^n \omega_i} (1-p)^{n-\sum_{i=1}^n \omega_i}.$$

Gdy $\omega \in A$, to $P^n(\{\omega\}) = p^k (1-p)^{n-k}$. Ponieważ zdarzenie A składa się z $\binom{n}{k}$ elementów, to

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

Warto ten wzór porównać z wzorem na prawdopodobieństwo wylosowania dokładnie k elementów w tracie losowania n elementów ze zwracaniem, wzór (2.1).

Przykład – 3.13. Salę oświetla 150 żarówek: po 15 żarówek w 10 rzędach. Wiadomo, że prawdopodobieństwo zepsucia się pojedynczej żarówki w ciągu nachodzącego tygodnia wynosi $p = 0.06$. Wiemy też, że żarówki psują się niezależnie od siebie. Jakie jest prawdopodobieństwo tego, że po upływie tygodnia w co najmniej 9 rzędach będzie świecić co najmniej po 13 żarówek?

A – w ustalonym rzędzie świeci co najmniej 13 żarówek.

$$p_{rz} = P(A) = \sum_{i=0}^2 \binom{15}{i} p^i (1-p)^{15-i}.$$

Odpowiedź na pytanie: $p_s = \sum_{i=9}^{10} \binom{10}{i} p_{rz}^i (1 - p_{rz})^{10-i}$.

Odpowiedź numeryczna: $p_{rz} = 0.9428666763$, $p_s = 0.8917349692$.

Gdy $p = 0.05$, to $p_{rz} = 0.9637997614$, $p_s = 0.9513932113$.

3.5 Pytania

Pytanie 3.1. Sformułować i wykazać twierdzenie mówiące o tym, że $P(\cdot | W)$ określa przestrzeń probabilistyczną.

ROZWIĄZANIE

Pytanie 3.2. Czy jest prawdą, że $P(A | W_1 \cup W_2) = P(A | W_1) + P(A | W_2)$? ROZWIĄZANIE

Pytanie 3.3. Jak można osłabić założenia w twierdzeniu o prawdopodobieństwie całkowitym? ROZWIĄZANIE

Pytanie 3.4. W przykładzie 3.2 obliczyć $P(K_1 | K_3)$. ROZWIĄZANIE

Pytanie 3.5. Rzucamy kostką do gry. Wskazać dwa nietrywialne zdarzenia niezależne w przestrzeni probabilistycznej odpowiadającej temu eksperymentowi. ROZWIĄZANIE

Pytanie 3.6. Uogólnić sytuację opisaną jako schemat Bernoulliego przy założeniu, że każde doświadczenie może mieć trzy różne wyniki ROZWIĄZANIE

Rozdział 4

Rozkład prawdopodobieństwa w \mathbb{R}^n

Definicja – 4.1 (Rozkład prawdopodobieństwa). n -wymiarowym rozkładem prawdopodobieństwa nazywamy miarę probabilistyczną Q określoną na σ -algebrze zbiorów borelowskich $\mathcal{B}(\mathbb{R}^n)$.

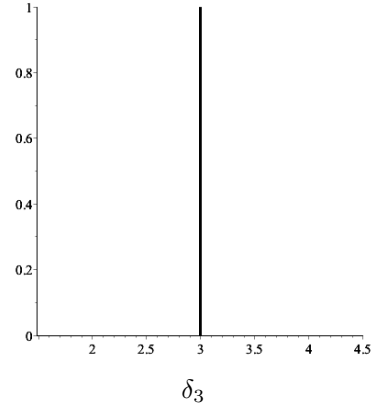
Mówiąc rozkład, najczęściej mamy na myśli rozkład jednowymiarowy.

Przykłady.

(1) δ -Diraca, δ_c . Dla ustalonego $c \in \mathbb{R}^n$ definiujemy:

$$\delta_c(A) = \begin{cases} 1, & \text{gdy } c \in A, \\ 0, & \text{gdy } c \notin A, \end{cases}$$

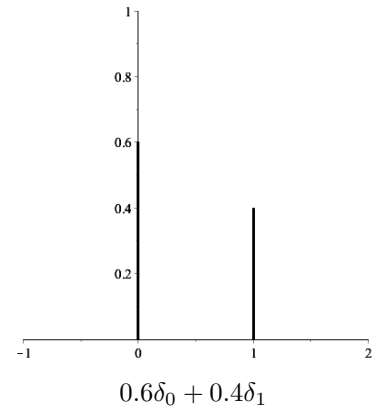
dla $A \in \mathcal{B}(\mathbb{R}^n)$.



(2) Jeżeli Q_1, \dots, Q_k są rozkładami prawdopodobieństwa, to funkcja $Q : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}$:

$$Q(A) = \sum_{i=1}^k \lambda_i Q_i(A),$$

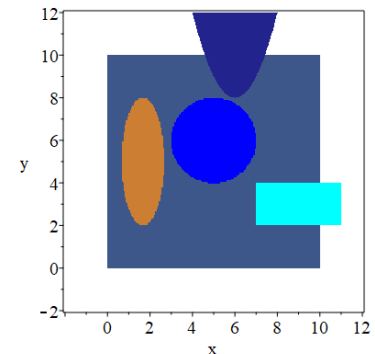
gdzie $\sum_{i=1}^k \lambda_i = 1$ oraz $\lambda_i > 0$, jest miarą probabilistyczną na $\mathcal{B}(\mathbb{R}^n)$, czyli jest n -wymiarowym rozkładem.



(3) Prawdopodobieństwo geometryczne (rozkład jednostajny): Niech $W \in \mathcal{B}(\mathbb{R}^n)$ będzie takim zbiorem, że $0 < \mu_{L_n}(W) < \infty$. Definiujemy

$$Q(A) = \frac{\mu_{L_n}(A \cap W)}{\mu_{L_n}(W)},$$

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



$$W = [0, 10]^2, \quad Q(A) = \text{pole}(A \cap W)/100$$

4.1 Rozkład jednowymiarowy. Dystrybuanta

Rozkłady (jednowymiarowe) mają ścisły związek z dystrybuantami.

Definicja – 4.2. *Dystrybuantą* nazywamy funkcję $F: \mathbb{R} \rightarrow \mathbb{R}$, spełniającą następujące cztery warunki:

1. Dla każdego $x \in \mathbb{R}$, $0 \leq F(x) \leq 1$.

2. F jest funkcją niemalejącą, to znaczy:

$$x < y \Rightarrow F(x) \leq F(y),$$

3. F jest prawostronnie ciągła, to znaczy:

$$\lim_{x \rightarrow a^+} F(x) = F(a)$$

dla każdego $a \in \mathbb{R}$,

4. $\lim_{x \rightarrow \infty} F(x) = 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$.

Twierdzenie – 4.3. *Jeżeli Q jest rozkładem prawdopodobieństwa, to funkcja F zdefiniowana wzorem:*

$$F(x) = Q(-\infty, x] = Q((-\infty, x]), \quad (4.1)$$

jest dystrybuantą. Mówimy wtedy, że rozkład Q ma dystrybuantę F , co często zaznaczamy, pisząc F_Q , zamiast F .

Dowód. Ad 1. Wynika natychmiast z własności prawdopodobieństwa.

Ad 2. Jeżeli $x < y$, to $(-\infty, x] \subset (-\infty, y]$, a więc $F(x) = Q(-\infty, x] \leq Q(-\infty, y] = F(y)$.

Ad 3. Niech $a \in \mathbb{R}$ oraz $x_n \searrow a$, to znaczy $\forall n \ x_{n+1} < x_n$, $\lim_{n \rightarrow \infty} x_n = a$.

$(-\infty, a] = \bigcap_{n=1}^{\infty} (-\infty, x_n]$ oraz $(-\infty, x_{n+1}) \subset (-\infty, x_n)$. $F(a) = Q(-\infty, a] = \lim_{n \rightarrow \infty} Q(-\infty, x_n] = \lim_{n \rightarrow \infty} F(x_n)$.

To oznacza, że $\lim_{x \rightarrow a^+} F(x) = F(a)$.

Ad 4. Podobnie jak wyżej. Wynika z faktu, że $\bigcup_{n=1}^{\infty} (-\infty, n] = \mathbb{R}$ oraz $\bigcap_{n=1}^{\infty} (-\infty, -n] = \emptyset$ (ćwiczenie). □

Zachodzi także twierdzenie odwrotne (dowód pomijamy).

Twierdzenie – 4.4. *Jeżeli F jest dystrybuantą, to istnieje dokładnie jeden rozkład Q , dla którego zachodzi wzór*

$$F(x) = Q(-\infty, x] = Q((-\infty, x]). \quad (4.2)$$

Tak więc istnieje wzajemnie jednoznaczna odpowiedniość (bijekcja) pomiędzy zbiorem rozkładów i zbiorem dystrybuant. Gdy rozkładowi Q odpowiada dystrybuanta F , piszemy często F_Q oraz Q_F . Zachodzą więc związki.

$$F_Q(x) = Q(-\infty, x], \quad F(x) = Q_F(-\infty, x] \quad \text{dla każdego } x \in \mathbb{R}. \quad (4.3)$$

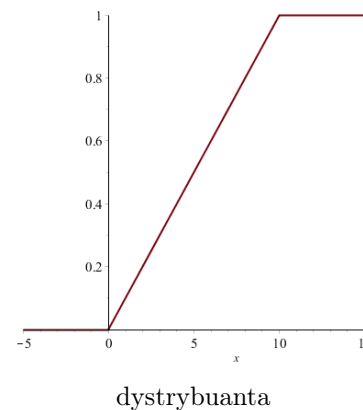
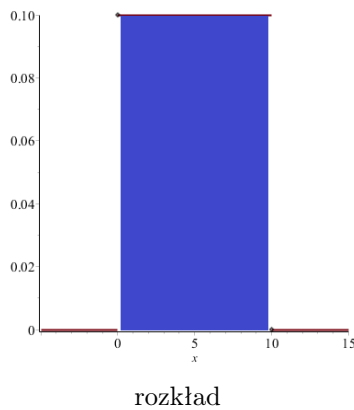
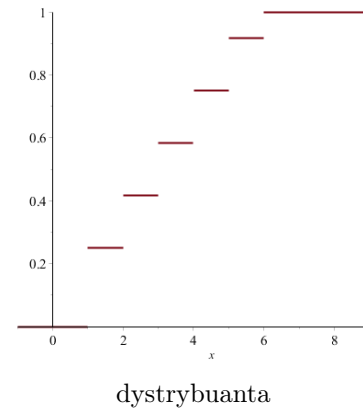
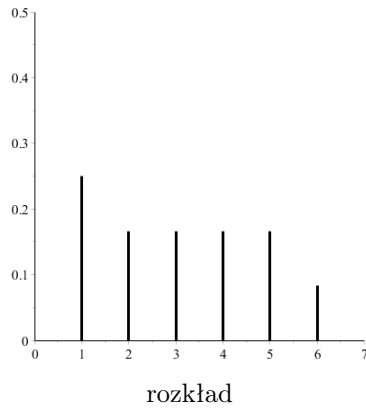
Ponieważ dla każdych $a, b \in \mathbb{R}$, $a < b$ zachodzi

$$Q(a, b] = Q((-\infty, b] \setminus (-\infty, a]) = Q(-\infty, b] - Q(-\infty, a],$$

to otrzymujemy następujący związek:

$$Q(a, b] = F_Q(b) - F_Q(a). \quad (4.4)$$

Przykłady dwóch rozkładów i ich dystrybuant:



Pytanie. W których punktach dystrybuanta jest ciągła?

Twierdzenie – 4.5. Niech Q będzie rozkładem prawdopodobieństwa, zaś F – jego dystrybuantą. Wówczas dla dowolnego $a \in \mathbb{R}$:

$$F \text{ jest ciągła w punkcie } a \iff Q(a) = 0.$$

Bardziej ogólnie:

$$Q(a) = F(a) - F(a)^-,$$

gdzie $F(a)^-$ oznacza lewostronną granicę funkcji F w punkcie a (ponieważ F jest niemalejąca, więc granica ta istnieje).

Dowód. Weźmy ciąg $x_n \nearrow a$ (to znaczy, że $\{x_n\}$ jest ciągiem rosnącym, zbieżnym do a). Wtedy $(-\infty, a) = \bigcup_n (-\infty, x_n]$, a więc:

$$F(a)^- = \lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} Q(-\infty, x_n] = Q(-\infty, a).$$

Stąd:

$$Q(a) = Q((-\infty, a] \setminus (-\infty, a)) = Q(-\infty, a] - Q(-\infty, a) = F(a) - F(a)^-. \quad \square$$

Dla $n > 1$ można też zdefiniować dystrybuantę i podać jej związek z rozkładem, jednak definicja nie jest automatycznym powtórzeniem sytuacji jednowymiarowej, gdyż w \mathbb{R}^n nie ma naturalnego porządku.

4.2 Rozkłady dyskretne i rozkłady ciągłe

Najczęściej rozważamy rozkłady dyskretne oraz rozkłady ciągłe (istnieją też inne rozkłady).

Definicja – 4.6 (Rozkład dyskretny). Rozkład n -wymiarowy Q nazywamy *rozkładem dyskretnym*, jeżeli istnieje zbiór borelowski $K \subset \mathbb{R}^n$ taki, że:

$$Q(K) = 1 \quad \text{oraz} \quad x \in K \Rightarrow Q(x) > 0.$$

Uwaga – 4.7. Występujący w powyższej definicji zbiór K jest skończony lub przeliczalny.

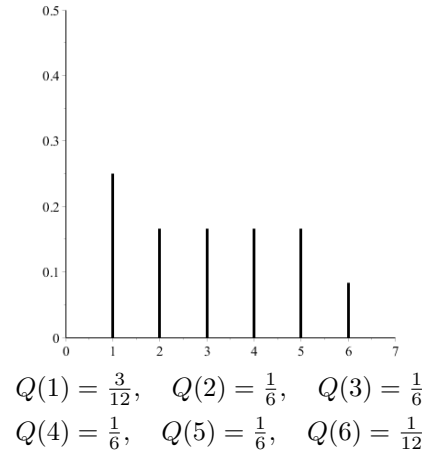
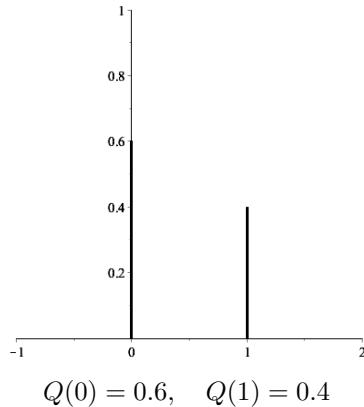
Dowód. Zauważmy, że K można przedstawić jako przeliczalną sumę zbiorów skończonych.

Dokładniej: $K = \bigcup_{i=1}^{\infty} K_i$, gdzie $K_i = \{x \in \mathbb{R}^n : Q(x) \geq \frac{1}{i}\}$.

Widzimy, że $1 \geq Q(K_i) \geq \#K_i \cdot \frac{1}{i}$, a więc $\#K_i \leq i$. □

Z powyższej uwagi wynika, iż możemy zbiór K ustawić w ciąg, powiedzmy $K = \{x_i : i = 1, \dots, m\}$, gdzie m jest liczbą naturalną lub $m = \infty$, i oznaczyć $p_i = Q(x_i)$. Mamy wtedy:

$$\sum_{i=1}^m p_i = 1 \quad \text{oraz} \quad p_i > 0 \text{ dla wszystkich } i.$$



Zdefiniowane w ten sposób ciągi $\{x_i\}$ i $\{p_i\}$ wyznaczają jednoznacznie rozkład Q . Mianowicie, dla każdego zbioru borelowskiego A mamy $Q(A) = Q(A \cap K)$ (dlaczego?) i dalej:

$$Q(A) = \sum_{i: x_i \in A} p_i. \quad (4.5)$$

W związku z powyższym, często używa się sformułowania: rozkład dyskretny zadany przez ciągi $\{x_i\}$ i $\{p_i\}$.

Dystrybuanta rozkładu dyskretnego. Niech rozkład dyskretny Q będzie zadany przez ciągi $\{x_i\}$ oraz $\{p_i\}$. Wtedy otrzymujemy:

$$F_Q(x) = \sum_{i: x_i \leq x} p_i.$$

Trywialnym przykładem rozkładu dyskretnego jest rozkład jednopunktowy δ_c . Nietrywialnymi przykładami są kombinacje barycentryczne takich rozkładów.

Rozkład $(0, 1, p)$. Jest to rozkład skupiony w dwóch punktach 0 oraz 1 mający parametr $0 < p < 1$. Mianowicie:

$$Q(0) = 1 - p, \quad Q(1) = p.$$

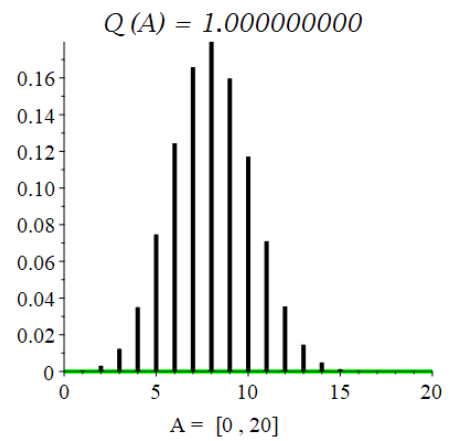
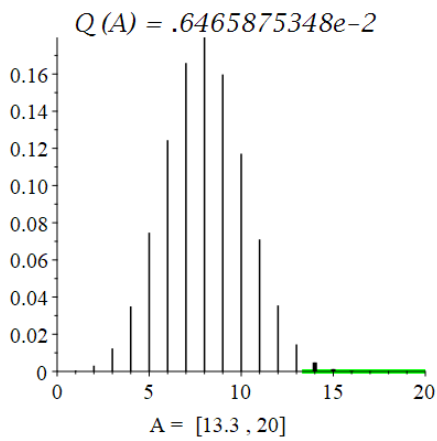
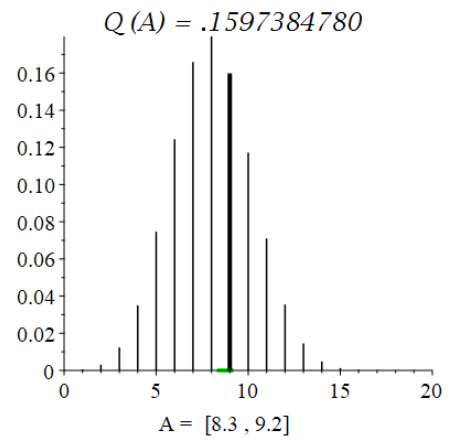
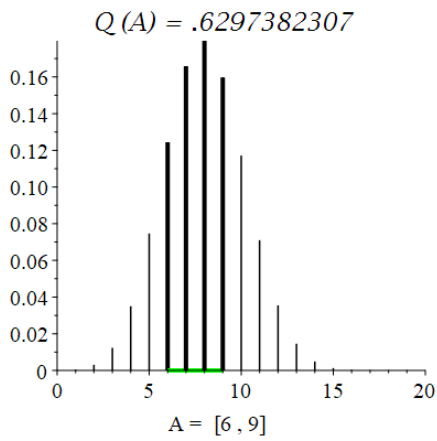
Jest często używany, jako model doświadczenia, które może dać dokładnie dwa wyniki nazywane często sukcesem – 1 i porażką – 0.

Rozkład dwumianowy, $B(n, p)$. Jest to rozkład skupiony w punktach $0, 1, \dots, n$, przy czym:

$$Q(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{dla } k = 0, 1, \dots, n.$$

Pamiętamy, że powyższy wzór określa prawdopodobieństwo dokładnie k sukcesów w n niezależnych próbach Bernoulliego.

Przykład – 4.8. Możemy zinterpretować graficznie wzór (4.5) dla rozkładu $B(20, 0.4)$.



Definicja – 4.9 (Rozkład ciągły). Rozkład n -wymiarowy Q nazywamy *rozkładem ciągłym*, jeżeli istnieje funkcja całkowalna $f: \mathbb{R}^n \rightarrow \mathbb{R}$ taka, że dla każdego zbioru borelowskiego $A \subset \mathbb{R}^n$:

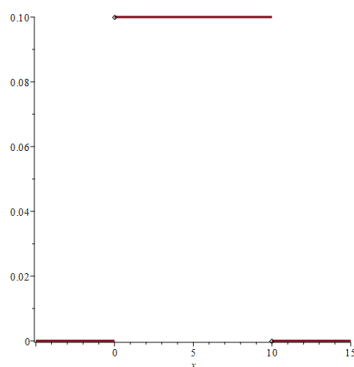
$$Q(A) = \int_A f(x) dx = \mu_{L_{n+1}}(A_f) \quad (4.6)$$

gdzie $\int_A f(x) dx$ oznacza całkę względem miary Lebesgue'a po zbiorze A z funkcji f .

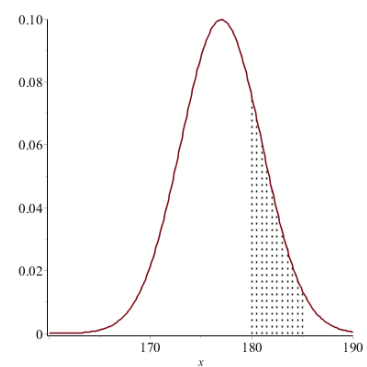
$$A_f = \{(x, y) \in \mathbb{R}^{n+1} : x \in A, 0 \leq y \leq f(x)\}.$$

Funkcję f nazywamy wówczas *gęstością rozkładu* Q .

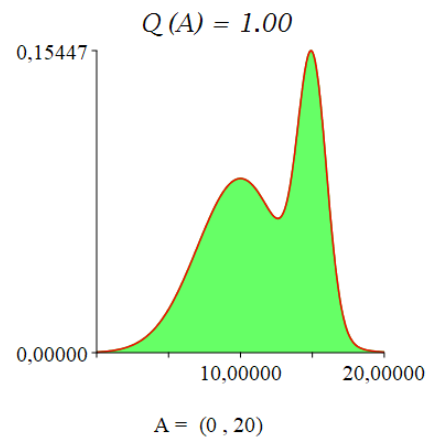
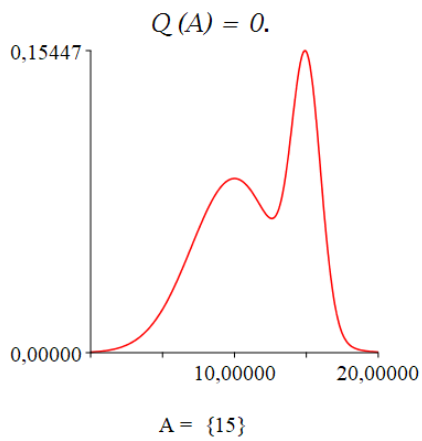
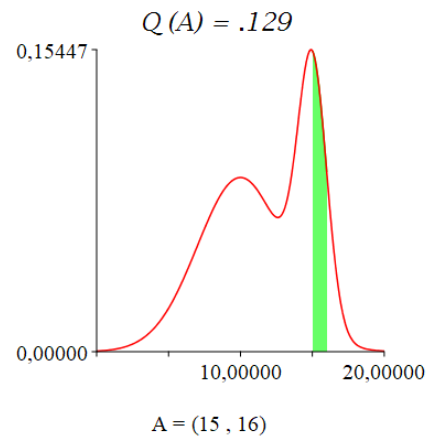
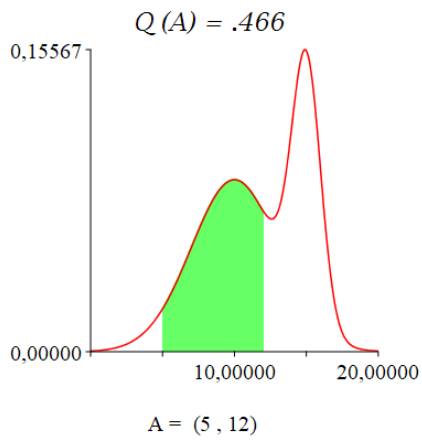
Przykład – 4.10. Przykłady gęstości rozkładów ciągłych wraz z interpretacją wzoru (4.6):



$$A = (2, 4), \quad Q(A) = \frac{2}{10}$$



$$A = (180, 185), \quad Q(A) - \text{zakreskowane pole}$$



Widać, że (bierzemy $A = \mathbb{R}^n$):

$$\int_{\mathbb{R}^n} f(x) dx = 1$$

oraz

$$f(x) \geq 0 \text{ prawie wszędzie,}$$

co rozumiemy następująco:

$\mu_{L_n}(\{x \in \mathbb{R}^n : f(x) < 0\}) = 0$, gdzie μ_{L_n} oznacza miarę Lebesgue'a.

Gdyby $\mu_{L_n}(\{x \in \mathbb{R}^n : f(x) < 0\}) > 0$, to $Q(\{x \in \mathbb{R}^n : f(x) < 0\}) = \int_{\{x \in \mathbb{R}^n : f(x) < 0\}} f(x) dx < 0$.

Uwaga – 4.11. Jeżeli funkcja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ spełnia dwa powyższe warunki, to jest ona gęstością pewnego rozkładu Q .

Wystarczy wziąć: $Q(A) = \int_A f(x) dx$, dla $A \in \mathbf{B}(\mathbb{R}^n)$.

Dystrybuanta rozkładu ciągłego. Niech rozkład ciągły Q ma gęstość f . Wtedy wprost z definicji otrzymujemy:

$$F_Q(x) = Q(-\infty, x] = \int_{-\infty}^x f(t) dt.$$

Zauważmy, że w rozkładzie ciągłym zbiory jednopunktowe mają miarę zero, a więc dystrybuanta jest ciągła w każdym punkcie.

Można podać przykład dystrybuanty, która jest funkcją ciągłą w każdym punkcie, ale jej rozkład **nie jest** ciągły.

Jeżeli pewna funkcja f mierzalna spełnia powyższy wzór, to jest ona gęstością rozkładu, którego dystrybuantą jest F . Jeżeli więc wiemy, że dystrybuanta jest funkcją ciągłą i różniczkowalną, ewentualnie poza skończoną liczbą punktów, to jej pochodna jest gęstością rozważanego rozkładu. Wiadomo ponadto, że w każdym punkcie x , który jest punktem ciągłości f , funkcja górnej granicy całkowania, a więc dystrybuanta, jest różniczkowalna oraz zachodzi wzór:

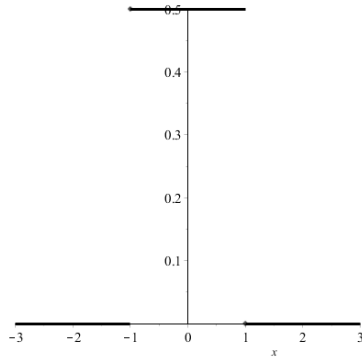
$$F'(x) = f(x).$$

Przykładem rozkładu ciągłego jest:

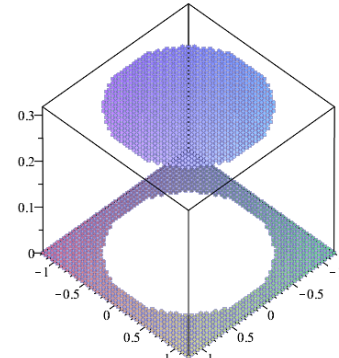
Rozkład jednostajny, $U(G)$. Niech $G \subset \mathbb{R}^n$ będzie zbiorem borelowskim o dodatniej i skończonej mierze Lebesgue'a, to znaczy $0 < \mu_{L^n}(G) < \infty$. Określmy funkcję:

$$f(x) = \begin{cases} 0, & \text{gdy } x \notin G \\ \frac{1}{\mu_{L^n}(G)}, & \text{gdy } x \in G. \end{cases}$$

Jest oczywiste, że f spełnia warunki wymagane od gęstości, jest więc gęstością pewnego rozkładu prawdopodobieństwa. Rozkład ten nazywamy rozkładem jednostajnym.



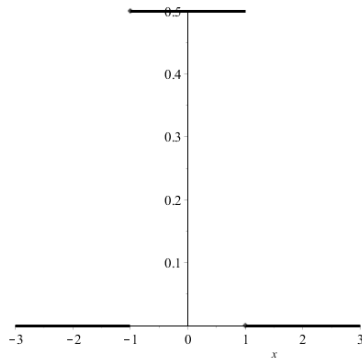
$$G = (-1, 1) \subset \mathbb{R}$$



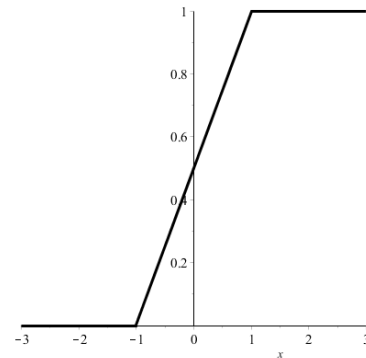
$$G = \{(x, y) : x^2 + y^2 < 1\} \subset \mathbb{R}^2$$

Przykład – 4.12. Niech F będzie dystrybuantą rozkładu jednostajnego na odcinku (a, b) , $U(a, b)$. Otrzymujemy:

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & b \leq x. \end{cases}$$



$$f(x) = \frac{1}{2} \mathbb{1}_{[-1,1]}$$



$$F(x) = \frac{x+1}{2} \text{ dla } -1 < x < 1$$

Zauważmy, że F nie jest różniczkowalna w punktach a, b i w tych samych punktach f nie jest ciągła. Nie jest też istotne, ile wynosi $f(a)$ oraz $f(b)$.

4.3 Pytania

Pytanie 4.1. Niech F_1, \dots, F_k będą dystrybuantami, $\lambda_1, \dots, \lambda_k > 0$, $\lambda_1 + \dots + \lambda_k = 1$. Czy $F = \lambda_1 F_1 + \dots + \lambda_k F_k$ jest dystrybuantą? ROZWIĄZANIE

Pytanie 4.2. Q jest rozkładem dyskretnym zadany przez ciągi: $\{x_i\} = (1, 2, 3, \dots)$, $\{p_i\} = (\frac{1}{6}, \frac{5}{6} \cdot \frac{1}{6}, (\frac{5}{6})^2 \cdot \frac{1}{6}, \dots)$, patrz wzór (4.5). Obliczyć $Q(A)$, gdzie A jest zbiorem liczb parzystych.

Wykazać, że zbiory liczb parzystych i liczb pierwszych są zdarzeniami zależnymi. ROZWIĄZANIE

Pytanie 4.3. Wskazać rozkład ciągły, którego gęstość jest dodatnia we wszystkich punktach $x \in \mathbb{R}$. ROZWIĄZANIE

Pytanie 4.4. Wykazać, że dla dowolnej dystrybuanty zbiór punktów nieciągłości jest co najwyżej przeliczalny.

ROZWIĄZANIE

Pytanie 4.5. Niech F będzie dystrybuantą, a $h : \mathbb{R} \rightarrow \mathbb{R}$ funkcją. Zaproponować założenia o h gwarantujące, że $F \circ h$ jest dystrybuantą. ROZWIĄZANIE

Pytanie 4.6. Wskazać dystrybuantę rozkładu jednostajnego na:

(a) przedziale $(-2, 2)$, (b) sumie przedziałów $(-4, -2)$, $(2, 4)$. ROZWIĄZANIE

Rozdział 5

Zmienne i wektory losowe

Rozważając matematyczny model zjawiska o charakterze losowym nie zawsze potrafimy (lub chcemy) opisywać odpowiadającą mu przestrzeń probabilistyczną. Operujemy jednak wielkościami, które są interpretowane jako wartości odpowiednich funkcji określonych na tej przestrzeni. O ile spełniają odpowiednie warunki, funkcje te nazywane są zmiennymi lub wektorami losowymi.

5.1 Definicje i własności

Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną.

Definicja – 5.1 (wektor losowy). Funkcję $X: \Omega \rightarrow \mathbb{R}^n$ nazywamy *wektorem losowym*, jeżeli jest ona funkcją mierzalną względem σ -algebry Σ , to znaczy:

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \Sigma$$

dla każdego zbioru borelowskiego $B \in \mathcal{B}(\mathbb{R}^n)$.

Zmienna losowa jest to jednowymiarowy wektor losowy.

Wyróżnianie zmiennych losowych nie ma formalnego uzasadnienia, stosuje się je ze względów tradycyjnych. Dość często określenie „wektor losowy” i „zmienna losowa” są używane zamiennie.

Zbiory $X^{-1}(B)$, gdzie $B \in \mathcal{B}(\mathbb{R}^n)$, będziemy nazywać *zbiorami opisywanymi przez wektor losowy X* . Podkreślamy wyraźnie, że są to zbiory postaci $\{\omega \in \Omega : X(\omega) \in B\}$, co skrótowo będziemy zapisywać $\{X \in B\}$. Tak więc, na przykład, wyrażenie $P(X < \varepsilon)$ oznacza: $P(\{\omega \in \Omega : X(\omega) < \varepsilon\})$.

Warunek mierzalności oznacza, że wszystkie zdarzenia opisane przez X są elementami Σ , czyli, że mamy dostępną informację na temat takich zdarzeń.

Przykład – 5.2. Niech $\Omega = \{(i, j) : i, j = 1, \dots, 6\}$. Niech $\Sigma = \sigma(F_1, \dots, F_6)$, gdzie $F_k = \{(i, j) : \max(i, j) = k\}$.

Niech $X: \Omega \rightarrow \mathbb{R}$ będzie funkcją określoną jako $X(i, j) = i$. X nie jest zmienną losową, gdyż na przykład zbioru

$$\{(i, j); X(i, j) = 1\} = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)\}$$

nie można przedstawić jako sumy pewnych zbiorów F_i . Inaczej mówiąc. Nie potrafimy za pomocą dostępnej informacji (znajomości Σ) zinterpretować wyniku zaobserwowanego na pierwszej kostce.

Niech $Y: \Omega \rightarrow \mathbb{R}$ będzie funkcją określoną jako; $Y(i, j) = 1$, gdy $i = 6$ oraz $j = 6$, oraz $Y(i, j) = 0$ w przeciwnym przypadku. Y nie jest zmienną losową, gdyż $Y^{-1}(1) = \{(6, 6)\}$ nie można przedstawić jako sumy pewnych zbiorów F_i .

Niech $Z: \Omega \rightarrow \mathbb{R}$ będzie funkcją określoną jako; $Z(i, j) = 1$, gdy $i = 6$ lub $j = 6$, oraz $Z(i, j) = 0$ w przeciwnym przypadku. Z jest zmienną losową, gdyż $Z^{-1}(1) = F_6$, $Z^{-1}(0) = F_1 \cup \dots \cup F_5$. Dla każdego zbioru borelowskiego B mamy $Z^{-1}(B) = Z^{-1}(B \cap \{0, 1\}) = Z^{-1}(B \cap \{0\}) \cup Z^{-1}(B \cap \{1\}) \in \Sigma$.

Przykład – 5.3. Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną, $A \subset \Omega$. Określamy *funkcję charakterystyczną zbioru A*

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

Wtedy: $A \in \Sigma \iff I_A$ jest zmienną losową.

Uwaga – 5.4. Gdy $\Sigma = \mathcal{P}(\Omega)$, to każda funkcja $X: \Omega \rightarrow \mathbb{R}^n$ jest zmienną losową.

Odwzorowania mierzalne, a więc także wektory losowe, mają szereg pożytecznych własności dotyczących działań algebraicznych, złożeń, kresów, zbieżności. Znane są one z innych kursów i będziemy z nich wielokrotnie korzystać w dalszej części wykładu. Poniżej podajemy ważniejsze z nich w języku wektorów i zmiennych losowych.

Twierdzenie – 5.5. *Zachodzą następujące własności:*

1. Suma, różnica, iloczyn, iloraz (o ile jest wykonalny) wektorów losowych jest wektorem losowym.
2. Zestawienie wektorów losowych jest wektorem losowym.
3. Minimum i maksimum zmiennych losowych jest zmienną losową.
4. Kres, dolny i górny, ciągu zmiennych losowych jest zmienną losową.
5. Granica ciągu wektorów losowych jest wektorem losowym.
6. Złożenie $g \circ X$ wektora losowego X z funkcją borelowską g jest wektorem losowym. (Funkcja g jest borelowska $\iff g^{-1}(B)$ jest zbiorem borelowskim, gdy B jest zbiorem borelowskim).

W szczególności zmiennymi losowymi są:

7. Funkcje proste (funkcje schodkowe) zdefiniowane jako:

$$X(x) = c_i, \text{ dla } x \in A_i,$$

przy czym $A_1, \dots, A_k \in \Sigma$ są zbiorami parami rozłącznymi i dającymi w sumie całą przestrzeń Ω , a liczby c_1, \dots, c_k są dowolne. Mianowicie X jest sumą:

$$X(x) = \sum_{i=1}^k c_i I_{A_i}.$$

8. Funkcje X^+ oraz X^- zdefiniowane dla zmiennej losowej X jako:

$$X^+(x) = \max(X(x), 0), \quad X^-(x) = -\min(X(x), 0).$$

9. Niech $X, Y : \Omega \rightarrow \mathbb{R}$. Wtedy:

X, Y są zmiennymi losowymi $\iff (X, Y)$ jest wektorem losowym.

Definicja – 5.6. *Rozkład wektora losowego $X : \Omega \rightarrow \mathbb{R}^n$ jest to rozkład określony wzorem:*

$$P_X(B) = P(X^{-1}(B)), \text{ dla } B \in \mathcal{B}(\mathbb{R}^n).$$

Mierzalność X gwarantuje sensowność tej definicji – ponieważ P jest określone na zdarzeniach z Σ , musimy mieć gwarancję, że $X^{-1}(B) \in \Sigma$.

Łatwo sprawdzić, że powyższy wzór określa rzeczywiście rozkład.

Często piszemy $X \sim Q$, gdy $P_X = Q$.

Dystrybuantę rozkładu P_X będziemy nazywać dystrybuantą zmiennej losowej X i oznaczać często przez F_X .

Zazwyczaj używając pojęcia zmienna losowa (wektor losowy) oraz operując jej (jego) rozkładem nie zwracamy zbytnej uwagi na przestrzeń probabilistyczną na której jest ten obiekt określony. W wielu przypadkach postępowanie takie jest usprawiedliwione dzięki następującemu twierdzeniu:

Twierdzenie – 5.7. *Niech Q będzie n -wymiarowym rozkładem prawdopodobieństwa. Wtedy istnieje przestrzeń probabilistyczna (Ω, Σ, P) oraz wektor losowy $X : \Omega \rightarrow \mathbb{R}^n$, taki, że $Q = P_X$.*

Dowód. Wystarczy wziąć: $\Omega = \mathbb{R}^n$, $\Sigma = \mathcal{B}(\mathbb{R}^n)$, $P = Q$ oraz $X = id_{\mathbb{R}^n}$. □

Przykład – 5.8. W urnie są dwie białe i trzy czarne kule. Losujemy pojedynczo kule do momentu wyciągnięcia białej kuli. Niech X oznacza liczbę losowań. Wyznamy rozkład X .

(a) losowanie ze zwracaniem. Widać, że X przyjmuje wartości $1, 2, 3, \dots$. Łatwo też sprawdzić, że $p_k = P(X = k) = \left(\frac{3}{5}\right)^{k-1} \cdot \frac{2}{5}$, $k = 1, 2, 3, \dots$. Tak więc P_X ma rozkład dyskretny zadany jest ciąg $\{k\}, \left\{\left(\frac{3}{5}\right)^{k-1} \cdot \frac{2}{5}\right\}$, $k = 1, 2, 3, \dots$.

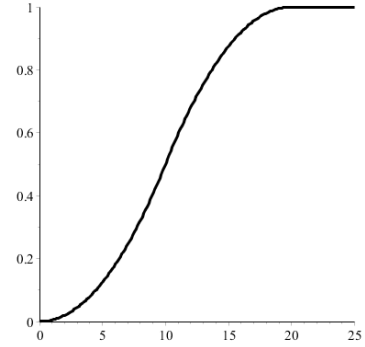
(b) losowanie bez zwracania. Widać, że X przyjmuje wartości 1, 2, 3, 4. łatwo też sprawdzić, że:

$$\begin{aligned} P(X = 1) &= P(B_1) = \frac{2}{5}, \\ P(X = 2) &= P(C_1 \cap B_2) = P(B_2 | C_1)P(C_1) = \frac{2}{4} \cdot \frac{3}{5} = \frac{3}{10}, \\ P(X = 3) &= P(C_1 \cap C_2 \cap B_3) = P(B_3 | C_1 \cap C_2)P(C_1 \cap C_2) = \frac{2}{3} \cdot \frac{2}{4} \cdot \frac{3}{5} = \frac{2}{10}, \\ P(X = 4) &= 1 - (P(X = 1) + P(X = 2) + P(X = 3)) = \frac{1}{10}. \end{aligned}$$

Przykład – 5.9. Pan Adam jeździ rano do pracy i wraca z pracy wieczorem autobusami, które jeżdżą dokładnie co 10 minut. Niech X oznacza sumę czasów rano i wieczorem spędzonych przez pana Adama na przystankach.

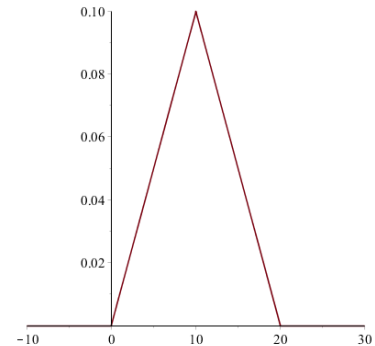
Znajdziemy rozkład X .

Wyznaczamy dystrybuantę F . Czyli dla każdego $x \in \mathbb{R}$ wyznaczmy $F(x) = P(X \leq x)$. Oczywiście $F(x) = 0$ dla $x \leq 0$ oraz $F(x) = 1$ dla $x \geq 20$. Dla pozostałych x korzystamy z modelu prawdopodobieństwa geometrycznego. Dla $0 \leq x \leq 10$ mamy $F(x) = \frac{x^2/2}{100}$. Dla $10 \leq x \leq 20$ mamy $F(x) = \frac{100 - (20-x)^2/2}{100}$.



Zauważmy, że jest to funkcja ciągła, łatwo też pokazać, że jest we wszystkich punktach różniczkowalna. Jej pochodna F' jest więc gęstością rozkładu P_X , czyli krótko mówiąc jest gęstością zmiennej losowej X .

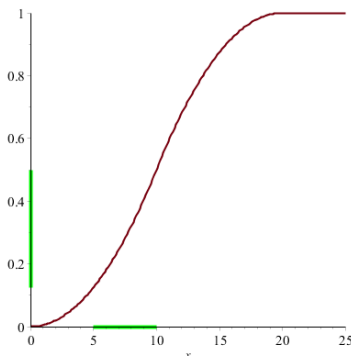
$$f(x) = F'(x) = \begin{cases} 0, & \text{dla } x \leq 0, \\ \frac{x}{100}, & \text{dla } 0 < x \leq 10, \\ \frac{20-x}{100}, & \text{dla } 10 < x \leq 20, \\ 0, & \text{dla } 20 < x. \end{cases}$$



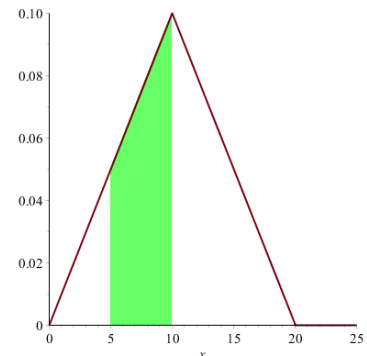
Znając gęstość lub dystrybuantę można wyznaczać prawdopodobieństwa zdarzeń opisanych przez X .

$$P(a < X < b) = P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x) dx.$$

Jakie jest prawdopodobieństwo tego, że pan Adam spędzi na przystankach co najmniej pięć, lecz nie więcej niż dziesięć minut?

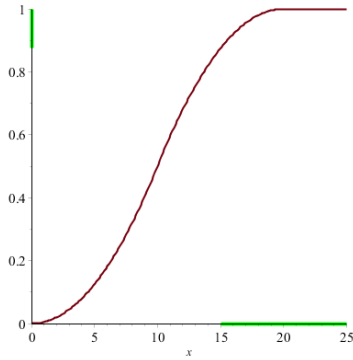


$$P(5 \leq X \leq 10) = F(10) - F(5) = \frac{1}{2} - \frac{1}{8} = \frac{3}{8}.$$

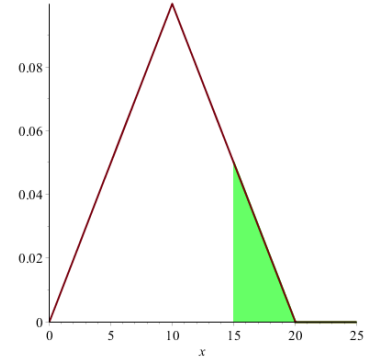


$$P(5 \leq X \leq 10) = \int_5^{10} f(x) dx = \frac{3}{8}.$$

Jakie jest prawdopodobieństwo tego, że pan Adam spędzi na przystankach więcej niż 15 minut?



$$P(X \geq 15) = 1 - F(15) = 1 - \frac{7}{8} = \frac{1}{8}.$$



$$P(X \geq 15) = \int_{15}^{\infty} f(x) dx = \frac{1}{8}.$$

Jakie jest prawdopodobieństwo tego, że pan Adam spędzi na przystankach dokładnie tyle samo czasu, powiedzmy po c minut na każdym?

Tego zdarzenia nie opisuje zmienna X ! Jednak, szukane prawdopodobieństwo jest nie większe niż $P(X = 2c) = \int_{2c}^{2c} f(x) dx = 0$.

Ile wynosi prawdopodobieństwo tego, że pan Adam spędzi na przystankach rano i wieczorem mniej niż 5 minut na każdym?

Tego zdarzenia nie opisuje zmienna losowa X , niemniej odpowiedź jest oczywista: $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

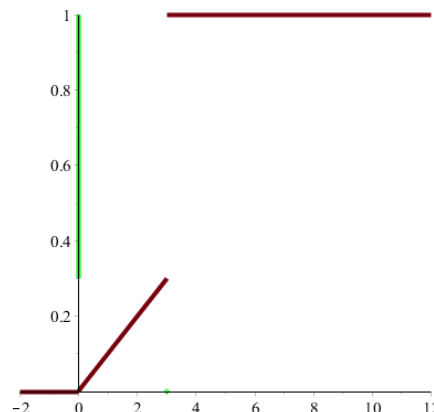
Przykład – 5.10 (kontynuacja przykładu 5.9). Pewnego dnia pan Adam zmienia strategię. Gdy wieczorem wychodzi z pracy czeka na przystanku maksymalnie 3 minuty i gdy jego autobus nie przyjedzie idzie do domu na piechotę. Jaki jest rozkład czasu X spędzonego na przystanku wieczorem?

Zauważmy, że $X = \min(Y, 3)$, gdzie Y ma rozkład jednostajny na odcinku $(0, 10)$ ($P_Y = U(0, 10)$). Niech $F = F_X$. Wtedy

$$F(x) = P(X \leq x) = P(\min(Y, 3) \leq x) = \begin{cases} 0 & \text{dla } x \leq 0, \\ \frac{x}{10} & \text{dla } 0 < x < 3, \\ 1 & \text{dla } 3 \leq x. \end{cases}$$

Ta dystrybuanta **nie jest ciągła w punkcie $x = 3$** , więc mamy do czynienia z rozkładem, który **nie jest ani dyskretny ani ciągły**.

Przy okazji: $P(X = 3) = \frac{7}{10}$.



5.2 Rozkłady brzegowe i warunkowe

Przykład – 5.11. Rozważmy rzut dwiema kostkami symetrycznymi. Niech X oznacza numer kostki na której wypadła większa liczba, lub 0, gdy na obydwu kostkach wypadła ta sama liczba, a Y oznacza maksimum uzyskanych oczek. Znajdziemy rozkład wektora losowego (X, Y) .

Z dodatnimi prawdopodobieństwami (X, Y) może potencjalnie przyjmować 18 wartości. W tym przypadku można zobrazować to tabelką zawierającą licznosci zdarzeń $(X = i, Y = j)$, $i = 0, 1, 2$, $j = 1, 2, 3, 4, 5, 6$.

| | | | | | | |
|-----------------|---|---|---|---|---|---|
| $X \setminus Y$ | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 0 | 1 | 2 | 3 | 4 | 5 |
| 2 | 0 | 1 | 2 | 3 | 4 | 5 |

W takim razie prawdopodobieństwa $p_{i,j} = P(X = i, Y = j)$ też tworzą macierz:

| | | | | | | |
|-----------------|------|------|------|------|------|------|
| $X \setminus Y$ | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 |
| 1 | 0 | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 |
| 2 | 0 | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 |

Przykład – 5.12 (kontynuacja przykładu poprzedniego). Sumując wiersze i kolumny otrzymujemy rozkłady zmiennych X oraz Y .

| | | | | | | | |
|-----------------|------|------|------|------|------|-------|------|
| $X \setminus Y$ | 1 | 2 | 3 | 4 | 5 | 6 | X |
| 0 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/6 |
| 1 | 0 | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 5/12 |
| 2 | 0 | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 5/12 |
| Y | 1/36 | 3/36 | 5/36 | 7/36 | 9/36 | 11/36 | |

Nazywamy je rozkładami brzegowymi. Ogólnie.

Definicja – 5.13 (Rozkład brzegowy). Dla danego rozkładu $n + m$ wymiarowego $Q : \mathcal{B}(\mathbb{R}^n \times \mathbb{R}^m) \rightarrow \mathbb{R}$ określamy rozkłady brzegowe $Q_1 : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}$, $Q_2 : \mathcal{B}(\mathbb{R}^m) \rightarrow \mathbb{R}$ za pomocą formuły:

$$Q_1(A) = Q(A \times \mathbb{R}^m), \quad Q_2(B) = Q(\mathbb{R}^n \times B).$$

W dalszym ciągu rozważamy przypadek $n = m = 1$. Uogólnienie na dowolne n, m jest oczywiste.

Niech (X, Y) będzie dwuwymiarowym wektorem losowym określonym na przestrzeni probabilistycznej (Ω, Σ, P) i niech Q będzie jego rozkładem, czyli $Q = P_{(X, Y)}$. Wtedy: dla dowolnego $A \in \mathcal{B}(\mathbb{R})$ mamy:

$P_X(A) = P(X \in A) = P(X \in A, Y \in \mathbb{R}) = P((X, Y) \in A \times \mathbb{R}) = P_{(X, Y)}(A \times \mathbb{R}) = Q(A \times \mathbb{R}) = Q_1(A)$. Tak więc $P_X = Q_1$. Podobnie $P_Y = Q_2$.

Uwaga – 5.14. Rozkłady brzegowe wektora losowego (X, Y) pokrywają się z rozkładami zmiennych losowych X oraz Y .

Niech Q będzie dyskretnym rozkładem 2-wymiarowym skupionym na zbiorze (co najwyżej przeliczalnym) K . Możemy taki rozkład jednoznacznie scharakteryzować przez podanie dwóch macierzy; punktów oraz ich prawdopodobieństw. Mianowicie, weźmy **najmniejsze** zbiory co najwyżej przeliczalne $K_1, K_2 \subset \mathbb{R}$ takie, że $K \subset K_1 \times K_2$ i ustawmy je w ciągi, powiedzmy $K_1 = \{x_i\}_{i=1}^M$, $K_2 = \{y_j\}_{j=1}^N$, $M, N \leq \infty$ oraz niech $p_{ij} = Q(x_i, y_j)$. Widać, że:

$$p_{ij} \geq 0 \text{ dla wszystkich } i, j, \text{ oraz } \sum_{i,j} p_{ij} = 1.$$

Z doboru zbiorów K_1, K_2 wynika też, że $\forall i \exists j : p_{ij} > 0$ oraz $\forall j \exists i : p_{ij} > 0$. Gdyby tak nie było, można by zmniejszyć K_1 lub K_2 .

Para $(\{x_i, y_j\}, \{p_{ij}\})$ w pełni charakteryzuje rozkład Q .

Wtedy rozkłady brzegowe Q_1, Q_2 są określone odpowiednio przez pary ciągów. Q_1 przez $(\{x_i\}, \{p_{i\cdot}\})$, Q_2 przez $(\{y_j\}, \{p_{\cdot j}\})$, przy czym:

$$p_{i\cdot} = Q_1(x_i) = Q(\{x_i\} \times \mathbb{R}) = \sum_j p_{ij}, \quad p_{\cdot j} = Q_2(y_j) = Q(\mathbb{R} \times \{y_j\}) = \sum_i p_{ij}.$$

Ponieważ występujące powyżej sumy mają wszystkie składniki nieujemne oraz przynajmniej jeden składnik dodatni, wszystkie liczby $p_{i\cdot}$ oraz $p_{\cdot j}$ są dodatnie. W dalszej części mówiąc o rozkładach dyskretnych wielowymiarowych będziemy zawsze zakładać, że mają powyższą własność.

Niech X, Y oznaczają liczbę oczek które wypadną w dwóch rzutach kostką.

| $X \setminus Y$ | 1 | 2 | 3 | 4 | 5 | 6 | X |
|-----------------|------|------|------|------|------|------|-----|
| 1 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/6 |
| 2 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/6 |
| 3 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/6 |
| 4 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/6 |
| 5 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/6 |
| 6 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/6 |
| Y | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | |

Określmy pojęcie rozkładu warunkowego tylko w kontekście zmiennych losowych. Niech (X, Y) będzie wektorem o dyskretnym rozkładzie 2-wymiarowym danym przez $(\{(x_i, y_j)\}, \{p_{ij}\})$. Czyli $P(X = x_i, Y = y_j) = p_{ij}$. Niech:

$$p_{i|j} = P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\cdot j}},$$

$$p_{j|i} = P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i \cdot}}.$$

Uwaga. Oznaczenia są formalnie niepoprawne. Ale są zwyczajowo stosowane.

Definicja – 5.17. Rozkład dany przez ciągi $\{x_i\}, \{p_{i|j}\}$ nazywamy *rozkładem warunkowym X pod warunkiem $Y = y_j$* . Oznaczamy go $P_{X|Y=y_j}$.

Rozkład dany przez ciągi $\{y_j\}, \{p_{j|i}\}$ nazywamy *rozkładem warunkowym Y pod warunkiem $X = x_i$* . Oznaczamy go $P_{Y|X=x_i}$.

Przykład – 5.18. Wróćmy do rozkładu:

| $X \setminus Y$ | 1 | 2 | 3 | 4 | 5 | 6 | X |
|-----------------|------|------|------|------|------|-------|------|
| 0 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/6 |
| 1 | 0 | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 5/12 |
| 2 | 0 | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 5/12 |
| Y | 1/36 | 3/36 | 5/36 | 7/36 | 9/36 | 11/36 | |

$P_{X|Y=1}$ jest rozkładem jednopunktowym δ_0 .

$P_{X|Y=6}$ jest rozkładem skupionym w punktach 0, 1, 2 z prawdopodobieństwami $\frac{1}{11}, \frac{5}{11}, \frac{5}{11}$.

$P_{Y|X=0}$ jest skupiony w punktach 1, 2, 3, 4, 5, 6 z równymi prawdopodobieństwami.

$$p_{ij} = p_{i|j}p_{\cdot j} = p_{j|i}p_{i \cdot}.$$

Znając rozkłady brzegowe i warunkowe można wyznaczyć rozkład wektora losowego.

Przykład – 5.19. Rzucamy kostką symetryczną uzyskując X oczek, a następnie rzucamy X razy monetą symetryczną uzyskując Y orłów. Interesuje nas rozkład wektora losowego (X, Y) . Oczywiście wektor ten może przyjmować wartości w punktach $(x_i, y_j), i = 1, \dots, 6, j = 0, 1, \dots, 6$ z prawdopodobieństwami $p_{ij} = p_{j|i}p_{i \cdot}$, gdzie $p_{j|i}, p_{i \cdot}$ są dane. Na przykład: $p_{53} = \binom{5}{3} \frac{1}{2^5} \frac{1}{6} = \frac{5}{96}, p_{35} = 0 \cdot \frac{1}{6} = 0$.

Niech (X, Y) będzie wektorem o ciągłym rozkładzie 2-wymiarowym danym przez gęstość f .

Formalnie nie możemy (jeszcze) mówić o rozkładzie warunkowym pod warunkiem, którego prawdopodobieństwo jest równe zero. Możemy jednak formalnie zdefiniować funkcje:

$$f_{X|Y=y}(x) = f(x | y) = \begin{cases} \frac{f(x,y)}{\int_{\mathbb{R}} f(x,y) dx} = \frac{f(x,y)}{f_Y(y)}, & \text{gdy } f_Y(y) > 0, \\ 0, & \text{gdy } f_Y(y) = 0. \end{cases}$$

$$f_{Y|X=x}(y) = f(y | x) = \begin{cases} \frac{f(x,y)}{\int_{\mathbb{R}} f(x,y) dy} = \frac{f(x,y)}{f_X(x)}, & \text{gdy } f_X(x) > 0, \\ 0, & \text{gdy } f_X(x) = 0. \end{cases}$$

Są to gęstości (ćwiczenie).

Definicja – 5.20. Rozkład o gęstości $f(\cdot|y)$ nazywamy rozkładem warunkowym X pod warunkiem ($Y = y$) i oznaczamy $P_{X|Y=y}$.

Rozkład o gęstości $f(\cdot|x)$ nazywamy rozkładem warunkowym Y pod warunkiem ($X = x$) i oznaczamy $P_{Y|X=x}$.

Podobnie jak w przypadku dyskretnym:

$$f(x, y) = f(y|x)f_X(x) = f(x|y)f_Y(y).$$

Przykład – 5.21. Losujemy według rozkładu jednostajnego liczbę X z odcinka $[0, 1]$ a następnie według rozkładu jednostajnego liczbę Y z odcinka $[0, X]$. Według jakiego rozkładu została wylosowana liczba Y ?

Mamy kolejno:

$f_X = I_{[0,1]}$ – funkcja charakterystyczna odcinka $[0, 1]$.

Gęstość warunkowa $f(\cdot|x) = \frac{1}{x}I_{[0,x]}$, dla $0 < x \leq 1$.

$$f(x, y) = f(y|x)f_X(x) = \begin{cases} \frac{1}{x} & \text{dla } 0 < y \leq x \leq 1 \\ 0 & \text{dla pozostałych } (x, y) \in \mathbb{R}^2. \end{cases}$$

$$f_Y(y) = \int_{\mathbb{R}} f(x, y) dx = \int_y^1 \frac{1}{x} dx = \ln x|_y^1 = -\ln y, \text{ dla } 0 < y \leq 1,$$

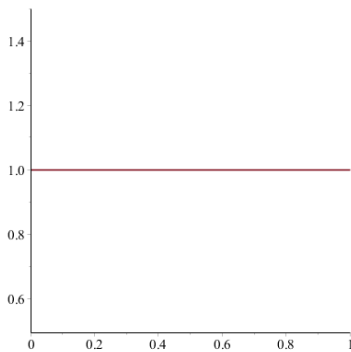
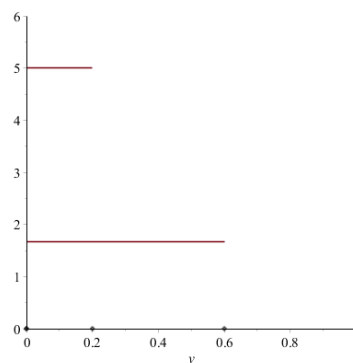
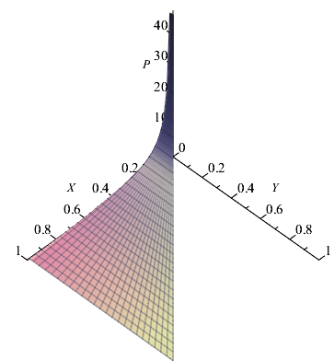
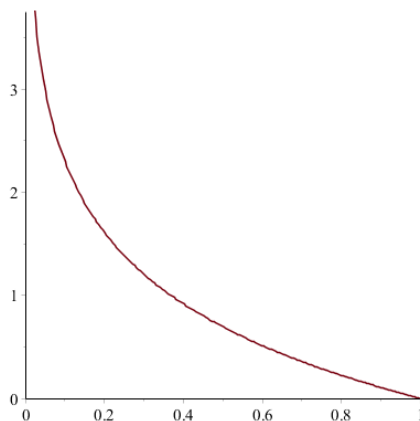
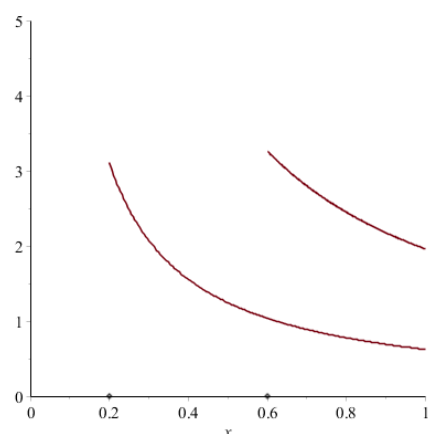
$$f_Y(y) = 0 \text{ dla pozostałych } y.$$

Można też wyznaczyć gęstość rozkładu warunkowego $P_{X|Y=y}$ dla $0 < y < 1$:

$$f(x|y) = \frac{1/x}{-\ln y} = -\frac{1}{x \ln y} \text{ dla } y < x < 1.$$

$$f_X(x) = 1 \text{ dla } 0 \leq x \leq 1,$$

$$f(y|x) = f_{Y|X=x}(y) = \frac{1}{x}, \text{ dla } 0 \leq y \leq x \leq 1.$$


 f_X

 $f_{Y|X=0.2}, f_{Y|X=0.6}$

 $f(x, y)$

 f_Y

 $f_{X|Y=0.2}, f_{X|Y=0.6}$

5.3 Niezależność zmiennych/wektorów losowych

Dany jest ciąg X_1, \dots, X_n zmiennych losowych określonych na przestrzeni probabilistycznej (Ω, Σ, P) .

Definicja – 5.22. Mówmy, że X_1, \dots, X_n są *niezależne* \iff dla każdego zbiorów borelowskich, $B_i \in \mathcal{B}(\mathbb{R})$, $i = 1, \dots, n$

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdot \dots \cdot P(X_n \in B_n).$$

O wektorach losowych, które nie są niezależne mówimy, że są *zależne*.

Przykład – 5.23. Rozważmy schemat klasyczny, gdzie $\Omega = \{1, \dots, 6\}^2$.

Zmienne losowe X_1, X_2 zdefiniowane jako $X_1(i, j) = i$, $X_2(i, j) = j$ są niezależne.

Zmienne losowe Y_1, Y_2 zdefiniowane jako $Y_1(i, j) = i$, $Y_2(i, j) = i + j$ są zależne: $P(Y_1 = 2, Y_2 = 10) = 0$, $P(Y_1 = 2) \cdot P(Y_2 = 10) = 1/6 \cdot 3/36$.

Następne twierdzenie udowodnimy dla 2 zmiennych losowych. Prawdziwe są analogiczne twierdzenia dla dowolnej skończonej liczby zmiennych losowych, jednak dowody stają się bardziej skomplikowane w zapisie.

Twierdzenie – 5.24. Zmienne losowe X_1, X_2 są niezależne $\iff P_{(X_1, X_2)} = P_{X_1} \times P_{X_2}$.

Dowód. „ \implies ” Pamiętamy, że iloczyn kartezjański dwóch miar probabilistycznych, powiedzmy Q_1, Q_2 , jest jednoznacznie określony przez warunek $(Q_1 \times Q_2)(A_1 \times A_2) = Q_1(A_1) \cdot Q_2(A_2)$, dla dowolnych mierzalnych A_1, A_2 .

Tutaj dla dowolnych $B_1, B_2 \in \mathcal{B}(\mathbb{R})$ mamy: $P_{(X_1, X_2)}(B_1 \times B_2) = P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1) \cdot P(X_2 \in B_2) = P_{X_1}(B_1) \cdot P_{X_2}(B_2)$. Czyli $P_{(X_1, X_2)}$ jest iloczynem kartezjańskim P_{X_1}, P_{X_2} .

„ \impliedby ” Oczywiście. □

Twierdzenie – 5.25. Niech (X, Y) będzie wektorem o dyskretnym rozkładzie 2-wymiarowym danym przez $(\{(x_i, y_j)\}, \{p_{ij}\})$. Czyli $P(X = x_i, Y = y_j) = p_{ij}$. Wtedy:

$$X, Y \text{ są niezależne} \iff \forall i, j : p_{ij} = p_i \cdot p_j.$$

Dowód. „ \implies ” $p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) = p_i \cdot p_j$.

„ \impliedby ” Niech A oraz B będą zbiorami borelowskimi.

$$P(X \in A, Y \in B) = \sum_{(i,j):(x_i, y_j) \in A \times B} p_{ij} = \sum_{i,j:x_i \in A, y_j \in B} p_i \cdot p_j = \sum_{i:x_i \in A} p_i \cdot \sum_{j:y_j \in B} p_j = P(X \in A)P(Y \in B). \quad \square$$

Twierdzenie – 5.26. Niech (X, Y) będzie wektorem o ciągłym rozkładzie 2-wymiarowym o gęstości f . Wtedy:

$$X, Y \text{ są niezależne} \iff \forall x, y \in \mathbb{R} : f(x, y) = f_X(x)f_Y(y).$$

Dowód. Twierdzenie Fubiniego □

Uwaga – 5.27. Jeżeli wektor losowy (X, Y) ma rozkład dyskretny, lub rozkład ciągły, a zmienne losowe X, Y są niezależne, to wszystkie rozkłady warunkowe są równe rozkładom brzegowym.

Prostą ilustracją powyższego twierdzenia i uwagi stanowi następujący:

Przykład – 5.28. Załóżmy, że wektor losowy (X, Y) ma rozkład jednostajny na prostokącie $[a, b] \times [c, d]$. W takim razie (X, Y) ma gęstość:

$$f_{(X,Y)} = \frac{1}{(b-a)(d-c)} I_{[a,b] \times [c,d]},$$

gdzie $I_{[a,b] \times [c,d]}$ oznacza funkcję charakterystyczną prostokąta $[a, b] \times [c, d]$. Jest ona iloczynem funkcji charakterystycznych przedziałów $[a, b]$ oraz $[c, d]$, mamy więc:

$$f_{(X,Y)}(x, y) = f_X(x) \cdot f_Y(y).$$

Wtedy też $f_{Y|X=x} = f_Y$ oraz $f_{X|Y=y} = f_X$.

5.4 Funkcje zmiennych/wektorów losowych

Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną, $X : \Omega \rightarrow \mathbb{R}^n$ wektorem losowym, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ funkcją borelowską ($\forall C \in \mathcal{B}(\mathbb{R}^m) \ g^{-1}(C) \in \mathcal{B}(\mathbb{R}^n)$).

Zgodnie ze zwyczajem oznaczamy złożenie $g \circ X$ symbolem $g(X)$,

$$g(X) = g \circ X : \Omega \ni \omega \rightarrow g(X(\omega)) \in \mathbb{R}^m.$$

Ważny problem praktyczny: znając rozkład X wyznaczyć rozkład $Y = g(X)$.

Rozwiązanie formalne. $P_Y(C) = P(g(X)^{-1}(C)) = P((g \circ X)^{-1}(C)) = P(X^{-1}(g^{-1}(C))) = P_X(g^{-1}(C))$.

$$P_Y(C) = P_X(g^{-1}(C)).$$

Z praktycznego punktu widzenia jest to wzór mało przydatny. W poszczególnych przypadkach możemy stosować różne sposoby.

Przykład – 5.29. Zmienna losowa X ma rozkład dany przez ciągi $-1, 0, 1, 2$ oraz $1/2, 1/8, 1/4, 1/8$. Wtedy zmienna X^2 ma rozkład dany przez ciągi $0, 1, 4$, oraz $1/8, 3/4, 1/8$.

Zauważmy, że funkcja $g(X)$ wektora losowego X o rozkładzie dyskretnym skupionym na zbiorze K (o którym wiemy, że jest skończony lub przeliczalny) ma rozkład dyskretny skupiony na zbiorze $g(K)$. Dla rozkładów ciągłych to nie jest prawdą: na przykład, gdy g jest funkcją stałą $g(X)$ ma rozkład jednopunktowy.

Przykład – 5.30. Znajdziemy rozkład $\cos X$, gdy X ma rozkład jednostajny na odcinku $(-\pi, \pi)$. Policzmy dystrybuantę tej funkcji w punktach $y \in (-1, 1)$. Niech $y = \cos x$, gdzie $x \in (0, \pi)$, inaczej $x = \arccos y$. Wtedy mamy:

$$F_{\cos X}(y) = P(\cos X \leq y) = P(-\pi, -x] + P[x, \pi) = \frac{-x + \pi}{2\pi} + \frac{\pi - x}{2\pi} = \frac{\pi - \arccos y}{\pi}.$$

Różniczkując, otrzymamy gęstość:

$$f_{\cos X}(y) = \frac{1}{\pi\sqrt{1-y^2}} \text{ dla } y \in (-1, 1)$$

oraz 0 dla pozostałych y .

Przykład – 5.31. Niech X, Y będą niezależnymi zmiennymi losowymi o rozkładzie $U(0, 1)$ każda. Znajdziemy rozkład zmiennej losowej $Z = \frac{X}{Y}$.

Znajdziemy najpierw dystrybuantę F_Z . Niech $z > 0$. $F_Z(z) = P(F_Z \leq z) = P(\frac{X}{Y} \leq z)$. Ponieważ (X, Y) ma rozkład geometryczny na kwadracie $[0, 1]^2$, możemy to prawdopodobieństwo interpretować jako prawdopodobieństwo geometryczne. Łatwo widać, że:

$$F_Z(z) = \begin{cases} 0, & \text{dla } z \leq 0, \\ \frac{z}{2}, & \text{dla } 0 < z < 1, \\ 1 - \frac{1}{2z}, & \text{dla } 1 \leq z. \end{cases}$$

Różniczkując, otrzymujemy gęstość:

$$f_Z(z) = \begin{cases} 0, & \text{dla } z \leq 0, \\ \frac{1}{2}, & \text{dla } 0 < z < 1, \\ \frac{1}{2z^2}, & \text{dla } 1 \leq z. \end{cases}$$

Przykład – 5.32. Niech X będzie dowolnie ustaloną zmienną losową, $F = F_X$ jej dystrybuantą, $a, b \in \mathbb{R}$ ustalonymi liczbami, $a \neq 0$. Policzmy dystrybuantę zmiennej losowej $Y = aX + b$.

Dla $a > 0$ mamy

$$F_Y(x) = P(Y \leq x) = P(aX + b \leq x) = P\left(X \leq \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right).$$

Podobnie, dla $a < 0$

$$\begin{aligned} F_Y(x) &= P(Y \leq x) = P(aX + b \leq x) = P\left(X \geq \frac{x-b}{a}\right) \\ &= 1 - P\left(X < \frac{x-b}{a}\right) = 1 - F_X\left(\frac{x-b}{a}\right)^-. \end{aligned}$$

Załóżmy teraz dodatkowo, że zmienna X ma gęstość f (dla uproszczenia zakładamy, że f jest ciągła, co w świetle następnego przykładu nie jest konieczne). Wtedy wiemy, że dystrybuanta F_X jest różniczkowalna; z powyższych wzorów także F_Y jest różniczkowalna, a więc Y ma rozkład ciągły o gęstości

$$g(x) = \frac{1}{|a|} f\left(\frac{x-b}{a}\right)$$

Powyższy wzór jest szczególnym przypadkiem następującego twierdzenia.

Twierdzenie – 5.33. *Niech X będzie wektorem losowym o n -wymiarowym rozkładzie ciągłym i niech $f : \mathbb{R}^n \rightarrow \mathbb{R}$ będzie jego gęstością. Zakładamy ponadto, że $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ jest dyfeomorfizmem. Wtedy wektor losowy $\varphi(X)$ ma również rozkład ciągły o gęstości g danej wzorem*

$$g(x) = |\text{Jac}_x \varphi^{-1}| f(\varphi^{-1}(x)).$$

Dowód. Korzystamy z definicji rozkładu ciągłego. Z twierdzenia o zmianie zmiennych mamy dla każdego zbioru borelowskiego A

$$P(\varphi(X) \in A) = P(X \in \varphi^{-1}(A)) = \int_{\varphi^{-1}(A)} f(x) dx = \int_A f(\varphi^{-1}(x)) |\text{Jac}_x \varphi^{-1}| dx. \quad \square$$

Funkcje niezależnych wektorów losowych są niezależne.

Formalnie:

Twierdzenie – 5.34. *Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną, $X : \Omega \rightarrow \mathbb{R}^n$, $Y : \Omega \rightarrow \mathbb{R}^m$ wektorami losowymi. Niech $g : \mathbb{R}^n \rightarrow \mathbb{R}^k$, $h : \mathbb{R}^m \rightarrow \mathbb{R}^l$ będą funkcjami borelowskimi.*

$$X, Y \text{ są niezależne} \implies g(X), h(Y) \text{ są niezależne.}$$

Dowód. Dla dowolnych zbiorów borelowskich C, D mamy:

$$P(g(X) \in C, h(Y) \in D) = P((g \circ X)^{-1}(C) \cap (h \circ Y)^{-1}(D)) = P(X^{-1}(g^{-1}(C)) \cap Y^{-1}(h^{-1}(D))) = P(X^{-1}(g^{-1}(C))) \cdot P(Y^{-1}(h^{-1}(D))) = P((g \circ X)^{-1}(C)) \cdot P((h \circ Y)^{-1}(D)) = P(g(X) \in C) \cdot P(h(Y) \in D). \quad \square$$

5.5 Pytania

Pytanie 5.1. Rzucono trzema kostkami i podano informację, czy, i na których kostkach wypadła „6”. Na odpowiedniej przestrzeni probabilistycznej podać przykłady funkcji, które są oraz które nie są zmiennymi losowymi.

ROZWIĄZANIE

Pytanie 5.2. Salę oświetla 150 żarówek: po 15 żarówek w 10 rzędach. Wiadomo, że prawdopodobieństwo zepsucia się pojedynczej żarówki w ciągu nachodzącego tygodnia wynosi $p = 0.06$. Wiemy też, że żarówki psują się niezależnie od siebie. Niech X oznacza liczbę rzędów, w których po tygodniu świeci co najmniej 14 żarówek. Wskazać rozkład X .

ROZWIĄZANIE

Pytanie 5.3. W przykładzie 3.2: X , odpowiednio Y , oznaczają liczbę kolejnych dni, w których nieprzerwanie sprząta Kaja, odpowiednio Leon, licząc od dnia zawarcia umowy. Znaleźć rozkład X oraz rozkład Y . Czy zmienne te są niezależne?

ROZWIĄZANIE

Pytanie 5.4. Niech $G \subset \mathbb{R}^2$ będzie zbiorem otwartym o mierze skończonej, (X, Y) wektorem losowym o rozkładzie $U(G)$. Podać przykład takiego zbioru G , że X, Y są: (a) niezależne, (b) zależne.

ROZWIĄZANIE

Pytanie 5.5. Dane są dwie zmienne losowe niezależne X, Y o rozkładzie $U(0, 1)$ każda. Znajdź dystrybuantę i gęstość zmiennych losowych $\min(X, Y)$, $\max(X, Y)$. Czy te zmienne są niezależne? Wskazać dystrybuantę rozkładu $P_{\min(X, Y) | X=a}$ dla $a \in (0, 1)$.

ROZWIĄZANIE

Pytanie 5.6. W schemacie klasycznym, gdy zbiór $\Omega = \{1, \dots, 6\}$, definiujemy zmienne losowe m, M jako $m(i, j) = \min(i, j)$, $M(i, j) = \max(i, j)$. Znaleźć: rozkład wektora losowego (m, M) oraz rozkład $P_{M|m=1}$.

ROZWIĄZANIE

Pytanie 5.7. Sformułować fakty 5.24–5.28 w wersji wielowymiarowej i zastanowić się nad ich uzasadnieniem.

Rozdział 6

Nadzieja matematyczna i wariancja

W pewnych zagadnieniach nie jest potrzebna (lub nie jest możliwa) znajomość rozkładu zmiennej losowej lub wektora losowego. Natomiast warto znać pewne parametry tego rozkładu. Najczęściej używanymi parametrami są liczby charakteryzujące tendencję centralną oraz wielkość rozrzutu. Kluczową rolę odgrywa tak zwana nadzieja matematyczna, zwana też wartością oczekiwaną, służąca do opisu tendencji centralnej oraz do budowy wielu innych parametrów, w tym parametrów charakteryzujących rozrzut: wariancji i odchylenia standardowego.

6.1 Nadzieja matematyczna – definicja i własności

Zanim podamy ogólną definicję $E(X)$, nadziei matematycznej zmiennej losowej X , opartą na pojęciu całki, rozpatrzmy dwie najbardziej typowe sytuacje szczególne.

Definicja – 6.1 (Rozkład dyskretny skończony). X ma rozkład zadany przez ciągi: $x_1, x_2, \dots, x_n, p_1, p_2, \dots, p_n$, czyli $P(X = x_i) = p_i$. Wtedy

$$E(X) = \sum_{i=1}^n x_i p_i.$$

Komentarz. Powyższy wzór z sumą nieskończoną również jest prawdziwy dla dowolnych rozkładów dyskretnych. Wiele własności opisanych dalej jest też prawdziwych w przypadku ogólnym. Wymaga to jednak przyjęcia założeń gwarantujących sumowalność szeregów – z reguły sformułowania twierdzeń ją implikują (patrz twierdzenie 6.11 i wniosek 6.12).

Definicja – 6.2 (Rozkład ciągły). X ma rozkład zadany przez gęstość f , czyli $P(X \in (a, b)) = \int_a^b f(x) dx$. Wtedy

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx.$$

Motywacja. Dla uproszczenia załóżmy, że $f : [a, b] \rightarrow \mathbb{R}$ jest funkcją ciągłą. Niech $a = x_0 < x_1 < \dots < x_n = b$ będzie podziałem odcinka $[a, b]$, $\xi_i \in [x_{i-1}, x_i]$ będą takie, że

$$(x_i - x_{i-1})f(\xi_i) = \int_{x_{i-1}}^{x_i} f(x) dx.$$

Określmy zmienną losową \hat{X} przyjmującą wartości ξ_i z prawdopodobieństwami $p_i = P(\hat{X} = \xi_i) = \int_{x_{i-1}}^{x_i} f(x) dx$.

Intuicja. \hat{X} przybliża X , więc $E(\hat{X})$ powinna przybliżać $E(X)$. Mamy:

$$E(\hat{X}) = \sum_{i=1}^n \xi_i p_i = \sum_{i=1}^n \xi_i (x_i - x_{i-1}) f(\xi_i) = \sum_{i=1}^n \xi_i f(\xi_i) (x_i - x_{i-1}).$$

A więc

$$E(\hat{X}) \rightarrow \int_a^b x f(x) dx.$$

Przykład – 6.3. Załóżmy, że zmienna losowa X ma rozkład jednostajny na przedziale $[a, b]$. Oznacza to, że gęstość $f_X = \frac{1}{b-a} I_{[a,b]}$. Mamy więc:

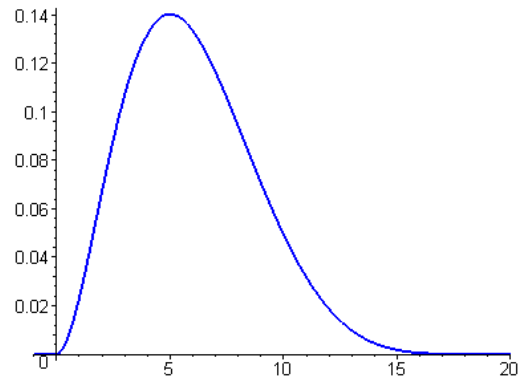
$$E(X) = \int_{\mathbb{R}} x \frac{1}{b-a} I_{[a,b]} dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{a+b}{2}.$$

Przykład – 6.4. Zmienna losowa X ma gęstość:

$$f(x) = \frac{x^2(20-x)^6}{c}, \text{ dla } 0 \leq x \leq 20,$$

gdzie $c = \frac{128000000000}{63}$, $f(x) = 0$ dla pozostałych x . Tego typu rozkłady mają na przykład okresy używania samochodu przez jednego właściciela.

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^{20} x \cdot \frac{x^2(20-x)^6}{c} dx = 6.$$



Powyższe dwie definicje stanowią szczególne przypadki definicji opartej na pojęciu całki. Zaczniemy od przypomnienia definicji całki.

Definicja – 6.5 (całka względem miary). Niech Ω będzie niepustym zbiorem, Σ σ -algebrą na zbiorze Ω , $\mu : \Sigma \rightarrow \mathbb{R} \cup \infty$ miarą, a $f : \Omega \rightarrow \mathbb{R}$ funkcją.

1. Niech $f = I_A$ będzie funkcją charakterystyczną zbioru $A \in \Sigma$, $f = I_A$. Definiujemy:

$$\int_{\Omega} f d\mu = \mu(A).$$

2. Niech f będzie mierzalną funkcją prostą nieujemną, to znaczy:

$$f(x) = c_i, \text{ dla } x \in A_i,$$

gdzie A_1, \dots, A_k są zbiorami mierzalnymi parami rozłącznymi i dającymi w sumie całą przestrzeń Ω , a liczby c_1, \dots, c_k są nieujemne. Definiujemy:

$$\int_{\Omega} f d\mu = \sum_{i=1}^k c_i \int_{\Omega} I_{A_i} d\mu = \sum_{i=1}^k c_i \mu(A_i).$$

3. Niech f będzie funkcją mierzalną nieujemną. Dowodzi się, że istnieją funkcje proste nieujemne mierzalne f_n , $n = 1, 2, 3, \dots$ o następujących własnościach:

$$f_n(x) \leq f_{n+1}(x),$$

dla wszystkich n oraz x , i

$$\lim_{n \rightarrow \infty} f_n(x) = f(x).$$

Definiujemy:

$$\int_{\Omega} f d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu.$$

4. Niech f będzie dowolną funkcją mierzalną. Jak łatwo widać, zachodzi związek:

$$f = f^+ - f^-,$$

gdzie $f^+(x) = \max(f(x), 0)$, $f^-(x) = -\min(f(x), 0)$. Są one nieujemnymi funkcjami mierzalnymi. Definiujemy:

$$\int_{\Omega} f d\mu = \int_{\Omega} f^+ d\mu - \int_{\Omega} f^- d\mu,$$

o ile ma sens.

Komentarze.

1. Definicja jest poprawna, gdyż dowodzi się, że wielkość zdefiniowana w punkcie 3 nie zależy od wyboru ciągu funkcji f_n .

2. Może się zdarzyć, że definiowane przez nas wielkości są nieskończone. Tak będzie w punkcie 1 oraz 2, gdy miara choćby jednego zbioru A_i będzie nieskończona, a odpowiadająca mu liczba c_i (wysokość słupka o podstawie A_i) będzie dodatnia. W teorii miary przyjmujemy konwencję: $0 \cdot \infty = 0$, co oznacza, że gdy $c_i = 0$, to składnik $c_i \mu(A_i) = 0$.

3. Całka z funkcji nieujemnej f zdefiniowana w punkcie 3 może być nieskończona, mimo że wszystkie całki $\int_{\Omega} f_n d\mu$ są skończone.

4. Całka z pewnych funkcji mierzalnych przyjmujących wartości o różnych znakach może nie istnieć. Jest tak wtedy, gdy obie całki $\int_{\Omega} f^+ d\mu$ oraz $\int_{\Omega} f^- d\mu$ są nieskończone. W każdym innym przypadku całka istnieje, chociaż może być nieskończona.

5. Stosuje się różne oznaczenia całki $\int_{\Omega} f d\mu$. Na przykład: $\int_{\Omega} f(x) d\mu(x)$, $\int_{\Omega} f(s) \mu(ds)$.

6. Dla zbioru mierzalnego $A \in \Sigma$ definiuje się: $\int_A f d\mu = \int_{\Omega} I_A \cdot f d\mu$.

Wprowadźmy definicję ogólną nadziei matematycznej.

Definicja – 6.6 (Nadzieja matematyczna, wartość oczekiwana). Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną, $X : \Omega \rightarrow \mathbb{R}$ zmienną losową.

$$E(X) = \int_{\Omega} X dP.$$

Uwagi:

1. Nadzieja matematyczna może nie istnieć.
2. Nadzieja matematyczna może istnieć, ale być nieskończona.
3. $E(X) \in \mathbb{R} \iff E(|X|) \in \mathbb{R}$. Bo $X = X^+ - X^-$, $|X| = X^+ + X^-$.
4. **Własności nadziei matematycznej wynikają z własności całek: liniowość, nierówności, zbieżność (ćwiczenie: wypisać znane własności).**
5. Nadzieja matematyczna uogólnia pojęcie prawdopodobieństwa. Mianowicie

$$E(I_A) = P(A), \text{ dla każdego } A \in \Sigma.$$

6. Niech X ma rozkład dyskretny. Wtedy definicja ogólna pokrywa się z definicją poprzednią.

Powyższa definicja uogólnia poprzednie definicje.

Przykład – 6.7. X ma rozkład zadany przez ciągi: $x_1, x_2, \dots, x_n, p_1, p_2, \dots, p_n$.

Czyli $P(X = x_i) = p_i$.

Więc X jest kombinacją liniową funkcji charakterystycznych zbiorów $A_i = \{\omega \in \Omega : X(\omega) = x_i\}$. Oczywiście $P(A_i) = p_i$. Mamy kolejno:

$$X = \sum_{i=1}^n x_i I_{A_i}.$$

$$E(X) = E\left(\sum_{i=1}^n x_i I_{A_i}\right) = \sum_{i=1}^n x_i E(I_{A_i}) = \sum_{i=1}^n x_i p_i.$$

Poniższe twierdzenie pozwala wyznaczać nadzieję matematyczną zmiennej losowej lub funkcji zmiennej losowej poprzez całkowanie odpowiedniej funkcji względem rozkładu zmiennej.

Twierdzenie – 6.8 (całkowanie względem transportu miary). Niech $X : \Omega \rightarrow \mathbb{R}^n$ będzie wektorem losowym, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ funkcją borelowską. Wtedy:

$$E(g(X)) = \int_{\mathbb{R}^n} g dP_X,$$

przy czym obydwie strony istnieją jednocześnie.

Dowód. I. $g = I_B$, gdzie $B \in \mathcal{B}(\mathbb{R}^n)$. Wtedy: $E(g(X)) = \int_{\Omega} I_B \circ X dP = \int_{X^{-1}(B)} 1 dP = P(X^{-1}(B)) = P_X(B) = \int_{\mathbb{R}^n} I_B dP_X$.

II. g – funkcja schodkowa, czyli $g = \sum_{i=1}^k c_i I_{A_i}$, gdzie A_i są rozkładem \mathbb{R}^n na sumę zbiorów rozłącznych. Wtedy z liniowości: $E(g(X)) = E\left(\sum_{i=1}^k c_i I_{A_i}\right) = \sum_{i=1}^k c_i E(I_{A_i}) = \sum_{i=1}^k c_i \int_{\mathbb{R}^n} I_{A_i} dP_X = \int_{\mathbb{R}^n} g dP_X$.

III. g – dowolna funkcja borelowska nieujemna. Wtedy g jest granicą punktową ciągu rosnącego funkcji schodkowych g_t . Oczywiście $g_t(X)$ zmierza punktowo do $g(X)$.

Z własności całek: $E(g(X)) = \lim E(g_t(X)) = \lim \int_{\mathbb{R}^n} g_t dP = \int_{\mathbb{R}^n} g dP$.

IV. g – dowolna funkcja borelowska. Wtedy $g = g^+ - g^-$. Więc $g(X) = g(X)^+ - g(X)^-$. A więc, gdy jedna ze stron istnieje (nie jest symbolem nieoznaczonym), to istnieje druga strona i zachodzi równość. \square

Wniosek – 6.9. Kładąc $g(x) = x$ z powyższego twierdzenia otrzymujemy dla zmiennej losowej X wzór:

$$E(X) = \int_{\mathbb{R}} \text{Id} dP_X = \int_{\mathbb{R}} x dP_X(x).$$

A więc nadzieja matematyczna zmiennej losowej jest jednoznacznie wyznaczona przez rozkład tej zmiennej.

Uwaga – 6.10. Można spotkać definicję wartości oczekiwanej dowolnego jednowymiarowego rozkładu Q . Mianowicie określa się $E(Q) = \int_{\mathbb{R}} x dQ(x)$. Wtedy oczywiście $E(X) = E(P_X)$.

Twierdzenie – 6.11. 1. X ma rozkład dyskretny zadany przez ciągi: $x_1, x_2, \dots \in \mathbb{R}^n, p_1, p_2, \dots$
 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ jest funkcją borelowską. Wtedy

$$E(g(X)) = \sum g(x_i) p_i,$$

przy czym obydwie strony istnieją jednocześnie¹.

2. X ma rozkład ciągły zadany przez gęstość $f: \mathbb{R}^n \rightarrow \mathbb{R}$.
 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ jest funkcją borelowską. Wtedy

$$E(g(X)) = \int_{\mathbb{R}^n} g(x) f(x) dx,$$

przy czym obydwie strony istnieją jednocześnie. Całkowanie odbywa się według miary Lebesgue'a.

Dowód. Dowód przebiega podobnie jak dla twierdzenia 6.8 (ćwiczenie). \square

Wniosek – 6.12. X jest zmienną losową o rozkładzie dyskretnym $(\{x_i\}, \{p_i\})$, lub ma gęstość $f: \mathbb{R} \rightarrow \mathbb{R}$. Wtedy

$$E(X) = \sum_i x_i p_i, \quad \text{lub} \quad E(X) = \int_{\mathbb{R}} x f(x) dx.$$

Przykład – 6.13. Znaleźć $E(X^2)$ dla zmiennej losowej X mającej rozkład jednostajny na odcinku $[-1, 1]$. Gęstość X wyraża się więc wzorem $f(x) = \frac{1}{2} I_{[-1, 1]}$.

Sposób 1. Szukam rozkładu X^2 i korzystam z definicji nadziei dla zmiennej X^2 . Najpierw wyznaczam dystrybucję F_{X^2} . Nietrywialna sytuacja jest, gdy $0 < x < 1$.

$$F_{X^2}(x) = P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq \sqrt{x}) = \int_{-\sqrt{x}}^{\sqrt{x}} f(x) dx = \sqrt{x}. \quad f_{X^2}(x) = (F_{X^2})'(x) = \frac{1}{2\sqrt{x}}.$$

$$E(X^2) = \int_{\mathbb{R}} x f_{X^2}(x) dx = \int_0^1 x f_{X^2}(x) dx = \int_0^1 \frac{1}{2} x^{\frac{1}{2}} dx = \frac{1}{3}.$$

Sposób 2. Korzystam z poprzedniego twierdzenia. Tutaj $g(x) = x^2$.

$$E(X^2) = \int_{\mathbb{R}} g(x) f(x) dx = \int_{-1}^1 \frac{1}{2} x^2 dx = \frac{1}{3}.$$

¹Zbieżność szeregu rozumiana jest w sposób opisany w definicji 6.5.

Przykład – 6.14. Pośrednik w handlu świeżą żywnością kupuje u producenta maliny w cenie 5€/kg i sprzedaje na lokalnym rynku po 8€/kg. Pośrednik zna dość dobrze popyt na tym rynku — przyjmijmy dla uproszczenia, że wszyscy klienci detaliczni łącznie każdego dnia mogą zażądać dowolnej ilości malin z przedziału $[100, 200]$ kg (zgodnie z rozkładem jednostajnym). Ile kilogramów malin dziennie powinien kupować pośrednik, aby zmaksymalizować swoje zyski? Zakładamy, że maliny niesprzedane danego dnia psują się i trzeba je wyrzucić.

Założmy, że pośrednik kupuje a kilogramów malin, $100 \leq a \leq 200$. Niech X oznacza dzienne zapotrzebowanie na maliny. X jest więc zmienną losową o rozkładzie jednostajnym na przedziale $[100, 200]$, czyli gęstość X , f_X jest równa $1/100$ w tym przedziale i zero w pozostałych przypadkach. W takim razie pośrednik sprzedaje $Y = \min(a, X)$ malin, a jego dzienny zysk wyniesie $Z = 8Y - 5a$ euro.

Chcemy więc wskazać taką wartość a , przy której średni zysk będzie największy. W pierwszym momencie wydaje się, że wystarczy znaleźć rozkład Y i skorzystać z odpowiedniego wzoru. Jednak trudnością jest to, że rozkład zmiennej losowej Y nie jest ani dyskretny ani ciągły (dlaczego?). Niemniej skorzystamy z poprzedniego twierdzenia dla zmiennej X oraz funkcji $g(x) = \min(a, x)$ i obliczymy wartość oczekiwaną $E(Y)$.

$$E(Y) = \int_{\mathbb{R}} \min(a, x) \frac{1}{100} I_{[100, 200]}(x) dx = \frac{1}{100} \left(\int_{100}^a x dx + \int_a^{200} a dx \right).$$

Jak widać jest to funkcja kwadratowa zmiennej a , a więc $E(Z)$ też jest funkcją kwadratową a i można łatwo wyznaczyć punkt, w którym przyjmuje ona wartość największą. Mianowicie:

$E(Z) = 8E(Y) - 5a = -\frac{1}{25}a^2 + 11a - 400$ i największą wartość przyjmuje, gdy $a = \frac{275}{2} = 137.5$. Wtedy $E(Z) = 356.25$ (euro) jest oczekiwanym zarobkiem pośrednika.

Przykład – 6.15. W celu zbadania dużej populacji osób, podzielono ją na grupy, a następnie pobrano od każdej osoby krew oraz przeprowadzano analizę łączną dla poszczególnych grup, wykonując odpowiedni test na próbkach powstałych przez zmieszanie krwi osób należących do tej samej grupy. Gdy w pewnej grupie wykryto wirus chorobowy, przeprowadzano odrębną analizę dla każdej osoby z tej grupy. Założmy, że liczebność populacji wynosi N , liczebność grup wynosi n , zaś k niech będzie liczbą grup (oczywiście $N = nk$).

Zakładamy też, że prawdopodobieństwo tego, że dany człowiek jest zarażony interesującym nas wirusem wynosi p oraz że obecność wirusa u danej osoby jest niezależna od jego obecności u innych osób.

Na przykład. $N = 1000$, $k = 20$, $n = 50$, $p = 0.01$.

1. Ile analiz będzie trzeba przeprowadzić?

2. Dobrać wielkość grupy n , tak aby liczba wszystkich (bardzo kosztownych) analiz była, w pewnym sensie, minimalna.

Określamy X :

X – liczba wszystkich potrzebnych analiz.

Bezpośrednie wyznaczenie rozkładu X jest trudne.

$$X = X_1 + X_2 + \dots + X_k.$$

X_1, X_2, \dots, X_k są niezależne i mają taki sam rozkład. Wystarczy znaleźć rozkład, na przykład, X_1 .

X_1 przyjmuje dwie wartości:

1. $X_1 = 1$ – wszystkie osoby w grupie pierwszej są zdrowe.

2. $X_1 = n + 1$ – w przeciwnym przypadku.

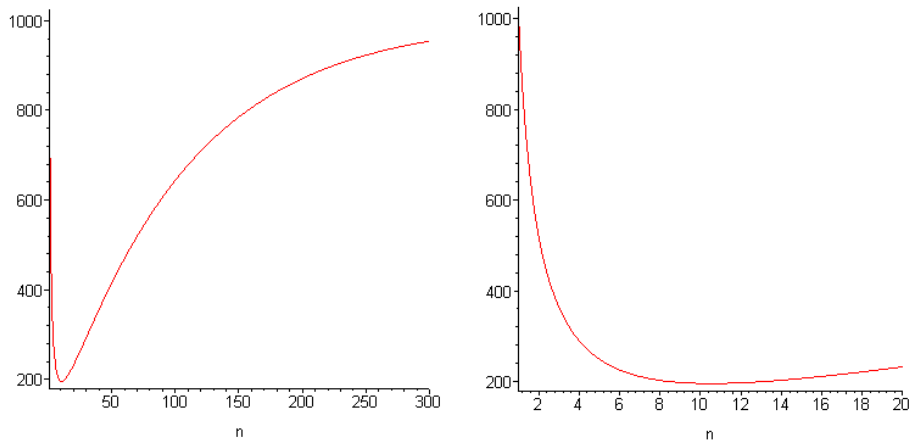
$$P(X_1 = 1) = (1 - p)^n, \quad P(X_1 = n + 1) = 1 - (1 - p)^n.$$

$$E(X_1) = 1 \cdot (1 - p)^n + (n + 1) \cdot (1 - (1 - p)^n) = n + 1 - n(1 - p)^n.$$

$$E(X) = E(X_1) + E(X_2) + \dots + E(X_k) = k \cdot E(X_1).$$

$$E(X) = k \cdot (n + 1 - n(1 - p)^n) = \frac{N}{n} (n + 1 - n(1 - p)^n).$$

Na przykład. $N = 1000$, $k = 20$, $n = 50$, $p = 0.01$. Wtedy $E(X) \approx 414.99$.



Optymalizacja. Nadzieja matematyczna $E(X)$ jako funkcja wielkości grupy.

Wtedy $n = 10$ – wartość optymalna oraz $E(X) \approx 195.68$.

Pytanie. Czy liczba analiz może przekroczyć 200? 250? 300?

Wspomnieliśmy już poprzednio, że nadzieja matematyczna może nie istnieć.

Przykład – 6.16. Niech X będzie zmienną losową o rozkładzie $U(0, 1)$. Niech $W = 1/X$. Łatwo wyznaczyć f_W , gęstość W (ćwiczenie). Mianowicie:

$$f_W(w) = \begin{cases} 0, & \text{dla } w < 1 \\ \frac{1}{w^2}, & \text{dla } 1 < w. \end{cases}$$

$$E(W) = \int_{\mathbb{R}} w f_W(w) dw = \int_0^{\infty} \frac{1}{w} dw = \infty.$$

Twierdzenie – 6.17 (Nadzieja iloczynu). Niech X, Y będą **niezależnymi** zmiennymi losowymi określonymi na tej samej przestrzeni probabilistycznej (Ω, Σ, P) .

Jeżeli $X \geq 0, Y \geq 0$, lub $E(X), E(Y) \in \mathbb{R}$, to

$$E(XY) = E(X)E(Y).$$

Dowód. W dowodzie skorzystamy z twierdzenia Fubniego oraz z twierdzeń 5.24 i 6.8. W pierwszym przypadku mamy

$$\begin{aligned} E(XY) &= \int_{[0, \infty)^2} xy dP_{(X,Y)}(x, y) = \int_{[0, \infty)^2} xy d(P_X \times P_Y)(x, y) = \\ &= \int_{[0, \infty)} x dP_X(x) \int_{[0, \infty)} y dP_Y(y) = E(X)E(Y). \end{aligned}$$

Można było stosować twierdzenie Fubniego, ponieważ – jak wiemy – zachodzi ono dla wszystkich funkcji mierzalnych nieujemnych.

Założmy teraz, że $E(|X|) < \infty, E(|Y|) < \infty$. Stosujemy udowodniony powyżej wzór dla funkcji nieujemnych $|X|$ oraz $|Y|$ i mamy

$$\int_{\mathbb{R}^2} |xy| dP_{(X,Y)}(x, y) = E(|XY|) = E(|X|)E(|Y|) < \infty.$$

Ale to oznacza, że znowu możemy zastosować twierdzenie Fubniego i podobnie jak poprzednio

$$\begin{aligned} E(XY) &= \int_{\mathbb{R} \times \mathbb{R}} xy dP_{(X,Y)}(x, y) = \int_{\mathbb{R} \times \mathbb{R}} xy d(P_X \times P_Y)(x, y) = \\ &= \int_{\mathbb{R}} x dP_X(x) \int_{\mathbb{R}} y dP_Y(y) = E(X)E(Y). \end{aligned}$$

□

6.2 Wariancja i odchylenie standardowe

Nadzieja matematyczna przynosi informację o tendencji centralnej rozkładu zmiennej losowej. Interesuje nas także informacja jak daleko od nadziei mogą znajdować się wartości tej zmiennej (jak duży jest rozrzut). Najważniejszą miarą rozrzutu jest wariancja oraz jej pierwiastek – odchylenie standardowe. Zaczniemy jednak od definicji momentów.

Niech X będzie zmienną losową określoną na przestrzeni probabilistycznej (Ω, Σ, P) . Zakładamy, że $m = E(X) \in \mathbb{R}$. Niech $k \geq 1$

Definicja – 6.18. $E(X^k)$ – moment rzędu k .

$E((X - m)^k)$ – centralny moment rzędu k .

$E(|X|^k)$ – moment bezwzględny rzędu k .

$E(|X - m|^k)$ – centralny bezwzględny moment rzędu k .

Twierdzenie – 6.19. Niech $l \geq k$. Jeżeli $E(X^l) \in \mathbb{R}$ to $E(X^k) \in \mathbb{R}$.

Dowód. $E(|X^k|) = \int_{\Omega} |X^k| dP = \int_{|X| < 1} |X^k| dP + \int_{|X| \geq 1} |X^k| dP \leq \int_{|X| < 1} 1 dP + \int_{|X| \geq 1} |X^l| dP \leq 1 + \int_{\Omega} |X^l| dP = 1 + E(|X^l|)$. \square

Zakładamy, że zmienna losowa X ma skończoną nadzieję matematyczną $m = E(X)$. Wtedy interpretujemy:

$X - m$ – odchylenie od średniej.

$E(|X - m|)$ – oczekiwana wartość odchylenia (średni błąd).

Definicja – 6.20. Centralny moment rzędu 2 nazywamy *wariancją*. Stosowane są różne oznaczenia.

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

Odchyleniem standardowym nazywamy pierwiastek z wariancji

$$\sigma(X) = \sqrt{D^2(X)} = \sqrt{\text{Var}(X)}.$$

Uwaga – 6.21.

$$D^2(X) = 0 \iff X = m.$$

Obliczanie momentów

$$D^2(X) = E((X - m)^2) = E(X^2) - 2mE(X) + E(m^2) = E(X^2) - m^2. \quad (6.1)$$

$$D^2(cX) = c^2 D^2(x), \text{ gdy } c \in \mathbb{R}. \quad (6.2)$$

$$E(X^k) = \int_{\mathbb{R}} x^k dP_X(x) = \begin{cases} \sum x_i^k p_i, & \text{w przypadku dyskretnym,} \\ \int_{\mathbb{R}} x^k f(x) dx, & \text{w przypadku ciągłym.} \end{cases} \quad (6.3)$$

$$E((X - m)^k) = \int_{\mathbb{R}} (x - m)^k dP_X(x) = \begin{cases} \sum (x_i - m)^k p_i, & \text{w przypadku dyskretnym,} \\ \int_{\mathbb{R}} (x - m)^k f(x) dx, & \text{w przypadku ciągłym.} \end{cases} \quad (6.4)$$

$$D^2(X) = \int_{\mathbb{R}} (x - m)^2 dP_X(x) = \begin{cases} \sum (x_i - m)^2 p_i, & \text{w przypadku dyskretnym,} \\ \int_{\mathbb{R}} (x - m)^2 f(x) dx, & \text{w przypadku ciągłym.} \end{cases} \quad (6.5)$$

$$D^2(X) = \int_{\mathbb{R}} x^2 dP_X(x) - m^2 = \begin{cases} \sum x_i^2 p_i - \left(\sum x_i p_i\right)^2, & \text{w przypadku dyskretnym,} \\ \int_{\mathbb{R}} x^2 f(x) dx - \left(\int_{\mathbb{R}} x f(x) dx\right)^2, & \text{w przypadku ciągłym.} \end{cases} \quad (6.6)$$

6.3 Kowariancja i korelacja

Iloczyn skalarny – przypomnienie. X – przestrzeń wektorowa, $\varphi : X \times X \rightarrow \mathbb{R}$ – iloczyn skalarny, to znaczy $\varphi(x, x) \geq 0$, $\varphi(x, x) = 0 \iff x = 0$, φ – dwuliniowe, φ – symetryczne.

Twierdzenie – 6.22 (Nierówność Cauchy’ego–Schwarza). *Dla każdych $x, y \in X$*

$$\begin{aligned}\varphi(x, y)^2 &\leq \varphi(x, x)\varphi(y, y). \\ \varphi(x, y)^2 = \varphi(x, x)\varphi(y, y) &\iff x, y \text{ liniowo zależne.}\end{aligned}$$

Niezerowe wektory x, y są liniowo zależne dokładnie wtedy, gdy istnieje takie $t \neq 0$, że $y = tx$. Wtedy zachodzi równość $\varphi(x, y) = t\varphi(x, x)$, co oznacza, że t ma taki sam znak jak $\varphi(x, y)$, co zapisujemy: $\operatorname{sgn} t = \operatorname{sgn} \varphi(x, y)$.

Przypominamy też, że $\sqrt{\varphi(x, x)}$ jest normą x .

Niech dane będą dwie zmienne losowe X, Y . Oznaczmy $m_X = E(X)$, $m_Y = E(Y)$.

Definicja – 6.23 (Kowariancja).

$$\operatorname{cov}(X, Y) = E((X - m_X)(Y - m_Y)) = E(XY) - m_X m_Y.$$

Zauważmy, że odwzorowanie

$$\varphi : (X, Y) \mapsto E(XY)$$

jest iloczynem skalarnym na przestrzeni $L^2(\Omega) = \{X : \Omega \rightarrow \mathbb{R} : E(X^2) < \infty\}$ (ćwiczenie).

W takim razie $\operatorname{cov}(X, Y)$ jest iloczynem skalarnym odchyłeń $X - m_X, Y - m_Y$: $\operatorname{cov}(X, Y) = \varphi(X - m_X, Y - m_Y)$, a wariancje są kwadratami ich norm. Z nierówności Cauchy’ego–Schwarza:

Wniosek – 6.24.

$$\operatorname{cov}(X, Y)^2 \leq D^2(X)D^2(Y),$$

przy czym:

$$\operatorname{cov}(X, Y)^2 = D^2(X)D^2(Y) \iff X - m_X, Y - m_Y \text{ są liniowo zależne.}$$

Wtedy, jeżeli X oraz Y nie są stałymi, to istnieje liczba t taka, że $Y - m_Y = t(X - m_X)$, a $\operatorname{sgn} t = \operatorname{sgn} \operatorname{cov}(X, Y)$.

Ogólnie: $\operatorname{cov}(X, Y)^2 = D^2(X)D^2(Y) \iff$ wartości (X, Y) zawarte są w pewnej prostej.

Definicja – 6.25 (Korelacja). Mówimy, że zmienne losowe X, Y są *niekorelowane* gdy $\operatorname{cov}(X, Y) = 0$.

Współczynnikiem korelacji zmiennych X, Y niebędących stałymi p.w. nazywać będziemy:

$$\rho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{D^2(X)}\sqrt{D^2(Y)}}.$$

Na podstawie wniosku 6.24 oraz twierdzenia 6.17 otrzymujemy

Wniosek – 6.26.

$$-1 \leq \rho(X, Y) \leq 1.$$

$$|\rho(X, Y)| = 1 \iff \text{wartości } (X, Y) \text{ zawierają się w pewnej prostej.}$$

Jeżeli X, Y są niezależne, to $\operatorname{cov}(X, Y) = 0$.

Uwaga – 6.27 (Związek między wariancją sumy i kowariancją).

$$\begin{aligned}D^2(X + Y) &= E((X + Y - (m_X + m_Y))^2) \\ &= E(((X - m_X) + (Y - m_Y))^2) = D^2(X) + D^2(Y) + 2\operatorname{cov}(X, Y).\end{aligned}$$

Z powyższych wniosków oraz uwagi wynikają ważne następujące:

Twierdzenie – 6.28 (Wariancja sumy zmiennych niezależnych). *Jeżeli X, Y są niezależne, to*

$$D^2(X + Y) = D^2(X) + D^2(Y).$$

Twierdzenie – 6.29. *Twierdzenie Jeżeli X_1, X_2, \dots, X_n , są niezależne, to*

$$D^2(X_1 + \dots + X_n) = D^2(X_1) + \dots + D^2(X_n).$$

Wniosek – 6.30 (Nadzieja i wariancja sumy oraz średniej). *Niech X_1, X_2, \dots, X_n , będą niezależne i mają wspólną nadzieję matematyczną m oraz wariancję σ^2 . Niech*

$$S_n = X_1 + \dots + X_n, \quad \bar{X}_n = \frac{X_1 + \dots + X_n}{n}.$$

Wtedy:

$$\begin{aligned} E(S_n) &= nm, & D^2(S_n) &= n\sigma^2, & \sigma(S_n) &= \sqrt{n}\sigma. \\ E(\bar{X}_n) &= m, & D^2(\bar{X}_n) &= \frac{\sigma^2}{n}, & \sigma(\bar{X}_n) &= \frac{\sigma}{\sqrt{n}}. \end{aligned}$$

6.4 Pytania

Pytanie 6.1. Podać przykłady zmiennych losowych o rozkładzie dyskretnym oraz o rozkładzie ciągłym, dla których nadzieja matematyczna nie jest skończona. (ROZWIĄZANIE)

Pytanie 6.2. Podać przykłady zmiennych losowych o rozkładzie dyskretnym oraz o rozkładzie ciągłym, dla których nadzieja matematyczna jest skończona, ale wariancja jest nieskończona. (ROZWIĄZANIE)

Pytanie 6.3. Dane są niezależne zmienne losowe o rozkładzie $U(0, 1)$ każda. Obliczyć $E(\min(X, Y))$. (ROZWIĄZANIE)

Pytanie 6.4. Przeprowadzić dowód twierdzenia 6.17 w przypadku, gdy (X, Y) ma rozkład dyskretny na zbiorze skończonym. (ROZWIĄZANIE)

Pytanie 6.5. Dana jest taka funkcja $f : [a, b] \rightarrow \mathbb{R}$, że $J = \int_a^b f(x) dx < \infty$. Wskazać taką zmienną losową X , że $J = E(X)$. (ROZWIĄZANIE)

Pytanie 6.6. Zmienna losowa X ma rozkład $P(X = i) = 2^{-i}$ dla $i = 1, 2, 3, \dots$, a funkcja g jest dana wzorem; $g(i) = (-1)^{i+1} \frac{2^i}{i}$. Czy istnieje $E(g(X))$? Czy jest zbieżny szereg $\sum_{i=1}^{\infty} g(i)P(X = i)$? (ROZWIĄZANIE)

Pytanie 6.7. Pokazać, że

$$\varphi : (L^2(\Omega))^2 \ni (X, Y) \mapsto E(XY) \in \mathbb{R}$$

jest iloczynem skalarnym na przestrzeni $L^2(\Omega) = \{X : \Omega \rightarrow \mathbb{R} : E(X^2) < \infty\}$. (ROZWIĄZANIE)

Rozdział 7

Nierówność Czebyszewa i prawa wielkich liczb

Znajomość momentów pozwala oszacować prawdopodobieństwo tego, że zmienna losowa przyjmuje wartość w określonym zbiorze. W szczególności znajomość wariancji pozwala oszacować z góry prawdopodobieństwo tak zwanych ogonów, to znaczy zbiorów postaci $\{|X - E(X)| \geq \varepsilon\}$.

7.1 Nierówność Czebyszewa

Twierdzenie – 7.1. Niech $X : \Omega \rightarrow \mathbb{R}$ będzie zmienną losową, $\varepsilon > 0$.

1. Niech $k \geq 1$, $X \geq 0$. Wtedy $P(X \geq \varepsilon) \leq \frac{E(X^k)}{\varepsilon^k}$.
2. Niech $m = E(X) \in \mathbb{R}$. Wtedy $P(|X - m| \geq \varepsilon) \leq \frac{D^2(X)}{\varepsilon^2}$.
3. Niech $m = E(X) \in \mathbb{R}$, $\sigma = \sqrt{D^2(X)} > 0$, $c > 0$. Wtedy

$$P(|X - m| \geq c\sigma) \leq \frac{1}{c^2}.$$

Dowód. Dowód pierwszej nierówności:

$$E(X^k) = \int_{\Omega} X^k dP \geq \int_{X \geq \varepsilon} X^k dP \geq \int_{X \geq \varepsilon} \varepsilon^k dP = \varepsilon^k P(X \geq \varepsilon). \quad (7.1)$$

Drugą nierówność otrzymujemy stosując nierówność pierwszą dla zmiennej losowej $|X - m|$ oraz $k = 2$. Trzecia nierówność wynika z drugiej, gdy $\varepsilon = c\sigma$. □

Dlatego też

Wniosek – 7.2 (Reguła 3 σ).

$$P(|X - m| \geq 3\sigma) \leq \frac{1}{9}.$$

Zauważmy, że w dowodzie nierówności Czebyszewa, wzór (7.1), wykonaliśmy dwa szacowanie, które w wielu przypadkach mogą być bardzo niedokładne (gdy ε jest duże – pierwsze, gdy ε jest małe – drugie). Dlatego też nierówność tę oraz płynące z niej wnioski warto traktować jako niezbyt precyzyjne. W przypadku, gdy znane są rozkłady interesujących nas zmiennych rezultaty otrzymane za pomocą nierówności Czebyszewa mogą być istotnie poprawione. Jednak, gdy nie znamy rozkładów, nierówność Czebyszewa może być bardzo pomocna.

Nierówność Czebyszewa służy do szacowania prawdopodobieństw na podstawie znajomości samych momentów, najczęściej nadziei oraz wariancji:

Przykład – 7.3. Załóżmy, że zmienna losowa o rozkładzie ciągłym ma parametry $m = E(X) = 100$, $\sigma = \sqrt{D^2(X)} = 2$. Szacujemy prawdopodobieństwo tego, że:

(a) $X \geq 110$. $P(X \geq 110) = P(X - m \geq 10) \leq P(|X - m| \geq 10) \leq \frac{\sigma^2}{10^2} = 0.04$.

$$(b) X \leq 105. \quad P(X \leq 105) = 1 - P(X \geq 105) \geq 1 - \frac{\sigma^2}{5^2} = 1 - 4/25 = 0.84.$$

(c) Szukamy takiej liczby M , aby $P(X < M) \geq 0.99$.

Wiemy, że: $P(X < M) = 1 - P(X \geq M) = 1 - P(X - m \geq M - m) \geq 1 - P(|X - m| \geq M - m) \geq 1 - \frac{\sigma^2}{(M-m)^2}$.
Wystarczy więc znaleźć takie M , że $1 - \frac{\sigma^2}{(M-m)^2} \geq 0.99$. Czyli, że $\frac{\sigma^2}{(M-m)^2} \leq 0.01$. Stąd $M \geq m + \frac{\sigma}{\sqrt{0.01}} = 120$.

Przykład – 7.4 (Kontynuacja zadania o malinach, przykład 6.14). Można obliczyć, że przy optymalnej wartości a (137.5) odchylenie standardowe zmiennej Z , $\sigma \approx 89.92184106$. A więc zgodnie z regułą 3σ zysk pośrednika zawiera się w przedziale (86.4844768, 626.0155232) z prawdopodobieństwem większym niż $\frac{8}{9}$ (ćwiczenie). W istocie wynik ten jest wysoce niedokładny, gdyż, jak łatwo zauważyć, dla tej wartości a zysk zawiera się w przedziale (112.5, 412.5).

Przykład – 7.5 (Kontynuacja przykładu 6.15). W celu zbadania dużej populacji osób, podzielono ją na grupy, a następnie pobrano od każdej osoby krew oraz przeprowadzano analizę łączną dla poszczególnych grup, wykonując odpowiedni test na próbkach powstałych przez zmieszanie krwi osób należących do tej samej grupy. Gdy w pewnej grupie wykryto wirus chorobowy, przeprowadzano odrębną analizę dla każdej osoby z tej grupy. Załóżmy, że liczebność populacji wynosi N , liczność grup wynosi n , zaś k niech będzie liczbą grup (oczywiście $N = nk$). Zakładamy też, że prawdopodobieństwo tego, że dany człowiek jest zarażony interesującym nas wirusem wynosi p oraz że obecność wirusa u danej osoby jest niezależna od jego obecności u innych osób.

Wiemy, że: Dla $N = 1000$, oraz $p = 0.01$ optymalnymi ze względu na średnią liczbę analiz parametrami są: $n = 10$, $k = 100$. Wtedy oczekiwana liczba analiz wynosi $m = E(X) = 195.68$.

Pytanie: Czy liczba analiz może przekroczyć 200? 250? 300?

Policzmy na przykład $P(X \geq 300)$. Aby skorzystać z nierówności Czebyszewa musimy policzyć wariancję X .

Pamiętamy, że $X = X_1 + \dots + X_k$, gdzie X_i są niezależne o takim samym rozkładzie dwupunktowym $P(X_1 = 1) = (1 - p)^n$, $P(X_1 = n + 1) = 1 - (1 - p)^n$.

Z niezależności:

$$D^2(X) = D^2(X_1) + \dots + D^2(X_k).$$

Można policzyć (w przybliżeniu): $D^2(X_i) \approx 8.65$ oraz $D^2(X) \approx 865$. Podobnie jak w przykładzie poprzednim: $P(X \geq 300) \leq \frac{D^2(X)}{(300-m)^2} \approx 0.0795$.

Wynik ten można znacznie polepszyć. W istocie, $P(X \geq 300)$ jest dużo mniejsze. Odpowiednie oszacowanie będzie możliwe, gdy poznamy szczególny charakter rozkładu zmiennej X .

7.2 Słabe prawo wielkich liczb

Założenia Zakładamy, że: zmienne losowe X_1, X_2, X_3, \dots są określone na tej samej przestrzeni probabilistycznej (Ω, Σ, P) , są niezależne, mają skończone nadzieje matematyczne oraz skończone i niezerowe wariancje. Oznaczmy:

$$S_n = X_1 + \dots + X_n, \quad \bar{X}_n = \frac{X_1 + \dots + X_n}{n}.$$

Twierdzenie – 7.6 (Słabe prawo wielkich liczb). *Przy powyższych założeniach:*

1. Jeżeli istnieje takie $M \in \mathbb{R}$, że $D^2(X_i) \leq M$ dla wszystkich i , to dla każdego $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{S_n - E(S_n)}{n} \right| \geq \varepsilon \right) = 0.$$

2. Zakładamy dodatkowo, że wszystkie nadzieje matematyczne są sobie równe i równe m . Wtedy dla każdego $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{S_n}{n} - m \right| \geq \varepsilon \right) = 0.$$

3. Załóżmy, że każda zmienna losowa ma taki sam rozkład dwupunktowy, $P(X_i = 0) = 1 - p$, $P(X_i = 1) = p$. Wtedy dla każdego $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{S_n}{n} - p \right| \geq \varepsilon \right) = 0.$$

Dowód. Ad 1. Z niezależności zmiennych losowych mamy

$$D^2(S_n) = D^2(X_1) + \dots + D^2(X_n) \leq nM.$$

Stosując nierówność Czebyszewa do dla zmiennej losowej S_n , dostajemy

$$0 \leq P\left(\frac{|S_n - E(S_n)|}{n} \geq \varepsilon\right) = P(|S_n - E(S_n)| \geq n\varepsilon) \leq \frac{D^2(S_n)}{(n\varepsilon)^2} \leq \frac{M}{n\varepsilon^2},$$

co daje tezę (twierdzenie o trzech ciągach).

Ad 2. Wynika natychmiast z punktu 1, gdyż $E(S_n) = nm$.

Ad 3. Wynika natychmiast z punktu 2, gdyż $m = E(X_i) = p$, $D^2(X_i) = p(1-p)$. □

Interpretacja.

Punkt 2. Średnia po przestrzeni = średniej po czasie.

Punkt 3. Aksjomatyczna definicja prawdopodobieństwa jest zgodna z dawniej używaną definicją częstościową.

Mocne prawo wielkich liczb. Sprawę interpretacji omówimy jeszcze dokładniej w rozdziale 10, gdzie udowodnimy później tak zwane mocne prawo wielkich liczb (w dwóch wersjach), które – jak sama nazwa wskazuje – jest wynikiem mocniejszym niż udowodnione przed chwilą słabe prawo wielkich liczb. Jego dowód będzie oparty na nierówności Kołmogorowa, która, w pewnym szczególnym przypadku, wzmacnia nierówność Czebyszewa. Przytoczmy już teraz jedną z wersji mocnego prawa wielkich liczb wzmacniającą istotnie punkt 2. w poprzednim twierdzeniu.

Twierdzenie – 10.16 (Mocne prawo wielkich liczb).

Niech X_1, X_2, X_3, \dots będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie i skończonej wartości oczekiwanej m . Niech $S_n = X_1 + X_2 + \dots + X_n$.

Wtedy

$$P\left(\left\{\omega : \lim_{n \rightarrow \infty} \frac{S_n(\omega)}{n} \rightarrow m\right\}\right) = 1.$$

7.3 Pytania

Pytanie 7.1. Wykazać, że dla zmiennej losowej X mającej skończoną nadzieję matematyczną m i dla każdego dodatniego $\varepsilon > 0$

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

ROZWIĄZANIE

Pytanie 7.2. Niech $m = E(X)$, $\sigma^2 = D^2(X)$, $a < m < b$. Oszacować z góry $P(X \leq a)$ i z dołu $P(X < b)$.

ROZWIĄZANIE

Pytanie 7.3. Niech X ma rozkład o gęstości f : $f(x) = 0$ dla $x < 0$ oraz $f(x) = e^{-x}$ dla $x \geq 0$. Dla $\varepsilon > 0$ obliczyć $P(X \geq \varepsilon)$ i porównać otrzymany wynik z oszacowaniem wynikającym z nierówności Czebyszewa. ROZWIĄZANIE

Pytanie 7.4. Sprawdzić, że w przypadku rozkładu jednostajnego reguła 3σ się trywializuje. ROZWIĄZANIE

Pytanie 7.5. Niech $f : [0, 1] \rightarrow \mathbb{R}$ będzie funkcją ciągłą, X_1, X_2, X_3, \dots będzie ciągiem niezależnych zmiennych losowych o rozkładzie $B(1, p)$ każda, $S_n = X_1 + \dots + X_n$. Wykazać, że

$$\lim_{n \rightarrow \infty} E\left(f\left(\frac{S_n}{n}\right) - f(p)\right) = 0,$$

przy czym zbieżność ta jest jednostajna względem p . ROZWIĄZANIE

Pytanie 7.6. Wskazać ciąg wielomianów, które jednostajnie aproksymują daną funkcję ciągłą f na przedziale $[0, 1]$.

ROZWIĄZANIE

Rozdział 8

Wybrane rozkłady prawdopodobieństwa

Ważniejsze rozkłady dyskretne i ciągłe

1. Rozkład jednopunktowy, δ -Diraca, δ_c .
2. Rozkład jednostajny dyskretny.
3. Rozkład jednostajny ciągły.
4. Rozkład Bernoulliego (dwupunktowy), $(0, 1, p)$ ($B(1, p)$).
5. Rozkład dwumianowy, $B(n, p)$.
6. Rozkład Poissona, P_λ .
7. Rozkład geometryczny G_p .
8. Rozkład hipergeometryczny.
9. Rozkład Pascala.
10. Rozkład wykładniczy, E_λ .
11. Rozkład Erlanga.
12. Rozkład normalny, $N(m, \sigma)$.

List of probability distributions

Rozkład jednopunktowy, δ_c . Dla ustalonego punktu $c \in \mathbb{R}$ jest to rozkład Q skupiony w tym punkcie, to znaczy: $Q(c) = 1$.

Jeżeli $X : \Omega \rightarrow \mathbb{R}$ jest zmienną losową, to

$$P_X = \delta_c \iff P(X = c) = 1.$$

$$E(X) = c, \quad D^2(X) = 0.$$

Rozkład jednostajny dyskretny na zbiorze skończonym K . Jest to rozkład zadany na zbiorze skończonym $K = \{x_1, \dots, x_n\}$ jako $Q(x_i) = \frac{1}{n}$ dla $i = 1, \dots, n$.

Przykładowo, zmienna X opisująca liczbę oczek jaka wypadła na kostce symetrycznej ma rozkład jednostajny dyskretny.

$$E(X) = \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i, \quad D^2(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Rozkład jednostajny, $U(a, b)$. Niech $a, b \in \mathbb{R}$, $a < b$. Jest to rozkład o gęstości $\frac{1}{b-a}I_{[a,b]}$.

Niewiele zjawisk podlega rozkładowi jednostajnemu.

Komputer „potrafi” generować liczby według rozkładu jednostajnego, co z kolei służy do generowania liczb z zadanego z góry innego rozkładu (dyskretnego lub ciągłego), patrz punkt 12.1.

$$E(X) = \frac{a+b}{2}, \quad D^2(X) = \frac{(b-a)^2}{12}.$$

Dowód.

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

$$D^2(X) = E(X^2) - E(X)^2 = \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \dots = \frac{(b-a)^2}{12}. \quad \square$$

Rozkład Bernoulliego, dwupunktowy – $(0, 1, p)$ ($B(1, p)$). Niech $0 < p < 1$. Jest to taki rozkład Q , że $Q(0) = 1 - p$, $Q(1) = p$.

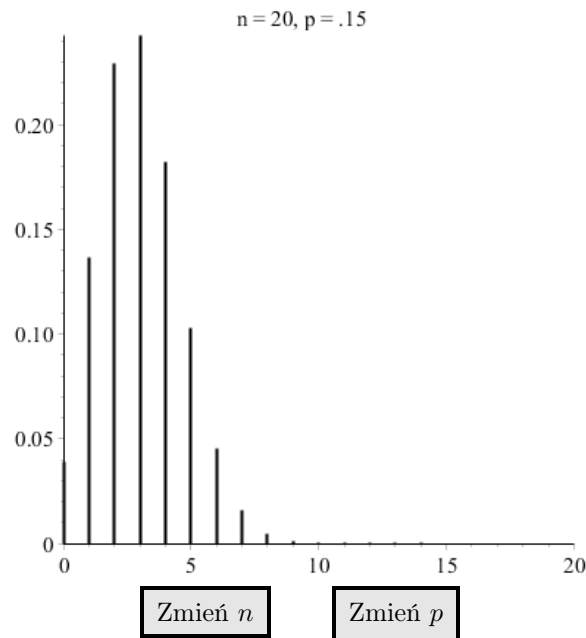
Gdy X jest wynikiem doświadczenia, które ma dokładnie dwa możliwe zakończenia (porażka – 0, lub sukces – 1), to X ma rozkład dwupunktowy.

$$E(X) = p, \quad D^2(X) = (1-p)p.$$

8.1 Rozkład dwumianowy – $B(n, p)$

Rozkład Q nazywamy rozkładem dwumianowym, jeżeli istnieją liczby $n > 0$ oraz p i q takie, że $0 < p, q < 1$, $p + q = 1$ oraz zachodzi równość:

$$Q(k) = \binom{n}{k} p^k q^{n-k} \quad \text{dla } k = 0, 1, \dots, n.$$



Twierdzenie – 8.1. Niech X_1, \dots, X_n będą niezależnymi zmiennymi losowymi o takim samym rozkładzie dwupunktowym $(0, 1, p)$. Wtedy suma:

$$S_n = X_1 + \dots + X_n$$

ma rozkład dwumianowy $B(n, p)$.

Dowód. Ustalmy $k, 0 \leq k \leq n$. Zdarzenie $\{S_n = k\}$ jest sumą rozłącznych zdarzeń polegających na tym, że dokładnie k spośród zmiennych losowych X_1, \dots, X_n przyjmuje wartość 1, a więc pozostałe $n - k$ zmiennych przyjmuje wartość 0. Niech A_{i_1, \dots, i_k} będzie jednym z takich zdarzeń, gdzie i_1, \dots, i_k oznaczają numery tych zmiennych, które przyjmują wartość 1. Z kolei każde zdarzenie A_{i_1, \dots, i_k} jest iloczynem n zdarzeń postaci $\{X_j = \varepsilon_j\}$, gdzie $\varepsilon_j = 1$ lub $\varepsilon_j = 0$, a prawdopodobieństwa tych zdarzeń są równe odpowiednio p i q . Z niezależności zmiennych X_1, \dots, X_n wynika, że:

$$P(A_{i_1, \dots, i_k}) = p^k q^{n-k}.$$

Ponieważ wskaźniki i_1, \dots, i_k można wybrać na $\binom{n}{k}$ sposobów, więc:

$$\begin{aligned} P(S_n = k) &= P\left(\bigcup_{i_1, \dots, i_k} A_{i_1, \dots, i_k}\right) = \sum_{i_1, \dots, i_k} P(A_{i_1, \dots, i_k}) \\ &= \sum_{i_1, \dots, i_k} p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}. \end{aligned} \quad \square$$

Dla zmiennej X o rozkładzie dwumianowym $B(n, p)$ zachodzi

$$E(X) = np, \quad D^2(X) = npq.$$

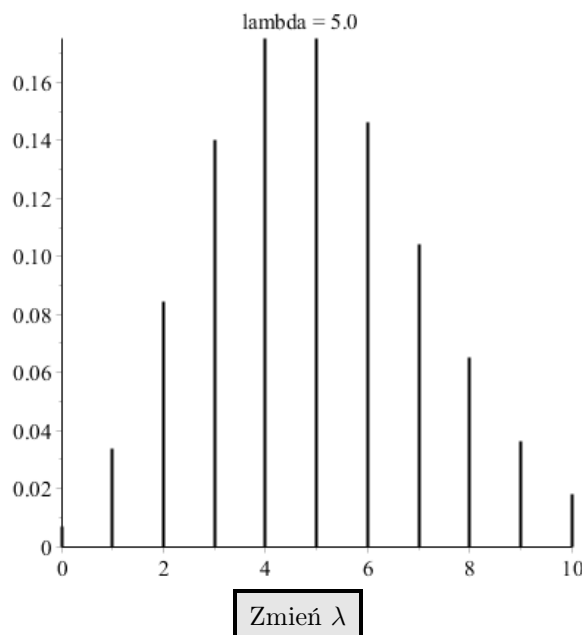
Dowód. Jest to wniosek z powyższego twierdzenia. □

Losowanie ze zwracaniem Przypuśćmy, że pewna populacja składa się z N elementów. Niech N_0 elementów tej populacji ma pewną własność, powiedzmy własność W . Niech $p = \frac{N_0}{N}$. Losujemy ze zwracaniem n elementów i oznaczamy przez X liczbę tych spośród nich, które mają własność W . Widać, że zmienna losowa X ma rozkład dwumianowy $B(n, p)$.

8.2 Rozkład Poissona – P_λ

Rozkład Q jest rozkładem Poissona, jeżeli istnieje taka liczba $\lambda > 0$, że:

$$Q(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{dla } k = 0, 1, 2, \dots$$



$$E(X) = \lambda, \quad D^2(X) = \lambda.$$

Dowód.

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^{k-1}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned}$$

Podobnie dla wariancji (ćwiczenie). □

Wiele zjawisk podlega rozkładowi Poissona. Zgodność taka została zaobserwowana w wielu konkretnych sytuacjach praktycznych.

Przykład – 8.2. Już ponad 100 lat temu próbowano ustalić ile wypadków śmiertelnych spowodowanych przez kontakt z koniem zachodzi w jednym korpusie armii pruskiej w ciągu jednego roku. Uzyskano odpowiednie dane z 10 korpusów za okres 20 lat. Okazało się, że: w 109 przypadkach nie zaszło żadne takie zdarzenie, w 65 przypadkach w korpusie zdarzył się w ciągu roku jeden śmiertelny wypadek, dwa śmiertelne wypadki w ciągu roku miały miejsce 22 razy, trzy wypadki – 3 razy, a raz w pewnym korpusie były 4 wypadki śmiertelne. W poniższej tabelce zawierającej te dane wyznaczono dodatkowo średnią ważoną liczby wypadków śmiertelnych w korpusie w ciągu jednego roku: $\lambda = 0.61$. Ostatni wiersz tabelki zawiera wartości rozkładu Poissona o parametrze równym λ w punktach 0, 1, 2, 3, 4. Widać, że są one bardzo zbliżone do obliczonych wcześniej częstości f_0, \dots, f_4 . Można na tej podstawie przypuszczać, że zmienna losowa X zdefiniowana jako liczba wypadków śmiertelnych w ciągu roku w jednym korpusie miała rozkład Poissona. Odpowiednie testy statystyczne zdecydowanie potwierdzają to przypuszczenie.

| k | 0 | 1 | 2 | 3 | 4 | SUMA |
|-----------------------------|-------|-------|-------|-------|-------|---------|
| n_k | 109 | 65 | 22 | 3 | 1 | 200 |
| $f_k = \frac{n_k}{200}$ | 0,545 | 0,325 | 0,11 | 0,015 | 0,005 | 1 |
| $k \cdot f_k$ | 0 | 0,325 | 0,22 | 0,045 | 0,02 | 0,61 |
| $P(X = k) = P_{\lambda}(k)$ | 0,543 | 0,331 | 0,101 | 0,021 | 0,003 | 0,99957 |

Dane o liczbie śmiertelnych wypadków spowodowanych przez konia w 10 korpusach armii pruskiej w ciągu 20 lat (Bortkiewicz), patrz [8], strona 75.

Wniosek. X – liczba śmiertelnych wypadków w ciągu jednego roku w korpusie armii pruskiej ma rozkład Poissona P_{λ} dla $\lambda = 0.61$.

Następujące twierdzenie mówi o tym, że rozkład Poissona jest w pewnym sensie granicą rozkładów dwumianowych. W szczególności, gdy mamy do czynienia z dużą ($n > 100$) liczbą niezależnych prób Bernoulliego, z jednakowym, małym ($p < 0.1$) prawdopodobieństwem sukcesu każda, to liczba wszystkich sukcesów ma niemal dokładnie rozkład Poissona z parametrem $\lambda = np$. Istnieją dość dokładne oszacowania błędu, jaki popełniamy przybliżając rozkład dwumianowy rozkładem Poissona.

Twierdzenie – 8.3. Niech liczby $p_n > 0$ tworzą taki ciąg, że: $\lim_{n \rightarrow \infty} np_n = \lambda > 0$ oraz niech k będzie nieujemną liczbą naturalną. Wtedy:

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

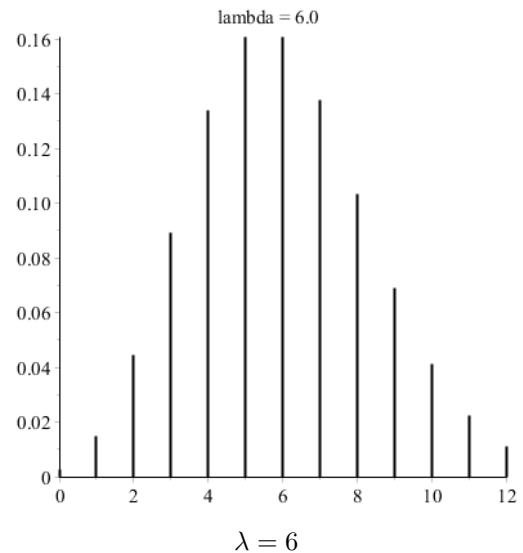
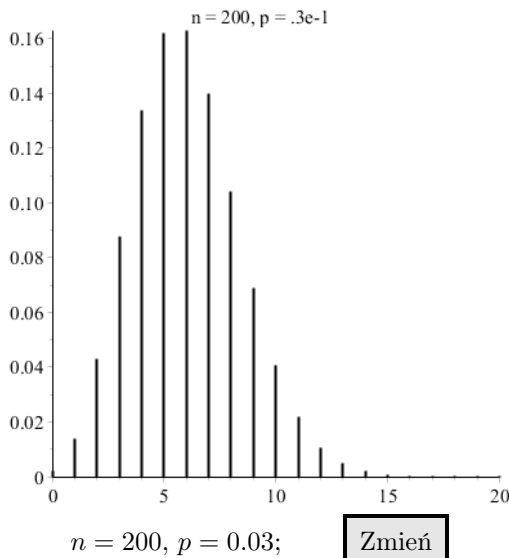
Dowód. Oznaczając $\lambda_n = np_n$, mamy równość

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{\lambda_n^k}{k!} \cdot \frac{n(n-1) \cdots (n-k+1)}{n^k} \cdot \left(1 - \frac{\lambda_n}{n}\right)^n \cdot \left(1 - \frac{\lambda_n}{n}\right)^{-k}.$$

Ponieważ k jest ustalone, to ostatni czynnik zmierza do 1. Drugi czynnik jest równy $1 \cdot (1 - \frac{1}{n}) \cdot \dots \cdot (1 - \frac{k-1}{n})$, więc też zmierza do 1. Istotne są czynniki pierwszy oraz trzeci i zmierzają one odpowiednio do $\frac{\lambda^k}{k!}$ oraz $e^{-\lambda}$. □

Warto porównać obydwa rozkłady dla wybranych parametrów.

| k | $n = 100,$ | $p = 0,01$ | $n = 50,$ | $p = 0,1$ | $n = 100,$ | $p = 0,1$ |
|-----|------------------|---------------------|------------------|---------------------|------------------|---------------------|
| | rozkład dwum. | rozkład Poissona | rozkład dwum. | rozkład Poissona | rozkład dwum. | rozkład Poissona |
| 0 | 0,3660 | 0,3679 | 0,0052 | 0,0067 | 0,0000 | 0,0000 |
| 1 | 0,3697 | 0,3679 | 0,0286 | 0,0337 | 0,0003 | 0,0005 |
| 2 | 0,1849 | 0,1839 | 0,0779 | 0,0842 | 0,0016 | 0,0023 |
| 3 | 0,0610 | 0,0613 | 0,1386 | 0,1404 | 0,0059 | 0,0076 |
| 4 | 0,0149 | 0,0153 | 0,1809 | 0,1755 | 0,0159 | 0,0189 |
| 5 | 0,0029 | 0,0031 | 0,1849 | 0,1755 | 0,0339 | 0,0378 |
| 6 | 0,0005 | 0,0005 | 0,1541 | 0,1462 | 0,0596 | 0,0631 |
| 7 | 0,0001 | 0,0001 | 0,1076 | 0,1044 | 0,0889 | 0,0901 |
| 8 | 0,0000 | 0,0000 | 0,0643 | 0,0653 | 0,1148 | 0,1126 |
| 9 | 0,0000 | 0,0000 | 0,0333 | 0,0363 | 0,1304 | 0,1251 |
| 10 | 0,0000 | 0,0000 | 0,0152 | 0,0181 | 0,1319 | 0,1251 |
| 11 | 0,0000 | 0,0000 | 0,0061 | 0,0082 | 0,1199 | 0,1137 |
| 12 | 0,0000 | 0,0000 | 0,0022 | 0,0034 | 0,0988 | 0,0948 |
| 13 | 0,0000 | 0,0000 | 0,0007 | 0,0013 | 0,0743 | 0,0729 |
| 14 | 0,0000 | 0,0000 | 0,0002 | 0,0005 | 0,0513 | 0,0521 |
| 15 | 0,0000 | 0,0000 | 0,0001 | 0,0002 | 0,0327 | 0,0347 |



Twierdzenie – 8.4. Suma niezależnych zmiennych losowych o rozkładach Poissona ma rozkład Poissona.

Dowód. Niech X ma rozkład P_λ , a Y rozkład P_μ . Niech $k \geq 0$ będzie ustalone.

$$\begin{aligned}
 P(X + Y = k) &= P\left(\bigcup_{i+j=k} (X = i, Y = j)\right) = \sum_{i+j=k} P(X = i)P(Y = j) = \sum_{i=0}^k P(X = i)P(Y = k - i) \\
 &= \sum_{i=0}^k e^{-\lambda} \frac{\lambda^i}{i!} e^{-\mu} \frac{\mu^{k-i}}{(k-i)!} = e^{-(\lambda+\mu)} \sum_{i=0}^k \frac{\lambda^i \mu^{k-i}}{i!(k-i)!} = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^k}{k!}.
 \end{aligned}$$

□

Przykład – 8.5. W ostatnich pięciu latach zanotowano 7, 6, 5, 5, 6 przypadków utonięć w Wiśle (dane fikcyjne). Oblicz prawdopodobieństwo tego, że w nadchodzącym roku liczba X utonięć w Wiśle będzie:

- (a) równa 0, (b) większa od 6, (c) co najmniej dziesięć.

Z charakteru zjawiska wynika, że zmienna losowa X ma rozkład Poissona P_λ . Średnia liczba utonięć w roku wynosi $\frac{7+6+5+5+6}{5} = 5.8$. Jeżeli założymy, że w poprzednich latach rozkład utonięć podlegał temu samemu rozkładowi, czyli rozkładowi o wartości oczekiwanej λ , to można przyjąć (wyjaśni to dokładniej statystyka), że $\lambda = 5.8$. Z komputera (Excel, Maple, Mathematica i dużo więcej), lub z tablic (nie polecam) otrzymujemy wartości prawdopodobieństw $P_\lambda(k)$ oraz wartości dystrybuanty $F_\lambda(k)$ dla $k = 0, 1, 2, \dots$. Mamy więc dla $\lambda = 5.8$:

- (a) $P(X = 0) = P_\lambda(0) \approx 0.0030$ – niestety bardzo małe.
 (b) $P(X > 6) = 1 - P(X \leq 6) = 1 - F_\lambda(6) \approx 1 - 0.6384 = 0.3616$ – dość duże.
 (c) $P(X \geq 10) = P(X > 9) = 1 - P(X \leq 9) = 1 - F_\lambda(9) \approx 1 - 0.9292 = 0.0708$ – już niezbyt duże.

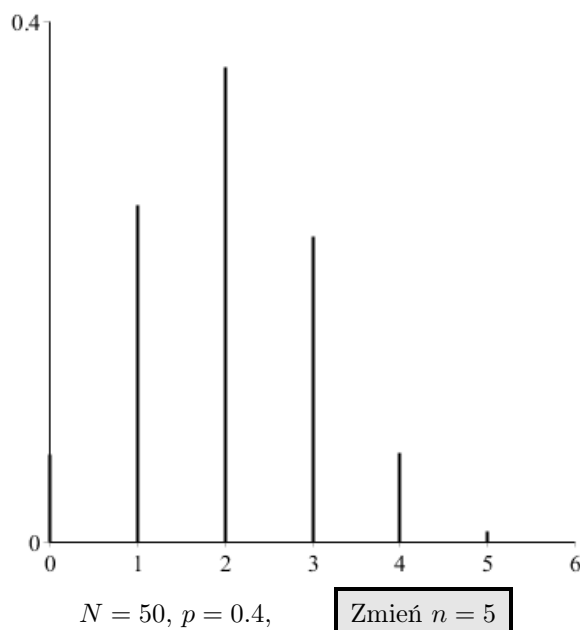
8.3 Rozkład hipergeometryczny

Rozkład Q nazywamy hipergeometrycznym, jeżeli istnieją liczby naturalne N , $N_0 \leq N$, $n \leq N$ takie, że dla każdego $k = 0, 1, 2, \dots, n$ zachodzi:

$$Q(k) = \frac{\binom{N_0}{k} \binom{N-N_0}{n-k}}{\binom{N}{n}}.$$

Oznaczając $p = \frac{N_0}{N}$ oraz $q = 1 - p$ otrzymujemy:

$$Q(k) = \frac{\binom{Np}{k} \binom{Nq}{n-k}}{\binom{N}{n}}.$$



Przypuśćmy, że pewna populacja składa się z N elementów, przy czym N_0 elementów ma własność W . Losujemy bez zwracania n elementów i oznaczamy przez X liczbę wylosowanych elementów mających własność W . Łatwo zauważyć, nawiązując do rozważań dotyczących losowania ze zwracaniem, że zmienna losowa X ma rozkład hipergeometryczny.

$$E(X) = np, \quad D^2(X) = npq \frac{N-n}{N-1}.$$

Dowód. Maple, Ćwiczenie 8.5. □

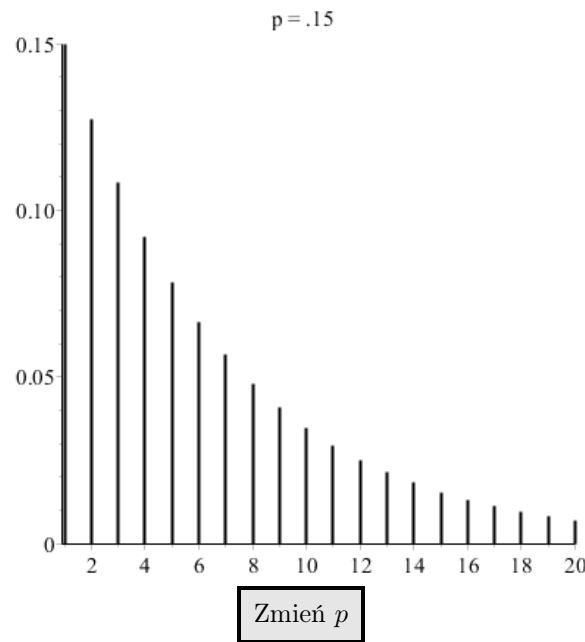
Przy losowaniu n elementów ze zwracaniem i przy losowaniu n elementów bez zwracania z populacji o liczebności N z wyróżnioną frakcją losujemy średnio **tyle samo** elementów z tych frakcji. Jednak przy losowaniu bez zwracania wariancja jest mniejsza.

Porównajmy odpowiednie rozkłady.

8.4 Rozkład geometryczny, G_p .

Rozkład Q jest rozkładem geometrycznym, jeżeli istnieją liczby p, q , $0 < p, q < 1$, $p + q = 1$ takie, że

$$Q(k) = q^{k-1}p, \quad \text{dla } k = 1, 2, 3, \dots$$



Twierdzenie – 8.6. Niech X_1, X_2, X_3, \dots będą niezależnymi zmiennymi losowymi o takim samym rozkładzie dwupunktowym $(0, 1, p)$. Wtedy funkcja

$$T = \min\{n \geq 1 : X_n = 1\},$$

nazywana czasem oczekiwania na pierwszy sukces w nieskończonym ciągu prób Bernoulliego, jest zmienną losową o rozkładzie geometrycznym G_p .

Dowód. Zauważmy, że zdarzenie $\{T = n\}$ jest takie samo jak zdarzenie $\{X_1 = 0, \dots, X_{n-1} = 0, X_n = 1\}$. Z niezależności zmiennych losowych X_i otrzymujemy

$$\begin{aligned} P(T = n) &= P(X_1 = 0, \dots, X_{n-1} = 0, X_n = 1) \\ &= P(X_1 = 0) \cdot \dots \cdot P(X_{n-1} = 0) \cdot P(X_n = 1) = q^{n-1}p. \end{aligned} \quad \square$$

Jeśli X jest zmienną o rozkładzie geometrycznym G_p , to

$$E(X) = \frac{1}{p}, \quad D^2(X) = \frac{1-p}{p^2}$$

Dowód. Wiadomo, że

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}, \quad \text{dla } |x| < 1.$$

Po zróżniczkowaniu otrzymujemy:

$$\sum_{i=1}^{\infty} ix^{i-1} = \frac{1}{(1-x)^2}, \quad \text{dla } |x| < 1.$$

Teraz, biorąc $x = 1 - p$ otrzymujemy:

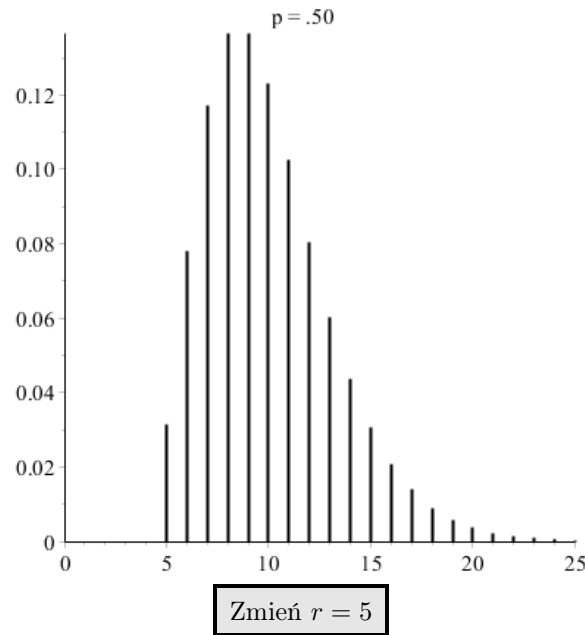
$$E(X) = \sum_{i=1}^{\infty} i(1-p)^{i-1}p = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

Wariancję oblicza się podobnie (ćwiczenie). □

8.5 Rozkład Pascala, ujemny rozkład dwumianowy

Rozkład Q nazywamy ujemnym rozkładem dwumianowym (lub rozkładem Pascala), jeżeli istnieją: liczba naturalna $r \geq 1$ oraz rzeczywista $p > 0$ takie, że

$$Q(r+k) = \binom{r+k-1}{r-1} p^r (1-p)^k, \quad \text{dla } k = 0, 1, 2, \dots$$



Zauważmy, że rozkład geometryczny jest szczególnym przypadkiem ujemnego rozkładu dwumianowego dla $r = 1$.

Twierdzenie – 8.7. Niech X_1, X_2, X_3, \dots będzie ciągiem niezależnych prób Bernoulliego o takim samym prawdopodobieństwie sukcesu p w każdej próbie. Określamy:

$$T_r := \min\{n : \exists 1 \leq k_1 < \dots < k_r = n \text{ takie, że } X_{k_i} = 1, \text{ dla } i = 1, \dots, r\}.$$

Wtedy, T_r jest zmienną losową o ujemnym rozkładzie dwumianowym. Inaczej: Czas oczekiwania na pierwszych r sukcesów w nieskończonym schemacie Bernoulliego ma ujemny rozkład dwumianowy.

Dowód. Dowód jest bardzo podobny do analogicznego twierdzenia o rozkładzie geometrycznym.

$$\{T_r = r + k\} = \bigcup \{X_1 = \varepsilon_1, \dots, X_{r+k-1} = \varepsilon_{r+k-1}, X_{r+k} = 1\},$$

gdzie sumowanie odbywa się po wszystkich $\{\varepsilon_1, \dots, \varepsilon_{r+k-1}\}$ takich, że spośród nich $r - 1$ ma wartość 1 oraz k ma wartość 0. Wtedy $P(\{X_1 = \varepsilon_1, \dots, X_{r+k-1} = \varepsilon_{r+k-1}, X_{r+k} = 1\}) = p^r(1 - p)^k$. \square

Można także udowodnić twierdzenie, które jeszcze inaczej pozwala spojrzeć na problem czasów oczekiwania:

Twierdzenie – 8.8. Niech T_1, \dots, T_r będzie ciągiem niezależnych zmiennych losowych o takim samym rozkładzie geometrycznym każda.

Wtedy suma $T_1 + \dots + T_r$ ma ujemny rozkład dwumianowy.

Dowód. Indukcja ze względu na r (ćwiczenie). \square

Dla zmiennej losowej X o ujemnym rozkładzie dwumianowym zachodzi

$$E(X) = \frac{r}{p}, \quad D^2(X) = \frac{r(1-p)}{p^2}.$$

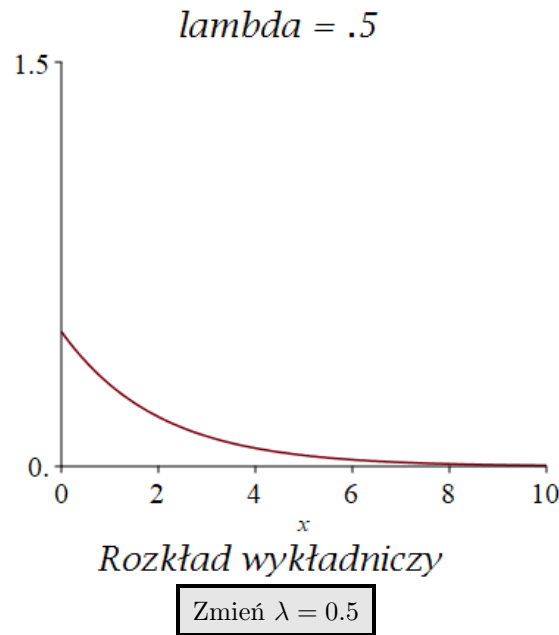
Dowód. Wynika z poprzedniego twierdzenia (ćwiczenie). \square

8.6 Rozkład wykładniczy, E_λ

Rozkład Q nazywamy rozkładem wykładniczym, jeżeli istnieje taka liczba $\lambda > 0$, że funkcja f określona wzorem

$$f(x) = \begin{cases} 0, & \text{dla } x < 0 \\ \lambda e^{-\lambda x}, & \text{dla } x \geq 0. \end{cases}$$

jest gęstością tego rozkładu.



Dystrybuanta

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & \text{dla } x < 0 \\ 1 - e^{-\lambda x}, & \text{dla } x \geq 0. \end{cases}$$

$$E(X) = \frac{1}{\lambda}, \quad D^2(X) = \frac{1}{\lambda^2}. \quad (8.1)$$

Dowód. Proste przeliczenie (ćwiczenie). □

Rozkład wykładniczy jest ciągłym odpowiednikiem rozkładu geometrycznego. Mówiąc nieściśle, czas oczekiwania na pierwszy sukces w nieskończonym ciągu niezależnych prób Bernoulliego ma w przybliżeniu rozkład wykładniczy o parametrze λ , o ile czas pomiędzy kolejnymi próbami jest bardzo mały, a prawdopodobieństwo sukcesu w pojedynczej próbie jest małe i proporcjonalne do tego czasu, przy czym parametr λ jest współczynnikiem tej proporcjonalności. Inaczej, gdy jednostką czasu jest δ oraz p jest bliskie zeru, to rozkład geometryczny o parametrze p i wykładniczy $\lambda = p$ są podobne. Zobacz sam.

Poniżej formułujemy odpowiednie twierdzenie.

Niech $\lambda > 0$ będzie ustalone.

Dla $\delta > 0$ oznaczamy $p = p_\delta = \lambda \cdot \delta$.

Niech X_1, X_2, X_3, \dots będzie ciągiem niezależnych zmiennych losowych, z których każda ma rozkład dwupunktowy o parametrze p .

Niech

$$T = \delta \min\{n \geq 1 : X_n = 1\}.$$

Niech F oznacza dystrybuantę rozkładu wykładniczego o parametrze λ .

Twierdzenie – 8.9. Dla każdego $t \in \mathbb{R}$

$$F_T(t) \longrightarrow F(t) \quad \text{gdy } \delta \longrightarrow 0.$$

Dowód. Dla $t \leq 0$ – trywialne.

Niech $t > 0$. Zmienna losowa $\frac{T}{\delta}$ ma rozkład geometryczny. Niech $n = [\frac{t}{\delta}]$.

$$\begin{aligned} F_T(t) &= P(T \leq t) = 1 - P(T > t) = 1 - P\left(\frac{T}{\delta} > \frac{t}{\delta}\right) = 1 - \sum_{k=n+1}^{\infty} (1-p)^{k-1} p \\ &= 1 - (1-p)^n = 1 - \left(1 - \frac{\lambda}{\delta-1}\right)^{\delta^{-1}t - r_\delta} \longrightarrow 1 - e^{-\lambda t} = F(t), \end{aligned}$$

dla $\delta \rightarrow 0$, gdyż $0 \leq r_\delta = \frac{t}{\delta} - n < 1$, więc $\left(1 - \frac{\lambda}{\delta-1}\right)^{-r_\delta}$ zmierza do 1. □

Ilustracja twierdzenia

Twierdzenie – 8.10. Niech T_1, \dots, T_r będą zmiennymi losowymi niezależnymi o takim samym rozkładzie wykładniczym o parametrze λ . Niech $S_r = T_1 + \dots + T_r$.

Wtedy S_r ma rozkład o gęstości f_r :

$$f_r(x) = \frac{\lambda(\lambda x)^{r-1}}{(r-1)!} e^{-\lambda x} \quad \text{dla } x > 0$$

oraz $f_r(x) = 0$ dla $x \leq 0$.

Powyższy rozkład nosi nazwę rozkładu Erlanga.

Dowód. Rachunkowy dowód polega na zastosowaniu indukcji oraz następującego wzoru na gęstość sumy niezależnych zmiennych losowych o rozkładach ciągłych (ćwiczenie). □

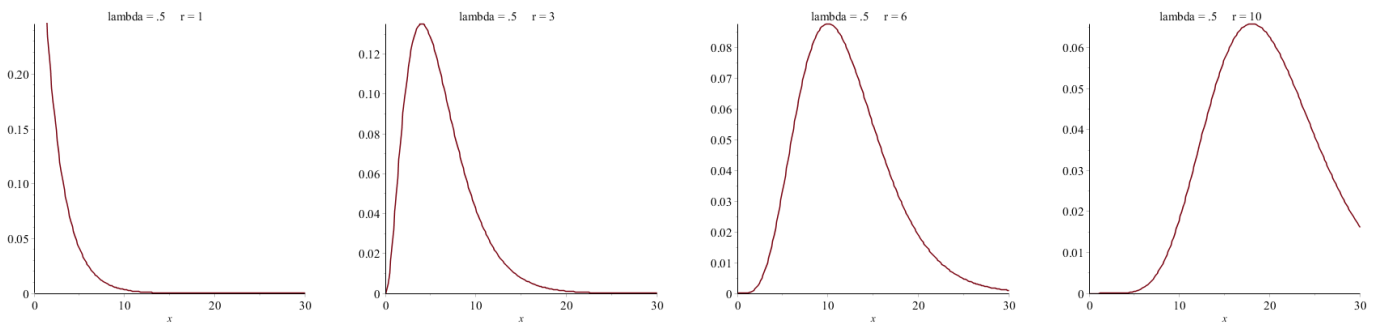
Twierdzenie – 8.11. Niech X oraz Y będą niezależnymi zmiennymi losowymi o gęstościach f_X oraz f_Y . Wtedy zmienna losowa $X + Y$ ma rozkład ciągły o gęstości:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(t) f_Y(z-t) dt.$$

Dowód. Wykorzystuje twierdzenie o zmianie zmiennych w całce podwójnej oraz twierdzenie Fubiniego. □

Indukcyjnie można pokazać, że dystrybuanta wyraża się wzorem (ćwiczenie):

$$F_r(t) = \int_0^t \frac{\lambda(\lambda x)^{r-1}}{(r-1)!} e^{-\lambda x} dx = 1 - e^{-\lambda t} \left(1 + \frac{\lambda t}{1!} + \dots + \frac{(\lambda t)^{r-1}}{(r-1)!} \right).$$



8.7 Proces Poissona

Twierdzenie – 8.12. Niech T_1, T_2, T_3, \dots będą zmiennymi losowymi niezależnymi o takim samym rozkładzie wykładniczym o parametrze λ . Niech $S_n = T_1 + \dots + T_n$. Kładziemy dodatkowo $S_0 = 0$. Definiujemy:

$$N_t := \max\{n : S_n \leq t\},$$

gdzie $t > 0$ jest ustaloną liczbą.

Wtedy zmienna losowa N_t ma rozkład Poissona o parametrze λt .

Komentarz. Zmienna N_t oznacza liczbę sukcesów, które mają miejsce na odcinku czasu $(0, t)$ w ciągu niezależnych prób Bernoulliego, o ile próby te mogą być powtarzane nieskończenie często, a prawdopodobieństwo pojawienia się sukcesu w bardzo małym odcinku czasu Δt wynosi w przybliżeniu $\lambda \Delta t$.

Dowód. Zauważmy, że zdarzenie $\{N_t = k\}$ jest równe zdarzeniu $\{S_k \leq t\} \setminus \{S_{k+1} \leq t\}$. Tak więc:

$$P(N_t = k) = F_k(t) - F_{k+1}(t),$$

gdzie F_k oznacza dystrybuantę zmiennej losowej S_k , która ma rozkład Erlanga. Poprzednio określiliśmy już dystrybuantę F_k .

Stąd łatwo widać, że: $P(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$. □

Proces stochastyczny. Rodzina zmiennych losowych określonych na tej samej przestrzeni probabilistycznej indeksowana przez czas nazywa się *procesem stochastycznym*. Najczęściej zbiorem czasów jest zbiór liczb naturalnych lub zbiór nieujemnych liczb rzeczywistych. Powyższa rodzina $\{N_t\}_{t \geq 0}$ jest właśnie takim przypadkiem i nazywa się *procesem Poissona*.

8.8 Pytania

Pytanie 8.1. Dwóch graczy wykonuje 10 rzutów kostką. Jakie jest prawdopodobieństwo tego, że obydwaj otrzymają tyle samo „6”? ROZWIĄZANIE

Pytanie 8.2. Czy/kiedy suma dwóch niezależnych zmiennych losowych o rozkładach dwumianowych ma rozkład dwumianowy? ROZWIĄZANIE

Pytanie 8.3. Ile rodzynek podczas wyrabiania ciasta trzeba średnio przeznaczyć na bułeczkę, aby losowo wybrana bułeczka z prawdopodobieństwem 0,95 lub większym zawierała co najmniej jedną rodzynkę? Jakie wtedy będzie prawdopodobieństwo tego, że losowo wybranej bułeczce będzie co najmniej 5 rodzynek? ROZWIĄZANIE

Pytanie 8.4. Przeprowadzić dowód twierdzenia 8.11. ROZWIĄZANIE

Pytanie 8.5. Znaleźć rozkład sumy niezależnych zmiennych losowych o rozkładzie $U(0, a)$ każda. ROZWIĄZANIE

Pytanie 8.6. Wykazać, że minimum dwóch niezależnych zmiennych losowych o rozkładach wykładniczych ma rozkład wykładniczy. ROZWIĄZANIE

Rozdział 9

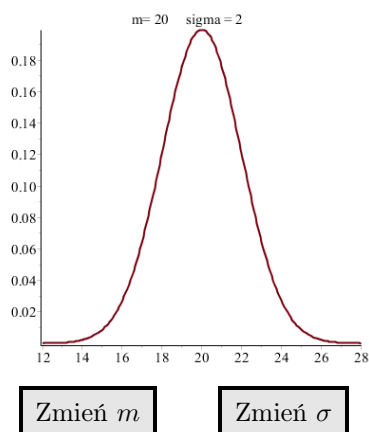
Rozkład normalny

Najczęściej spotykamy się z rozkładem Gaussa, zwanym także normalnym.

Rozkład normalny, $N(m, \sigma)$. Rozkład Q nazywamy rozkładem normalnym, jeżeli istnieją takie liczby rzeczywiste m oraz $\sigma > 0$, że funkcja $f: \mathbb{R} \rightarrow \mathbb{R}$, określona wzorem:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \quad \text{dla } x \in \mathbb{R},$$

jest gęstością tego rozkładu.



f jest rzeczywiście gęstością. Wiadomo (podstawowy fakt z analizy matematycznej, ćwiczenie), że

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi}.$$

Stosując podstawienie $t = \frac{x-m}{\sigma}$, otrzymujemy:

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = 1.$$

Interpretacja parametrów. m – punkt maksimum globalnego gęstości (ćwiczenie).

$m - \sigma$, $m + \sigma$ – punkty przegięcia gęstości (ćwiczenie).

$$E(X) = m, \quad D^2(X) = \sigma^2.$$

Dowód. Proste całkowanie przez podstawienie. □

Oznaczenie. $\Phi_{m,\sigma}$ – dystrybuanta rozkładu normalnego $N(m, \sigma)$, czyli

$$\Phi_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-m}{\sigma}\right)^2} dt.$$

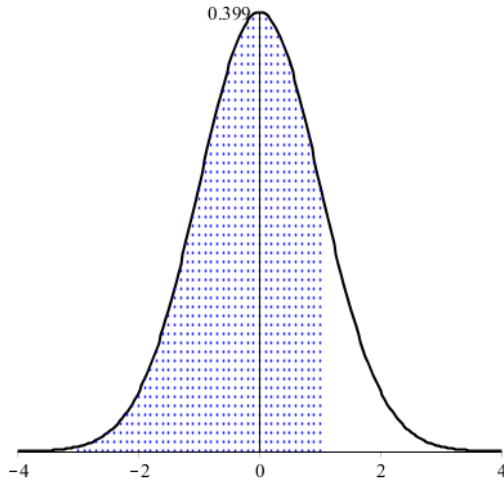
9.1 Standardowy rozkład normalny, $N(0, 1)$

Rozkład $N(0, 1)$ – standardowy rozkład normalny.

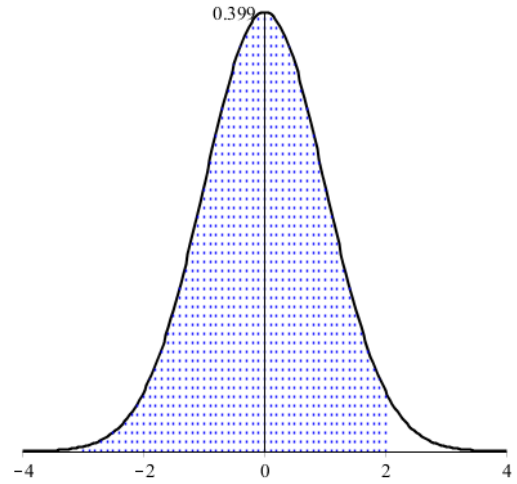
$$E(X) = 0, \quad D^2(X) = 1.$$

Oznaczenie. $\Phi = \Phi_{0,1}$ – dystrybuanta rozkładu normalnego standardowego, czyli

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt.$$



$$\Phi(1) = 0.8413447461$$



$$\Phi(2) = 0.9772498681$$

Pożyteczne wzory (ćwiczenie):

$$\Phi(0) = 0.5,$$

$$\Phi(-x) = 1 - \Phi(x),$$

$$P(|X| < \varepsilon) = \Phi(\varepsilon) - \Phi(-\varepsilon) = 2\Phi(\varepsilon) - 1, \text{ gdy } X \sim N(0, 1),$$

$$\Phi_{m,\sigma}(x) = \Phi\left(\frac{x - m}{\sigma}\right).$$

Przed erą komputerową w powszechnym użyciu były tablice rozkładu $N(0, 1)$.

| x | 0,00 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 | 0,09 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0,0 | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 | 0,5359 |
| 0,1 | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 | 0,5753 |
| 0,2 | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 | 0,6141 |
| 0,3 | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 | 0,6517 |
| 0,4 | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 | 0,6879 |
| 0,5 | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 | 0,7224 |
| 0,6 | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 | 0,7549 |
| 0,7 | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 | 0,7852 |
| 0,8 | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 | 0,8133 |
| 0,9 | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8264 | 0,8289 | 0,8315 | 0,8340 | 0,8365 | 0,8389 |
| 1,0 | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 | 0,8621 |
| 1,1 | 0,8643 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 | 0,8830 |
| 1,2 | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 | 0,9015 |
| 1,3 | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 | 0,9177 |
| 1,4 | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 | 0,9319 |
| 1,5 | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 | 0,9441 |
| 1,6 | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 | 0,9545 |
| 1,7 | 0,9554 | 0,9564 | 0,9573 | 0,9582 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 | 0,9633 |
| 1,8 | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 | 0,9706 |
| 1,9 | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 | 0,9767 |
| 2,0 | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 | 0,9817 |
| 2,1 | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 | 0,9857 |
| 2,2 | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 | 0,9890 |
| 2,3 | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 | 0,9916 |
| 2,4 | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 | 0,9936 |
| 2,5 | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 | 0,9952 |
| 2,6 | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 | 0,9964 |
| 2,7 | 0,9965 | 0,9966 | 0,9967 | 0,9968 | 0,9969 | 0,9970 | 0,9971 | 0,9972 | 0,9973 | 0,9974 |
| 2,8 | 0,9974 | 0,9975 | 0,9976 | 0,9977 | 0,9977 | 0,9978 | 0,9979 | 0,9979 | 0,9980 | 0,9981 |
| 2,9 | 0,9981 | 0,9982 | 0,9982 | 0,9983 | 0,9984 | 0,9984 | 0,9985 | 0,9985 | 0,9986 | 0,9986 |
| 3,0 | 0,9987 | 0,9987 | 0,9987 | 0,9988 | 0,9988 | 0,9989 | 0,9989 | 0,9989 | 0,9990 | 0,9990 |

9.2 Centralne twierdzenie graniczne

W tym punkcie zakładamy, że:

Założenie.

(Ω, Σ, P) jest przestrzenią probabilistyczną, zaś X_1, X_2, X_3, \dots – ciągiem niezależnych zmiennych losowych określonych na Ω . Wszystkie zmienne losowe X_i mają taki sam rozkład, a ich wspólna nadzieja matematyczna m oraz wariancja σ^2 istnieją i są skończone, przy czym $\sigma > 0$ (ten ostatni warunek oznacza, że zmienne losowe nie są stałymi).

$$S_n = X_1 + \dots + X_n.$$

Zmienną losową:

$$Z_n := \frac{S_n - E(S_n)}{\sqrt{D^2(S_n)}} = \frac{S_n - nm}{\sigma\sqrt{n}}$$

nazywamy standaryzacją sumy S_n .

Jak łatwo zauważyć:

$$E(Z_n) = 0 \quad \text{oraz} \quad D^2(Z_n) = 1.$$

Twierdzenie – 9.1 (Twierdzenie Lindeberga–Lévy’ego, centralne twierdzenie graniczne (CTG)).

Dla każdego $x \in \mathbb{R}$ zachodzi równość:

$$\lim_{n \rightarrow \infty} P(Z_n \leq x) = \Phi(x).$$

Twierdzenie – 9.2 (CTG dla sum). Rozkład zmiennej losowej S_n jest asymptotycznie równy rozkładowi $N(nm, \sigma\sqrt{n})$.

Inaczej:

$$\lim_{n \rightarrow \infty} \left(F_{S_n}(x) - \Phi_{nm, \sigma\sqrt{n}}(x) \right) = 0,$$

dla $x \in \mathbb{R}$.

Twierdzenie – 9.3 (CTG dla średnich). Rozkład zmiennej losowej $\frac{S_n}{n}$ jest asymptotycznie równy rozkładowi $N\left(m, \frac{\sigma}{\sqrt{n}}\right)$.

Inaczej:

$$\lim_{n \rightarrow \infty} \left(F_{\frac{S_n}{n}}(x) - \Phi_{m, \frac{\sigma}{\sqrt{n}}}(x) \right) = 0,$$

dla $x \in \mathbb{R}$.

Dowód oparty na teorii funkcji charakterystycznych będzie później, patrz punkt 11.3.

Centralne twierdzenie graniczne jest prawdziwe przy dużo ogólniejszych założeniach. W szczególności zmienne losowe nie muszą mieć takiego samego rozkładu, a nawet nie muszą być niezależne. Jednakże, różnym wersjom centralnego twierdzenia granicznego przyświeca ta sama idea:

Suma niewiele zależnych od siebie składników losowych, z których żaden nie dominuje istotnie nad pozostałymi, ma w przybliżeniu rozkład normalny.

Z twierdzenia 9.1 otrzymujemy natychmiast klasyczne twierdzenie:

Twierdzenie – 9.4 (de Moivre’a–Laplace’a). Niech X_1, X_2, X_3, \dots będzie ciągiem niezależnych prób Bernoulliego, z takim samym prawdopodobieństwem sukcesu p i porażki $q = 1 - p$ w każdej próbie ($0 < p < 1$). Wtedy:

$$P\left(\frac{S_n - np}{\sqrt{npq}} \leq x\right) \rightarrow \Phi(x),$$

dla każdego $x \in \mathbb{R}$.

Ilustracja twierdzenia

Przykład – 9.5 (Eksperyment, rzuty kostką). Wyobraźmy sobie eksperyment polegający na wielokrotnym rzucie kostką do gry. Suma uzyskanych oczek S jest zmienną losową mającą, zgodnie z CTG, w przybliżeniu rozkład $N(nm, \sigma\sqrt{n})$, gdzie m oraz σ są odpowiednio nadzieją matematyczną oraz odchyleniem standardowym zmiennej losowej X , reprezentującej wynik pojedynczego rzutu, a n jest liczbą wykonanych prób. Ponieważ X ma rozkład

dyskretny, skupiony w punktach 1, 2, 3, 4, 5, 6 przyjmowanych z jednakowym prawdopodobieństwem $\frac{1}{6}$, więc bez trudu można stwierdzić, że:

$$m = 3.5 \text{ oraz } \sigma = \frac{\sqrt{105}}{6} \approx 1.7078251.$$

Przypuśćmy, że wykonano 1000 rzutów ($n = 1000$). Wówczas na podstawie CTG suma S_{1000} ma w przybliżeniu rozkład $N(3500, 54.00617)$.

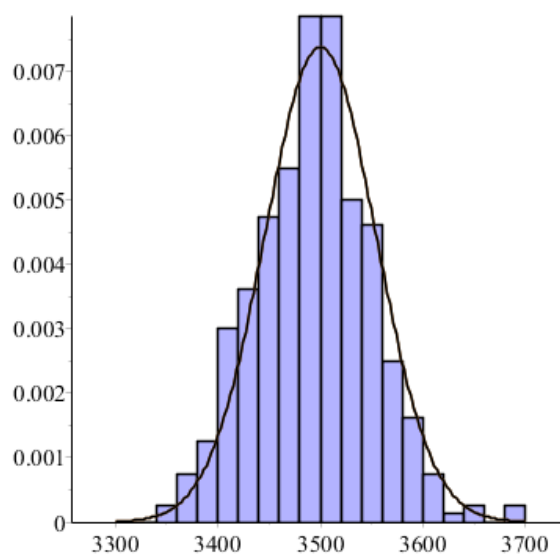
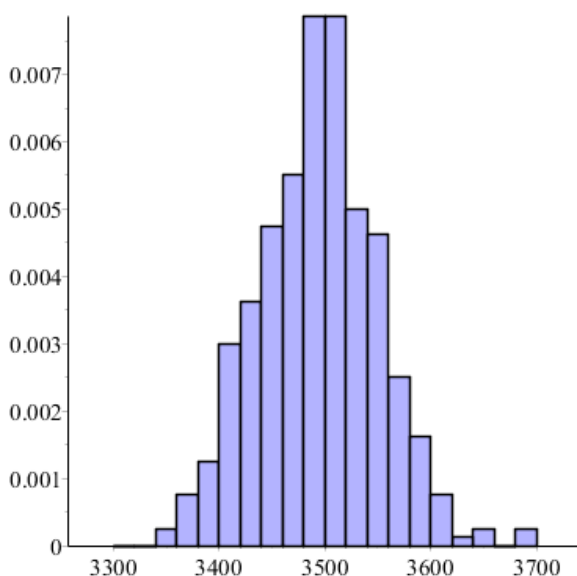
Zweryfikujmy doświadczalnie uzyskany wynik. W tym celu można przeprowadzić symulację tysiąca rzutów kostką za pomocą komputera, uzyskując odpowiednią wartość sumy wszystkich uzyskanych oczek.

Doświadczenie to powtórzmy 400 razy, uzyskując 400 wartości sumy oczek.

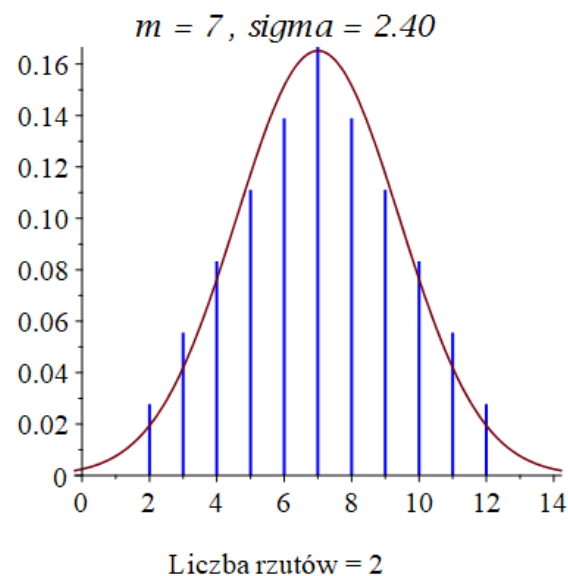
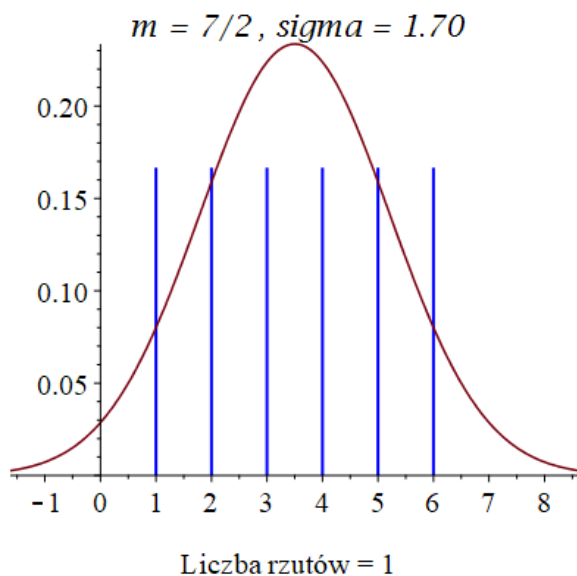
Wyniki:

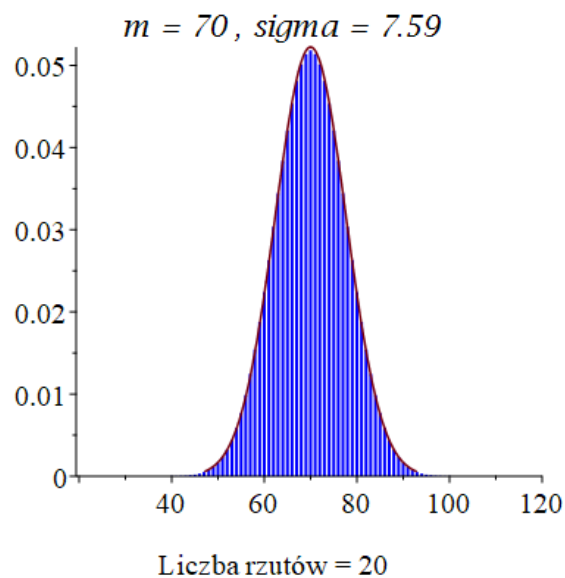
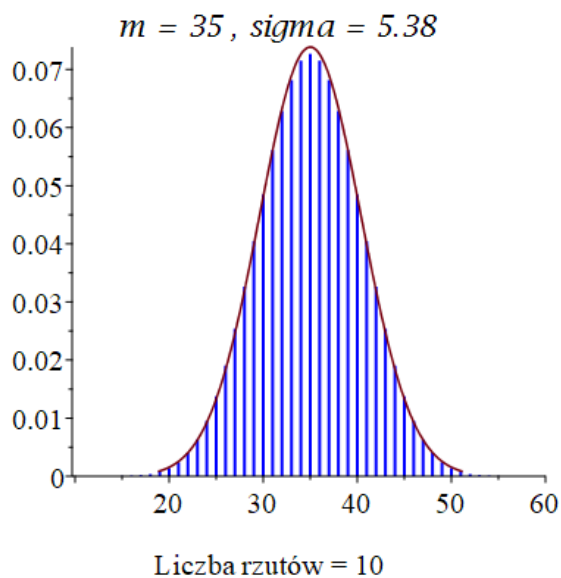
$$3567, 3423, 3424, \dots, 3671, 3558, 3582.$$

są przedstawione graficznie w postaci histogramu. W tym celu przedział $[3300, 3700]$ został podzielony na 20 równych przedziałów i została policzona liczba danych znajdujących się w każdym z tych przedziałów, n_i , $i = 1, \dots, 20$, a na rysunku zostały zaznaczone prostokąty o wysokościach $\frac{n_i}{20N}$ nad kolejnymi przedziałami. Tutaj $N = \sum_{i=1}^{20} n_i$. Widać, że suma pól = 1. Histogram porównano na wspólnym rysunku z gęstością rozkładu $N(3500, 54, 00617)$.

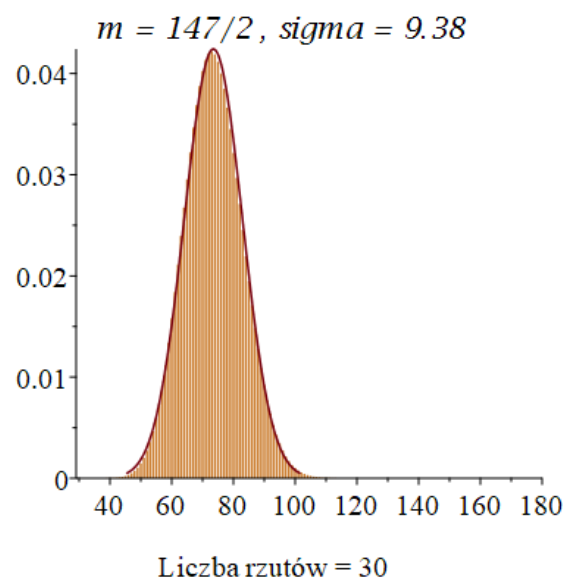
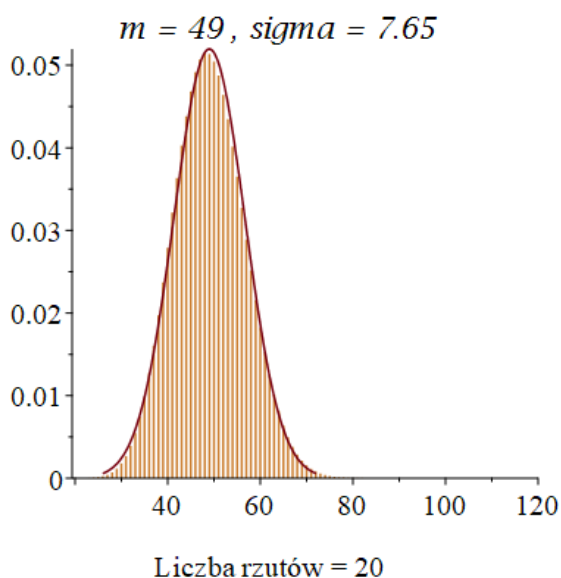
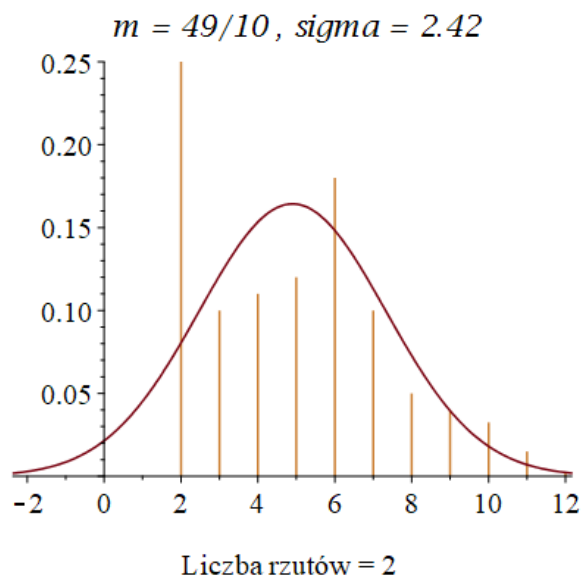
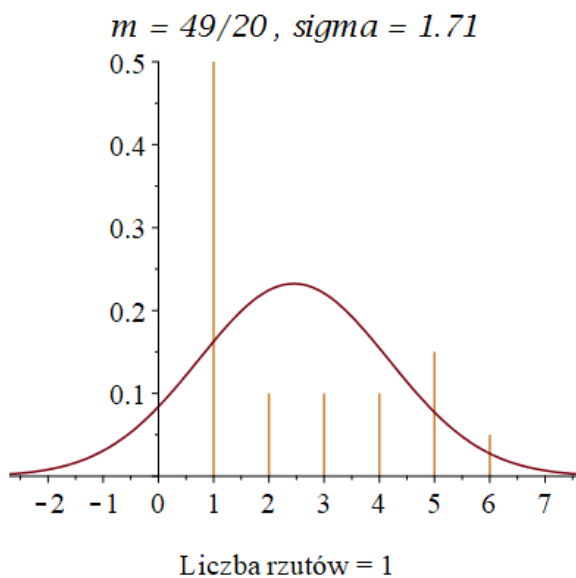


W istocie nie trzeba rzucać kostką aż 1000 razy aby suma oczek miała rozkład w przybliżeniu normalny. Faktycznie wystarczy ograniczyć liczbę rzutów do kilkunastu. Może o tym świadczyć porównanie rozkładu sumy z odpowiednim rozkładem normalnym z naturalnie dobranymi parametrami.





Nawet, gdy kostka jest wyraźnie sfalszowana („1” wypada dużo częściej niż „6” – patrz rozkład na następnym rysunku) suma oczek dość szybko „normalnieje”.



Przykład – 9.6. Rzucono 1000 razy symetryczną kostką do gry. Obliczyć prawdopodobieństwo tego, że „6” wypadła więcej niż 150 razy.

Zauważmy najpierw, że interesująca nas ilość „6” jest sumą S_n , $n = 1000$, niezależnych prób Bernoulliego o prawdopodobieństwie sukcesu $p = \frac{1}{6}$ w każdej próbie. Zgodnie z centralnym twierdzeniem granicznym, suma ta ma w przybliżeniu rozkład normalny $N(np, \sqrt{npq})$. Wstawiając wartości liczbowe otrzymujemy: $P(S_{1000} > 150) = 1 - P(S_{1000} \leq 150) \approx 1 - \Phi_{np, \sqrt{npq}}(150) = 1 - \Phi\left(\frac{150 - \frac{1000}{6}}{\sqrt{1000 \frac{5}{6} \frac{1}{6}}}\right) \approx 1 - \Phi(-1.41) = \Phi(1.41) \approx 0.9207$, gdzie ostatnia liczba pochodzi z tablic rozkładu normalnego.

Przykład – 9.7. Jakie jest prawdopodobieństwo, że przy 1000 rzutach monetą symetryczną różnica między ilością reszek i orłów będzie wynosić co najmniej 100? Podobnie jak poprzednio, ilość uzyskanych orłów jest sumą S_n , $n = 1000$, niezależnych prób Bernoulliego o prawdopodobieństwie sukcesu $p = \frac{1}{2}$ w pojedynczej próbie. Chcemy obliczyć $P(|S_n - (n - S_n)| \geq 100)$, czyli $P(|S_n - 500| \geq 50)$. Prawdopodobieństwo zdarzenia przeciwnego jest w bardzo dużym przybliżeniu równe:

$$F_{S_n}(550) - F_{S_n}(450) \approx \Phi_{500, 5\sqrt{10}}(550) - \Phi_{500, 5\sqrt{10}}(450) = \Phi(\sqrt{10}) - \Phi(-\sqrt{10}) = 2\Phi(\sqrt{10}) - 1 \approx 2\Phi(3.1622) - 1 \approx 0.9984.$$

Interesujące więc nas prawdopodobieństwo wynosi w przybliżeniu 0,0016.

Przykład – 9.8 (Kontynuacja przykładu 6.15 o liczbie analiz). W celu zbadania dużej populacji osób, podzielono ją na grupy, a następnie pobrano od każdej osoby krew oraz przeprowadzano analizę łączną dla poszczególnych grup, wykonując odpowiedni test na próbkach powstałych przez zmieszanie krwi osób należących do tej samej grupy. Gdy w pewnej grupie wykryto wirus chorobowy, przeprowadzano odrębną analizę dla każdej osoby z tej grupy. Załóżmy, że liczebność populacji wynosi N , liczność grup wynosi n , zaś k niech będzie liczbą grup (oczywiście $N = nk$).

Zakładamy też, że prawdopodobieństwo tego, że dany człowiek jest zarażony interesującym nas wirusem wynosi p oraz że obecność wirusa u danej osoby jest niezależna od jego obecności u innych osób.

Pamiętamy, że dla $N = 1000$, oraz $p = 0.01$ optymalnymi ze względu na średnią liczbę analiz parametrami są: $n = 10$, $k = 100$. Wtedy oczekiwana liczba analiz wynosi $m = E(X) = 195.68$. **Pytanie:** czy liczba analiz może przekroczyć 300?

Stosując nierówność Czebyszewa stwierdziliśmy, że: $P(X \geq 300) \leq 0.0795$.

Stosując CTG możemy założyć, że zmienna losowa X oznaczająca liczbę analiz ma rozkład normalny, $N(m, \sigma)$. Obliczyliśmy już poprzednio: $m = 195.68$, $\sigma^2 = 865$.

Mamy więc: $P(X \geq 300) = 1 - P(X < 300) = 1 - \Phi\left(\frac{300-m}{\sigma}\right) \approx 1 - 0.999607712 = 0,000392288$.

Uwaga – 9.9 (Reguła 1.96). *Jeżeli zmienna losowa X ma rozkład normalny $N(m, \sigma)$, to*

$$P(X \in (m - 1.96\sigma, m + 1.96\sigma)) \approx 0.95.$$

Dowód.

$$\begin{aligned} P(X \in (m - 1.96\sigma, m + 1.96\sigma)) &= \Phi_{m, \sigma}(m + 1.96\sigma) - \Phi_{m, \sigma}(m - 1.96\sigma) \\ &= 2\Phi(1.96) - 1 \approx 2 \cdot 0.975002104851780 - 1 = 0.950004209703559. \end{aligned} \quad \square$$

Dla dowolnej zmiennej losowej X o parametrach $m = E(X)$, $\sigma^2 = D^2(X)$ z reguły 3σ otrzymujemy:

$$P(X \in (m - 3\sigma, m + 3\sigma)) \geq \frac{8}{9}.$$

Gdy założymy normalność X , to

$$P(X \in (m - 3\sigma, m + 3\sigma)) = 2\Phi(3) - 1 \approx 2 \cdot 0.999 - 1 \approx 0.997.$$

Przykład – 9.10. Aby stwierdzić, jak wielu wyborców popiera obecnie partię ABC , losujemy spośród nich reprezentatywną próbkę i na niej przeprowadzamy badanie. Jak duża powinna być ta próbka, aby uzyskany wynik różnił się od rzeczywistego poparcia dla partii ABC nie więcej niż o $b = 3\%$ z prawdopodobieństwem co najmniej $1 - \alpha = 0,95$?

Niech $p \in (0, 1)$ oznacza faktyczne (lecz nieznanne) poparcie dla partii ABC . Jeżeli próbka składa się z n osób, z których S_n wyraziło poparcie dla ABC , to liczba $\frac{S_n}{n}$ jest poparciem wyznaczonym na podstawie próbki. Możemy

założyć, że S_n jest sumą niezależnych zmiennych losowych ξ_i o rozkładzie: $P(\xi_i = 0) = 1 - p$, $P(\xi_i = 1) = p$. Chcemy znaleźć takie n , żeby:

$$P\left(\left|\frac{S_n}{n} - p\right| \leq b\right) \geq 1 - \alpha.$$

Ponieważ średnia arytmetyczna $\frac{S_n}{n}$ ma w przybliżeniu rozkład normalny, więc:

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - p\right| \leq b\right) &= P\left(\frac{S_n}{n} \in (p - b, p + b)\right) \\ &= \Phi_{p, \sqrt{\frac{p(1-p)}{n}}}(p + b) - \Phi_{p, \sqrt{\frac{p(1-p)}{n}}}(p - b) = \Phi\left(\frac{b\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(-\frac{b\sqrt{n}}{\sqrt{p(1-p)}}\right). \end{aligned}$$

Czyli powinna być spełniona następująca nierówność:

$$2\Phi\left(\frac{b\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 1 - \alpha,$$

która jest z kolei równoważna:

$$n \geq \left(\frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}{b}\right)^2 (1-p)p.$$

Chociaż nie znamy p , wiemy, że $(1-p)p \leq \frac{1}{4}$. W takim razie n spełniająca nierówność:

$$n \geq \left(\frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}{b}\right)^2 0.25,$$

spełnia także poprzednią nierówność, a więc określa (z nadstatkiem) wystarczającą wielkość próbki.

Podstawiając $b = 0.03$, $\alpha = 0.05$, otrzymamy: $n \geq 1067$.

Jeżeli jeszcze przed losowaniem próbki mamy wstępne informacje o poparciu dla partii ABC – na przykład wiemy, że poparcie to jest mniejsze niż 20% – możemy powyższy wynik znacznie polepszyć. Ponieważ $p \leq 0.2$, więc $(1-p)p \leq 0.16$, co oznacza, że $n \geq 683$ jest wystarczającą wielkością próbki.

Przeprowadzono sondaż i okazało się, że na 1050 badanych osób 299 popiera partię ABC . Jakie jest prawdziwe poparcie dla ABC ? Podobnie jak w poprzednim przykładzie, na tak postawione pytanie nie potrafimy odpowiedzieć. Potrafimy jednak z dużym prawdopodobieństwem wskazać przedział (zwany przedziałem ufności), w którym to poparcie p się zawiera. Jest to na przykład przedział postaci $(\hat{p} - b, \hat{p} + b)$, gdzie $\hat{p} = \frac{S_n}{n}$ (w naszym przypadku $\hat{p} = \frac{299}{1050} \approx 0.2848$), natomiast b jest tak dobrane, aby

$$P(p \in (\hat{p} - b, \hat{p} + b)) \geq 1 - \alpha,$$

przy czym $\alpha > 0$ jest ustaloną przez nas małą liczbą dodatnią. Rozumując jak poprzednio widzimy, że ten warunek jest równoważny warunkowi

$$2\Phi\left(\frac{b\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 1 - \alpha,$$

a dalej warunkowi

$$b \geq \frac{\sqrt{p(1-p)}}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \quad (9.1)$$

Powiedzmy, że $\alpha = 0.05$. Ponieważ nie znamy p , nie możemy podać możliwie najmniejszego b . Możemy jedynie formalnie stwierdzić, że

$$b \geq \frac{\sqrt{\frac{1}{2}(1 - \frac{1}{2})}}{\sqrt{1050}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \frac{\frac{1}{2}}{32.4} 1.96 = 0.0302.$$

Czyli szukany przedział ufności $(\hat{p} - b, \hat{p} + b)$ jest równy $(0.2545, 0.3150)$. Praktycy postępują jednak inaczej. Ponieważ znamy przybliżoną wartość p , czyli \hat{p} , to we wzorze (9.1) podstawmy \hat{p} zamiast p . Otrzymamy:

$$b \geq \frac{\sqrt{\hat{p}(1 - \hat{p})}}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad (9.2)$$

i po podstawieniu wartości: $b \geq 0.0273$ oraz przedział ufności

$$(0.2575, 0.3120).$$

Warto zauważyć, że gdyby partia *ABC* miała mniejsze poparcie, na przykład 99 na 1050 ankietowanych, to przedział ufności byłby istotnie krótszy. Mamy teraz $\hat{p} = 0.0894$, $b = 0.0177$ oraz przedział ufności

$$(0.0766, 0.1120).$$

Przykład – 9.11. Ze zbioru N -elementowego losujemy w kolejnych momentach po jednym elemencie, przy czym jest to losowanie ze zwracaniem. Interesuje nas rozkład czasu oczekiwania T na wylosowanie r różnych elementów. Widać, że:

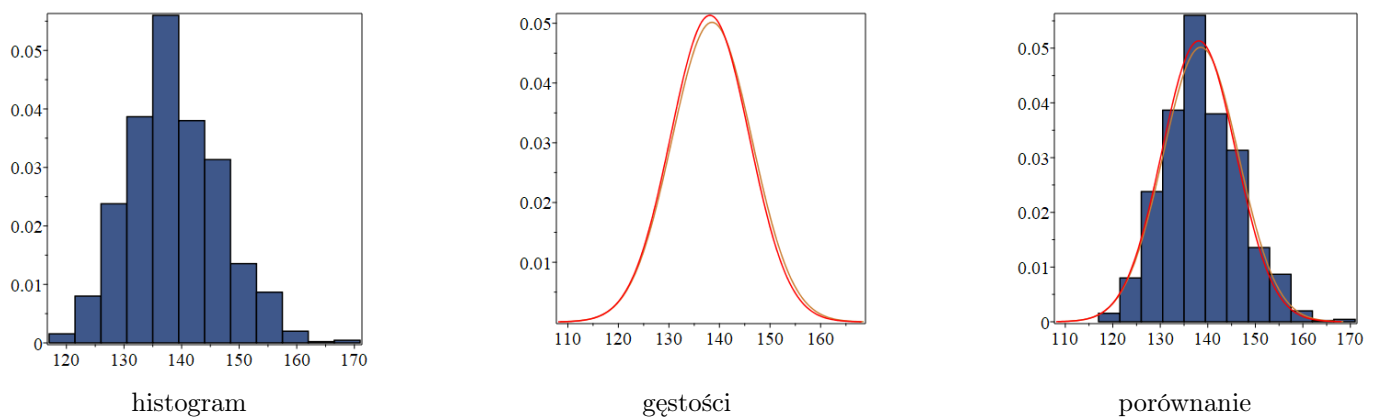
$$T = T_0 + \dots + T_{r-1},$$

gdzie T_n , $n = 0, 1, 2, \dots, r - 1$ są niezależnymi zmiennymi losowymi o rozkładach geometrycznych; $T_n \sim G_{\frac{N-n}{N}}$. Dla ustalonych N oraz r można obliczyć (komputer) $E(T)$ oraz $D^2(T)$, gdyż znamy te wielkości dla czasów T_n , a czasy te są niezależne. Na przykład, dla $N = 100$ oraz $r = 8$, $E(T) = 8.294833858$, $D^2(T) = 0.3105547438$. Gdy chcemy wylosować 190 różnych elementów spośród 200, potrzebujemy średnio prawie 590 losowań, a wariancja wynosi ponad 3 000.

Czy dla dużych r czas T ma rozkład normalny?

CTG w wersji, którą znamy, nie może być stosowane. Sprawdzamy to więc doświadczalnie wykonując 1000 symulacji naszego doświadczenia dla $N = 200$, $r = 100$.

Sporządzamy odpowiedni histogram, wyznaczamy średnią i wariancję z otrzymanej próby: $mD := 138.511$, $varD = 63.221100100100266$ oraz wyliczamy nadzieję matematyczną oraz wariancję T : $E(T) = 138.1306861$, $D^2(T) = 60.37514711$ i porównujemy na wspólnym wykresie:



Wydaje się (potwierdzają to testy statystyczne), że T ma rzeczywiście rozkład normalny.

9.3 Pytania

Pytanie 9.1. Czy suma niezależnych zmiennych losowych o rozkładach normalnych ma rozkład normalny? Czy założenie niezależności jest istotne? ROZWIĄZANIE

Pytanie 9.2. Sformułować odpowiednik reguły 1.96, gdy $\alpha = 0.01$. ROZWIĄZANIE

Pytanie 9.3. Niech X_1, X_2 będą niezależnymi zmiennymi losowymi o rozkładzie standardowym normalnym każda. Czy zmienne losowe $X_1 - X_2, X_1 + X_2$ są niezależne? Czy założenie normalności jest istotne? ROZWIĄZANIE

Pytanie 9.4. Partia *ABC* wie, że ma poparcie nie większe niż 10%. Zamawia sondaż, aby stwierdzić czy zdobędzie co najmniej 5% poparcie. Chciałaby mieć 99% pewności, że wynik sondażu oddaje prawdziwe preferencje wyborców z dopuszczalnym błędem nie większym niż 2%. Jak duża powinna być próbka ankietowanych osób? ROZWIĄZANIE

Pytanie 9.5. Pewna agencja prowadzi rekrutację pracowników w kilku różnych krajach, które stosują różne systemy punktowania, niemniej trudność używanych testów jest porównywalna. Agencja otrzymuje listy punktów uzyskanych przez kandydatów ze wszystkich krajach i na tej podstawie chce wybrać 100 najlepszych kandydatów. W jaki sposób może postąpić agencja, aby wybór był racjonalny? ROZWIĄZANIE

Rozdział 10

Zbieżność zmiennych losowych

10.1 Rodzaje zbieżności

Dany jest ciąg zmiennych losowych $X_n : \Omega \rightarrow \mathbb{R}$, $n = 1, 2, 3, \dots$ oraz zmienna losowa $X : \Omega \rightarrow \mathbb{R}$.

Rozróżniamy kilka rodzajów zbieżności $X_n \rightarrow X$, gdy $n \rightarrow \infty$. W trakcie tego kursu dyskutujemy głównie:

1. Zbieżność z prawdopodobieństwem 1.
2. Zbieżność stochastyczna.
3. Zbieżność według rozkładów.

Nie rozważa się zbieżności punktowej zmiennych losowych!

Powód. Jeżeli zmienną losową, powiedzmy X , zmienimy na zbiorze miary zero otrzymując zmienną, powiedzmy Y , to obydwie te zmienne mają taki sam rozkład, patrz uwaga poniżej, więc z punktu widzenia rachunku prawdopodobieństwa są sobie równe. Jednak, gdy $X_n \rightarrow X$ punktowo, to wtedy X_n nie może być zbieżny w niektórych punktach do Y . \square

Mówimy, że dwa wektory losowe X, Y określone na przestrzeni probabilistycznej (Ω, Σ, P) są *równe prawie wszędzie* ($X = Y$ pw.) względem miary P , gdy $P(X = Y) = 1$. Często wiadomo, o którą miarę chodzi i wtedy nie musimy tego podkreślać. Można łatwo pokazać następujący fakt.

Uwaga – 10.1. Jeżeli $X = Y$ pw., to $P_X = P_Y$.

Definicja – 10.2 (Zbieżność z prawdopodobieństwem 1). $X_n \xrightarrow{1} X \iff$

$$P(\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1.$$

Definicja – 10.3 (Zbieżność stochastyczna). $X_n \xrightarrow{s} X \iff$

$$\forall \varepsilon > 0 \lim_{n \rightarrow \infty} P(|X_n - X| \leq \varepsilon) = 1.$$

\iff (ćwiczenie)

$$\forall \varepsilon > 0 \lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0.$$

Definicja – 10.4 (Zbieżność według rozkładów). $X_n \xrightarrow{d} X \iff$

$$\forall a \in \mathbb{R}, \text{ punktu ciągłości } F_X : \lim_{n \rightarrow \infty} F_{X_n}(a) = F_X(a).$$

Ogólniej.

Niech F_n , $n = 1, 2, 3, \dots$, oraz F będą dystrybuantami.

Definicja – 10.5 (Zbieżność rozkładów). $F_n \xrightarrow{d} F \iff$

$$\forall a \in \mathbb{R}, \text{ punktu ciągłości } F : \lim_{n \rightarrow \infty} F_n(a) = F(a).$$

Wtedy:

$$X_n \xrightarrow{d} X \iff F_{X_n} \xrightarrow{d} F_X.$$

Posługując się wprowadzoną terminologią możemy zreformułować kilka wcześniej podanych twierdzeń.

Słabe Prawo Wielkich Liczb (zob. twierdzenie 7.6.2). Niech X_n będą niezależnymi zmiennymi losowymi o wspólnej nadziei matematycznej m i wspólnie ograniczonych wariancjach. Niech $S_n = X_1 + \dots + X_n$. Wtedy $\frac{S_n}{n} \xrightarrow{s} m$.

Centralne twierdzenie graniczne (zob. twierdzenie 9.1). Niech X_n będą niezależnymi zmiennymi losowymi o takim samym rozkładzie. Niech m oznacza nadzieję matematyczną, a σ odchylenie standardowe tego rozkładu. Niech $Z_n = \frac{S_n - m}{\sigma\sqrt{n}}$. Wtedy

$$F_{Z_n} \xrightarrow{d} \Phi,$$

gdzie Φ jest dystrybuantą rozkładu $N(0, 1)$.

Zbieżność rozkładów geometrycznych do rozkładu wykładniczego. Twierdzenie 8.9 mówi o zbieżności ciągu dystrybuant rozkładów czasu oczekiwania na pierwszy sukces w ciągu niezależnych prób Bernoulliego, gdy odcinki czasowe są coraz krótsze, a prawdopodobieństwo sukcesu zmniejsza się proporcjonalnie wraz z ich długością.

Twierdzenie – 10.6 (Przybliżenie rozkładu dwumianowego rozkładem Poissona.). Niech liczby $p_n > 0$ tworzą taki ciąg, że:

$$\lim_{n \rightarrow \infty} np_n = \lambda > 0.$$

Wtedy:

$$F_{B(n, p_n)} \xrightarrow{d} F_{P_\lambda}.$$

Dowód. Niech a będzie punktem ciągłości dystrybuanty rozkładu Poissona P_λ . Korzystając z twierdzenia 8.3 mamy:

$$\begin{aligned} F_{P_\lambda}(a) &= \sum_{k \leq a} e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k \leq a} \lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \lim_{n \rightarrow \infty} \sum_{k \leq a} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \lim_{n \rightarrow \infty} F_{B(n, p_n)}(a). \quad \square \end{aligned}$$

Twierdzenie – 10.7. X_1, X_2, X_3, \dots, X – zmienne losowe. Zachodzą implikacje:

1. $X_n \xrightarrow{1} X \implies X_n \xrightarrow{s} X$.
2. $X_n \xrightarrow{s} X \implies X_n \xrightarrow{d} X$.
3. $X_n \xrightarrow{d} X, X \equiv c \in \mathbb{R} \implies X_n \xrightarrow{s} X$.

Dowód. Ad 1. Niech $A = \{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega), n \rightarrow \infty\}$. Z założenia wiemy, że $P(A) = 1$. Ustalmy $\varepsilon > 0$.

Wiemy, że $A \subset A_\varepsilon$, gdzie $A_\varepsilon = \bigcup_{N+1}^{\infty} A_{\varepsilon N}$, gdzie $A_{\varepsilon N} = \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon, \text{ dla } n \geq N\}$. Zbiory $A_{\varepsilon N}$ tworzą ciąg wstępujący. Mamy więc: $1 = P(A) \leq P(A_\varepsilon) = \lim_{N \rightarrow \infty} P(A_{\varepsilon N}) \leq \lim_{N \rightarrow \infty} P(\{\omega : |X_N(\omega) - X(\omega)| < \varepsilon\})$. \square

Ad 2. Niech a będzie punktem ciągłości dystrybuanty F_X i niech $\varepsilon > 0$ będzie ustalone. Dla każdego n zachodzą dwie oczywiste inkluzje

$$\{X \leq a - \varepsilon\} \subset \{|X_n - X| \geq \varepsilon\} \cup \{X_n \leq a\}$$

oraz

$$\{X_n \leq a\} \subset \{|X_n - X| \geq \varepsilon\} \cup \{X \leq a + \varepsilon\}.$$

To oznacza, że

$$F_X(a - \varepsilon) \leq P(\{|X_n - X| \geq \varepsilon\}) + F_{X_n}(a) \leq 2P(\{|X_n - X| \geq \varepsilon\}) + F_X(a + \varepsilon).$$

Ponieważ $P(\{|X_n - X| \geq \varepsilon\}) \rightarrow 0$, dla $n \rightarrow \infty$, więc dla każdego $\varepsilon > 0$ mamy

$$F_X(a - \varepsilon) \leq \liminf_n F_{X_n}(a) \leq \limsup_n F_{X_n}(a) \leq F_X(a + \varepsilon).$$

Przechodząc z ε do zera i korzystając z ciągłości F_X w punkcie a otrzymujemy w powyższym wzorze same równości, co oznacza istnienie granicy i równość $\lim_{n \rightarrow \infty} F_{X_n}(a) = F_X(a)$. \square

Ad 3. Dystrybuanta rozkładu skupionego w jednym punkcie c jest nieciągła tylko w punkcie c . Weźmy dwa punkty ciągłości dystrybuanty F_X , mianowicie punkty $c - \varepsilon$ oraz $c + \varepsilon$, gdzie $\varepsilon > 0$. Dostajemy

$$\begin{aligned} P(\{|X_n - c| \leq \varepsilon\}) &= P(c - \varepsilon \leq X_n \leq c + \varepsilon) \geq P(c - \varepsilon < X_n \leq c + \varepsilon) \\ &= F_{X_n}(c + \varepsilon) - F_{X_n}(c - \varepsilon) \rightarrow 1 - 0 = 1, \end{aligned}$$

co oznacza, że dla każdego $\varepsilon > 0$ mamy $\lim_{n \rightarrow \infty} P(\{|X_n - c| \leq \varepsilon\}) = 1$, a to daje stochastyczną zbieżność X_n do c . \square

Twierdzenie odwrotne do twierdzenia 10.7.2 nie jest prawdziwe.

Przykład – 10.8. Niech X oraz Y będą dwiema niezależnymi próbami Bernoulliego o prawdopodobieństwie sukcesu $\frac{1}{2}$ każda. Ciąg $X_n = X$ ma w sposób trywialny dystrybuanty zbieżne do dystrybuanty F_Y .

Z drugiej strony, z niezależności zmiennych losowych X oraz Y , $P(|X_n - Y| \geq 1) = \frac{1}{2}$, więc X_n nie są zbieżne stochastycznie do Y .

Twierdzenie odwrotne do twierdzenia 10.7.1 nie jest prawdziwe.

Przykład – 10.9. Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną taką, że $\Omega = [0, 1]$, Σ składa się ze zbiorów borelowskich zawartych w odcinku $[0, 1]$, a P jest miarą Lebesgue'a na tym odcinku. Rozważamy ciąg zmiennych losowych $X_{11}, X_{21}, X_{22}, X_{31}, \dots$, zdefiniowanych na Ω w sposób następujący:

$$X_{kl}(\omega) = \begin{cases} 1, & \text{dla } \frac{l-1}{k} < \omega \leq \frac{l}{k}, \\ 0, & \text{dla pozostałych } \omega, \end{cases}$$

gdzie $k = 1, 2, 3, \dots$, $l = 1, \dots, k$. Zobacz sam.

Dla dowolnego $0 < \varepsilon < 1$ widzimy, że $P(|X_{kl}| \geq \varepsilon) = \frac{1}{k}$. Tak więc nasz ciąg jest zbieżny stochastycznie do 0.

Dla każdego ustalonego ω ciąg $X_{kl}(\omega)$ zawiera nieskończenie wiele zer i nieskończenie wiele jedynek, więc nie jest ciągiem zbieżnym. Tym bardziej ciąg X_{kl} nie jest zbieżny prawie wszędzie do żadnej granicy.

10.2 Mocne Prawa Wielkich Liczb

Celem tego punktu jest przedstawienie Mocnego Prawa Wielkich Liczb, które jest odpowiednikiem Słabego Prawa Wielkich Liczb i opiera się na pojęciu zbieżności z prawdopodobieństwem 1. Pierwszym krokiem będzie nierówność Kołmogorowa, która w pewnym szczególnym przypadku wzmacnia nierówność Czebyszewa.

Twierdzenie – 10.10 (Nierówność Kołmogorowa). *Niech X_1, X_2, X_3, \dots będą niezależnymi zmiennymi losowymi, $E(X_i) \in \mathbb{R}$, dla $i = 1, 2, 3, \dots$. Ustalmy $\varepsilon > 0$. Wtedy*

$$\forall n \geq 1 : P\left(\max_{1 \leq k \leq n} |S_k - E(S_k)| \geq \varepsilon\right) \leq \frac{D^2(S_n)}{\varepsilon^2}.$$

Uwaga. Z nierówności Czebyszewa wynika istotnie mniej, mianowicie:

$$P(|S_n - E(S_n)| \geq \varepsilon) \leq \frac{D^2(S_n)}{\varepsilon^2}.$$

Tymczasem:

$$\{|S_n - E(S_n)| \geq \varepsilon\} \subset \bigcup_{k=1}^n \{|S_k - E(S_k)| \geq \varepsilon\} = \left\{ \max_{1 \leq k \leq n} |S_k - E(S_k)| \geq \varepsilon \right\}.$$

Dowód. Bez straty ogólności zakładamy, że wszystkie $E(X_i) = 0$. Definiujemy zdarzenia: $A = \{ \max_{1 \leq k \leq n} |S_k| \geq \varepsilon \}$ oraz $A_1 = \{|S_1| \geq \varepsilon\}$. $A_k = \{|S_i| < \varepsilon, i = 1, 2, \dots, k-1, |S_k| \geq \varepsilon\}$, dla $k = 2, \dots, n$. Widać, że: $A = \bigcup_{k=1}^n A_k$, oraz $A_i \cap A_j = \emptyset$ dla $i \neq j$.

Szacujemy:

$$D^2(S_n) = \int_{\Omega} S_n^2 dP \geq \int_A S_n^2 dP = \sum_{k=1}^n \int_{A_k} S_n^2 dP.$$

Ale $S_n = S_k + Y_k$. Zauważmy najpierw, że:

$$\int_{A_k} S_k Y_k dP = \int_{\Omega} \mathbb{1}_{A_k} S_k Y_k dP = E(\mathbb{1}_{A_k} S_k \cdot Y_k) = E(\mathbb{1}_{A_k} S_k) \cdot E(Y_k) = 0,$$

gdyż zmienne losowe $\mathbb{1}_{A_k} S_k$ oraz Y_k są niezależne jako funkcje wektorów niezależnych.

$$\int_{A_k} S_n^2 dP = \int_{A_k} (S_k + Y_k)^2 dP = \int_{A_k} S_k^2 dP + 2 \int_{A_k} S_k Y_k dP + \int_{A_k} Y_k^2 dP \geq \int_{A_k} S_k^2 dP \geq P(A_k) \varepsilon^2,$$

z określenia zdarzenia A_k . Ostatecznie: $D^2(S_n) \geq \sum_{k=1}^n P(A_k) \varepsilon^2 = P(A) \varepsilon^2$. □

Twierdzenie – 10.11 (Zbieżność szeregu). *Niech X_1, X_2, X_3, \dots będzie ciągiem niezależnych zmiennych losowych, takich, że:*

$$\sum_{i=1}^{\infty} D^2(X_i) < \infty.$$

Wtedy

$$\sum_{i=1}^{\infty} (X_i - E(X_i))$$

jest zbieżny z prawdopodobieństwem 1.

Dowód. Bez straty ogólności zakładamy, że $E(X_i) = 0$. Wystarczy pokazać, że dla każdego ω ze zbioru, którego prawdopodobieństwo = 1 spełniony jest warunek Cauchy'ego zbieżności sum częściowych $S_n(\omega)$. Chcemy więc pokazać, że:

$$P \left(\bigcap_{\varepsilon > 0} \bigcup_N \bigcap_{k, l \geq N} \{ \omega : |S_k(\omega) - S_l(\omega)| < \varepsilon \} \right) = 1.$$

Wystarczy więc pokazać, że:

$$\forall \varepsilon > 0 : P \left(\bigcup_N \bigcap_{k, l \geq N} |S_k - S_l| < \varepsilon \right) = 1.$$

Oznaczmy:

$$A_{N, \varepsilon} = \bigcap_{k, l \geq N} \{|S_k - S_l| < \varepsilon\}, \quad B_{N, \varepsilon} = \bigcap_{k \geq 1} \{|S_{N+k} - S_N| < \varepsilon\}.$$

Ponieważ zbiory $A_{N, \varepsilon}$ tworzą ciąg wstępujący, więc wystarczy pokazać, że:

$$\forall \varepsilon > 0 : \lim_{N \rightarrow \infty} P(A_{N, \varepsilon}) = 1.$$

Ponieważ $|S_k - S_l| \leq |S_k - S_N| + |S_l - S_N|$, to $B_{N, \frac{\varepsilon}{2}} \subset A_{N, \varepsilon}$. Wystarczy więc wykazać, że:

$$\boxed{\forall \varepsilon > 0 : \lim_{N \rightarrow \infty} P(B_{N, \varepsilon}) = 1.}$$

Ustalmy $\varepsilon > 0$ oraz $N < M$. Teraz, z nierówności Kołmogorowa:

$$P \left(\bigcup_{k=1}^M |S_{N+k} - S_N| \geq \varepsilon \right) = P \left(\max_{1 \leq k \leq M} |S_{N+k} - S_N| \geq \varepsilon \right) \leq \frac{D^2(S_M - S_N)}{\varepsilon^2}.$$

Inaczej:

$$P\left(\bigcup_{k=1}^M |S_{N+k} - S_N| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{k=N+1}^M D^2(X_k).$$

Niech $M \rightarrow \infty$. Wtedy lewa strona ma granicę (ciąg zbiorów wstępujących): $P\left(\bigcup_{k=1}^{\infty} |S_{N+k} - S_N| \geq \varepsilon\right)$, a prawa strona ma granicę: $\frac{1}{\varepsilon^2} \sum_{k=N+1}^{\infty} D^2(X_k)$. Zachodzi więc też nierówność:

$$P\left(\bigcup_{k=1}^{\infty} |S_{N+k} - S_N| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{k=N+1}^{\infty} D^2(X_k).$$

Niech $N \rightarrow \infty$. Wtedy prawa strona, a więc i lewa strona dążą do zera.

Ponieważ $\Omega \setminus B_{N,\varepsilon} = \bigcup_{k=1}^{\infty} \{|S_{N+k} - S_N| \geq \varepsilon\}$, otrzymujemy żadaną tezę:

$$\forall \varepsilon > 0 \lim_{N \rightarrow \infty} P(B_{N,\varepsilon}) = 1. \quad \square$$

Przykład – 10.12. Zbadamy zbieżność szeregu $\sum_{n=1}^{\infty} \frac{a_n}{n}$, gdzie a_n są niezależnymi zmiennymi losowymi o wspólnym rozkładzie: $P(a_n = -1) = \frac{1}{2}$, $P(a_n = 1) = \frac{1}{2}$. Ponieważ $D^2\left(\frac{a_n}{n}\right) = \frac{D^2(a_n)}{n^2} = \frac{\frac{1}{4}}{n^2}$, to szereg $\sum_{n=1}^{\infty} D^2\left(\frac{a_n}{n}\right)$ jest zbieżny. Z powyższego twierdzenia wynika, że $\sum_{n=1}^{\infty} \frac{a_n}{n}$ jest zbieżny z prawdopodobieństwem 1.

W dalszej części będziemy korzystać z dwóch faktów z analizy matematycznej.

Lemat – 10.13 (Toeplitz). Niech $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}$, $x \in \mathbb{R}$. Wtedy

$$\lim_{n \rightarrow \infty} x_n = x \implies \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i = x.$$

Dowód. Niech $\varepsilon > 0$. Istnieje takie n_1 , że dla $n \geq n_1$: $|x_n - x| < \frac{\varepsilon}{2}$. Istnieje takie $n_0 > n_1$, że dla $n \geq n_0$: $\frac{1}{n} \sum_{i=1}^{n_1} |x_i - x| < \frac{\varepsilon}{2}$. Niech $n \geq n_0$. $\left| \frac{1}{n} \sum_{i=1}^n x_i - x \right| \leq \frac{1}{n} \sum_{i=1}^{n_1} |x_i - x| + \frac{1}{n} \sum_{i=n_1+1}^n |x_i - x| \leq \frac{\varepsilon}{2} + \frac{n - n_1}{n} \frac{\varepsilon}{2} \leq \varepsilon. \quad \square$

Lemat – 10.14 (Kronecker). Niech $\{x_n\}_{n=1}^{\infty} \subset \mathbb{R}$. Wtedy: Szereg $\sum_{i=1}^{\infty} x_i$ jest zbieżny $\implies \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n ix_i = 0$.

Dowód. Oznaczmy: $s_0 = 0$, $s_n = x_1 + \dots + x_n$. Wtedy: $\sum_{i=1}^n ix_i = s_1 - s_0 + 2(s_2 - s_1) + \dots + n(s_n - s_{n-1}) = -s_0 - s_1 - s_2 - \dots - s_{n-1} + ns_n = -\sum_{i=1}^n s_{i-1} + ns_n$.

Niech $s = \sum_{i=1}^{\infty} x_i$. Z lematu Toeplitza:

$$\frac{1}{n} \sum_{i=1}^n ix_i = -\frac{1}{n} \sum_{i=1}^n s_{i-1} + s_n \rightarrow -s + s = 0. \quad \square$$

Jesteśmy przygotowani do dowodu jednego z dwóch głównych twierdzeń tego rozdziału.

Twierdzenie – 10.15 (Mocne Prawo Wielkich Liczb, MPWL). *Niech X_1, X_2, X_3, \dots będzie ciągiem niezależnych zmiennych losowych i niech szereg $\sum_{n=1}^{\infty} \frac{D^2(X_n)}{n^2}$ będzie zbieżny.*

Zdefiniujmy $S_n = X_1 + X_2 + \dots + X_n$.

Wtedy:

$$\frac{S_n - E(S_n)}{n} \xrightarrow{1} 0.$$

Dowód. Korzystamy z twierdzenia o zbieżności szeregu oraz z lematu Kroneckera.

Ponieważ szereg $\sum_{i=1}^{\infty} D^2\left(\frac{X_i - E(X_i)}{i}\right) = \sum_{i=1}^{\infty} D^2\left(\frac{X_i}{i}\right)$ jest zbieżny, więc szereg $\sum_{i=1}^{\infty} \frac{X_i - E(X_i)}{i}$ jest zbieżny z prawdopodobieństwem 1.

$$\text{Ale w takim razie: } \frac{S_n - E(S_n)}{n} = \frac{1}{n} \sum_{i=1}^n i \frac{X_i - E(X_i)}{i} \xrightarrow{1} 0. \quad \square$$

Założenie o wariancjach można opuścić, gdy się założy, że wszystkie zmienne losowe mają ten sam rozkład, czyli są i.i.d (independent, identically distributed)

Twierdzenie – 10.16 (Mocne Prawo Wielkich Liczb dla zmiennych i.i.d.). *Niech X_1, X_2, X_3, \dots będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie i skończonej wartości oczekiwanej m .*

Niech $S_n = X_1 + X_2 + \dots + X_n$.

Wtedy

$$\frac{S_n}{n} \xrightarrow{1} m.$$

Dowód polega na zastąpieniu ciągu $\{X_n\}$ innym ciągiem, który spełnia założenia poprzedniego twierdzenia, ale ma średnie tak samo zbieżne jak średnie $\{X_n\}$.

Wykażemy jednak wcześniej dwa pomocnicze lematy i przypomnimy podstawowe twierdzenie o zbieżności całek. Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną, $A_1, A_2, A_3, \dots \in \Sigma$. Interesuje nas zbiór:

$$A = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i.$$

Zauważmy, że:

$$\omega \in A \iff \omega \text{ należy do nieskończenie wielu spośród zdarzeń } A_i.$$

Lemat – 10.17 (Lemat Borela–Cantellego).

$$(1) \text{ Jeżeli } \sum_{i=1}^{\infty} P(A_i) < \infty, \text{ to } P(A) = 0.$$

$$(2) \text{ Jeżeli } \sum_{i=1}^{\infty} P(A_i) = \infty \text{ oraz } A_1, A_2, A_3, \dots \text{ są niezależne, to } P(A) = 1.$$

Dowód. Ad 1. Dla każdego n : $P(A) \leq P\left(\bigcup_{i=n}^{\infty} A_i\right) \leq \sum_{i=n}^{\infty} P(A_i)$. Ponieważ jednak szereg $\sum_{i=1}^{\infty} P(A_i)$ jest zbieżny, to jego „końcówka” $\sum_{i=n}^{\infty} P(A_i) \rightarrow 0$, gdy $n \rightarrow \infty$. Stąd $P(A) = 0$. □

Ad 2. Zauważmy najpierw że korzystając z założenia o niezależności oraz ze standardowej nierówności $1 + x \leq e^x$ dla każdego $x \in \mathbb{R}$, otrzymujemy dla wszystkich $n < m$:

$$P\left(\bigcap_{i=n}^m (\Omega \setminus A_i)\right) = \prod_{i=n}^m P(\Omega \setminus A_i) = \prod_{i=n}^m (1 - P(A_i)) \leq \prod_{i=n}^m e^{-P(A_i)} = e^{-\sum_{i=n}^m P(A_i)}.$$

Ponieważ szereg $\sum_{i=1}^{\infty} P(A_i)$ jest rozbieżny, więc dla każdego ustalonego n :

$$P\left(\bigcap_{i=n}^{\infty} (\Omega \setminus A_i)\right) = \lim_{m \rightarrow \infty} P\left(\bigcap_{i=n}^m (\Omega \setminus A_i)\right) \leq \lim_{m \rightarrow \infty} e^{-\sum_{i=n}^m P(A_i)} = 0.$$

i stąd kolejno:

$$P\left(\bigcup_{i=n}^{\infty} A_i\right) = 1 \text{ oraz } P\left(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i\right) = 1. \quad \square$$

Lemat – 10.18. Niech Y będzie zmienną losową nieujemną, czyli $P(Y \geq 0) = 1$. Wtedy

$$\sum_{n=1}^{\infty} P(Y \geq n) \leq E(Y) \leq 1 + \sum_{n=1}^{\infty} P(Y \geq n).$$

Dowód. Udowodnimy pierwszą nierówność. Dowód drugiej – ćwiczenie. Mamy kolejno:

$$\begin{aligned} \sum_{n=1}^{\infty} P(Y \geq n) &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P(k \leq Y < k+1) \\ &= \sum_{k=1}^{\infty} \sum_{n=1}^k P(k \leq Y < k+1) = \sum_{k=1}^{\infty} k P(k \leq Y < k+1) \\ &= \sum_{k=0}^{\infty} \int_{\{k \leq Y < k+1\}} k dP \leq \sum_{k=0}^{\infty} \int_{\{k \leq Y < k+1\}} Y dP = E(Y). \quad \square \end{aligned}$$

Przypominamy podstawowe twierdzenie o przechodzeniu do granicy pod znakiem całki.

Twierdzenie – 10.19 (Twierdzenie Lebesgue'a). Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną, $X_1, X_2, X_3, \dots, X, Z$ zmiennymi losowymi określonymi na Ω .

$$1. 0 \leq X_1 \leq X_2 \leq X_3 \leq \dots \text{ pw.}, X_n \xrightarrow{1} X \implies E(X_n) \nearrow E(X).$$

$$2. |X_n| \leq Z, E(Z) \in \mathbb{R}, X_n \xrightarrow{1} X \implies E(X_n) \longrightarrow E(X).$$

Wracamy do Mocnego Prawa Wielkich Liczb.

Dowód twierdzenia 10.16. Definiujemy nowe zmienne losowe:

$$Y_n = \begin{cases} X_n, & \text{gdym } |X_n| < n, \\ 0, & \text{gdym } |X_n| \geq n. \end{cases}$$

Wykażemy najpierw, że ciąg $\{Y_n\}$ spełnia założenia Mocnego Prawa Wielkich Liczb, twierdzenie 10.15. Zauważmy najpierw, że: $Y_n = \mathbb{1}_{\{|X_n| < n\}} \cdot X_n$ gdzie $\mathbb{1}_A$ oznacza funkcję charakterystyczną (indykator) zbioru A . W związku z tym zmienne losowe Y_n są niezależne jako funkcje zmiennych losowych niezależnych.

Wykażemy, że $\sum_{n=1}^{\infty} \frac{D^2(Y_n)}{n^2} < \infty$. Mamy kolejno:

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{D^2(Y_n)}{n^2} &\leq \sum_{n=1}^{\infty} \frac{E(Y_n^2)}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} E(\mathbb{1}_{\{|X_n| < n\}} \cdot X_n^2) = \sum_{n=1}^{\infty} \frac{1}{n^2} E(\mathbb{1}_{\{|X_1| < n\}} \cdot X_1^2) \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^n E(\mathbb{1}_{\{k-1 \leq |X_1| < k\}} \cdot X_1^2) = \sum_{k=1}^{\infty} E(\mathbb{1}_{\{k-1 \leq |X_1| < k\}} \cdot X_1^2) \sum_{n=k}^{\infty} \frac{1}{n^2} \\ &\leq \sum_{k=1}^{\infty} E(\mathbb{1}_{\{k-1 \leq |X_1| < k\}} \cdot |X_1|) \cdot k \cdot \left(\frac{1}{k} + \frac{1}{k^2}\right) \leq \sum_{k=1}^{\infty} 2E(\mathbb{1}_{\{k-1 \leq |X_1| < k\}} \cdot |X_1|) = 2E(|X_1|) < \infty. \end{aligned}$$

Skorzystaliśmy tutaj z nierówności:

$$\sum_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{1}{k^2} + \frac{1}{k(k+1)} + \frac{1}{(k+1)(k+2)} + \dots = \frac{1}{k^2} + \frac{1}{k}.$$

Wykazaliśmy więc, że zmienne losowe Y_N spełniają Mocne Prawo Wielkich Liczb. Czyli:

$$\frac{\sum_{i=1}^n Y_i - \sum_{i=1}^n E(Y_i)}{n} \xrightarrow{1} 0. \quad (\text{MPWL})$$

Niech $A_n = \{X_n \neq Y_n\}$. Oczywiście $A_n \subset \{|X_n| \geq n\}$. Mamy więc:

$$\sum_{n=1}^{\infty} P(A_n) \leq \sum_{n=1}^{\infty} P(|X_n| \geq n) = \sum_{n=1}^{\infty} P(|X_1| \geq n) \leq E(|X_1|) < \infty.$$

Z lematu Borela–Cantellego $P\left(\bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} A_i\right) = 0$. Czyli $P\left(\bigcup_{n=0}^{\infty} \bigcap_{i=n}^{\infty} \{X_i = Y_i\}\right) = 1$.

Inaczej: $P(X_i = Y_i, \text{ dla prawie wszystkich } i) = 1$.

W szczególności:

$$\frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n Y_i}{n} \xrightarrow{1} 0. \quad (\text{B-C})$$

Zauważmy też, że:

$$\frac{1}{n} \sum_{i=1}^n E(Y_i) \rightarrow m. \quad (\text{L+T})$$

Rzeczywiście, korzystając z twierdzenia Lebesgue’a wiemy, że $E(\mathbb{1}_{\{|X_1| < i\}} X_1) \rightarrow m$, gdy $i \rightarrow \infty$. Mamy więc

$$\frac{1}{n} \sum_{i=1}^n E(Y_i) = \frac{1}{n} \sum_{i=1}^n E(\mathbb{1}_{\{|X_i| < i\}} X_i) = \frac{1}{n} \sum_{i=1}^n E(\mathbb{1}_{\{|X_1| < i\}} X_1) \rightarrow m, \text{ co wynika z lematu Toeplitza.}$$

Wykorzystując kolejno wzory (B–C), (MPWL), (L+T), mamy:

$$\frac{\sum_{i=1}^n X_i}{n} = \frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n Y_i}{n} + \frac{\sum_{i=1}^n Y_i - \sum_{i=1}^n E(Y_i)}{n} + \frac{1}{n} \sum_{i=1}^n E(Y_i) \xrightarrow{1} 0 + 0 + m = m. \quad \square$$

Jako natychmiastową konsekwencję mocnego prawa wielkich liczb otrzymujemy:

Wniosek – 10.20.

1. Niech X_1, X_2, X_3, \dots będzie ciągiem niezależnych prób Bernoulliego o prawdopodobieństwie sukcesu p ($X_i \sim B(1, p)$). Wtedy

$$\frac{S_n}{n} \xrightarrow{1} p.$$

2. Niech $A \in \Sigma$ i niech X_1, X_2, X_3, \dots będzie ciągiem niezależnych prób Bernoulliego o prawdopodobieństwie sukcesu $p = P(A)$. Wtedy

$$\frac{S_n}{n} \xrightarrow{1} P(A).$$

Komentarz, Przybliżone wyznaczanie nadziei. W wielu przypadkach chcemy poznać nadzieję matematyczną $m = E(X)$ zmiennej losowej X , ale analitycznie jest to trudne lub niemożliwe. Tymczasem, zarówno słabe prawo wielkich liczb jak i pośrednio centralne twierdzenie graniczne mówią, że m jest przybliżana w określonym sensie przez średnie $\frac{S_n}{n}$ ciągu niezależnych zmiennych losowych X_1, \dots, X_n mających ten sam rozkład co zmienna losowa X .

Niekiedy potrafimy generować na komputerze wielkości x_1, \dots, x_n , które można traktować jako niezależne realizacje zmiennej X , czyli $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$ dla pewnego $\omega \in \Omega$, gdzie X_1, \dots, X_n , są niezależnymi zmiennymi losowymi mającymi taki sam rozkład jak X .

Właśnie z mocnego prawa wielkich liczb wynika, że prawie zawsze dla dużych n średnia $\bar{x}_n = \frac{\sum_{i=1}^n x_i}{n}$ będąca realizacją $\frac{S_n}{n}$ jest blisko m .

Dla każdego zdarzenia A jego prawdopodobieństwo $P(A)$ jest równe wartości oczekiwanej $E(\mathbb{1}_A)$, gdzie $\mathbb{1}_A$ jest funkcją charakterystyczną zbioru A . W związku z tym mocne prawo wielkich liczb może być użyte także do szacowania $P(A)$ przy pomocy komputera.

Konkluzja. Aby wyznaczyć przybliżoną wartość nadziei matematycznej $m = E(X)$ zmiennej losowej X (prawdopodobieństwo zdarzenia A) wystarczy wygenerować odpowiednio dużo niezależnych realizacji tej zmiennej (zmiennej $\mathbb{1}_A$), a ich średnia jest poszukiwaną wielkością.

Co oznacza zwrot „odpowiednio dużo” będzie częściowo wyjaśnione w trakcie omawiania metod Monte Carlo.

Przykład – 10.21. Leon wchodzi do apteki i widzi, że jest dziesiąty w kolejce oraz, że czynne są cztery okienka obsługujące aktualnie klientów, Jego poprzednie obserwacje wskazują, że jedna osoba jest obsługiwana około 1–5 minut (przypuszcza więc, że czas obsługi jednego klienta X ma rozkład $N(3, 1)$). Zakłada, że czas podejścia pierwszej osoby z kolejki do zwalnianego okienka ma rozkład jednostajny $U(0, m_t)$, gdzie m_t jest minimalnym czasem obsługi osób aktualnie stojących przy okienkach. Ile czasu spędzi Leon w kolejce?

Czas oczekiwania Leona T jest zmienną losową, którą można wyrazić za pomocą sumy i minimów innych zmiennych losowych i w tej sytuacji trudno jest wyznaczyć bezpośrednio jej nadzieję. Można jednak łatwo wygenerować, powiedzmy 10 000 niezależnych realizacji zmiennej T . Ich średnia przybliży średni czas oczekiwania naszego klienta. W kilku wykonanych próbach średnie te wynosiły: 7.55012877309712, 7.52607461327209, 7.52962450258854, 7.53444237897340, 7.52881780301394. Wynik ten można było z pewnym przybliżeniem przewidzieć. Gdyby w aptece było czynne tylko jedno okienko i musiało ono obsłużyć 9 klientów stojących przed Leonem, a wcześniej skończyć obsługę poprzedniego klienta, to zajęłoby to średnio $1.5 + 27 = 28.5$ minut. Gdy są czynne 4 okienka, to obsługa trwa około 4 razy krócej.

Przykład – 10.22. Oszacujmy prawdopodobieństwo zbioru A występującego w nierówności Kołmogorowa:

$$A = \left\{ \max_{1 \leq k \leq n} |S_k - E(S_k)| \geq \varepsilon \right\}.$$

Załóżymy tutaj, że $S_k = X_1 + \dots + X_k$, gdzie X_1, \dots, X_n są niezależnymi zmiennymi losowymi o jednakowym rozkładzie, a ε oraz n są znane. Przy okazji przybliżymy (łatwiejsze do wyliczenia analitycznie) prawdopodobieństwo $P(A_n)$, gdzie

$$A_n = \{|S_n - E(S_n)| \geq \varepsilon\}$$

oraz wartość prawej strony nierówności Kołmogorowa: $\frac{D^2(S_n)}{\varepsilon^2}$.

Ponieważ chcemy znaleźć wartości średnie $E(\mathbb{1}_A)$ oraz $E(\mathbb{1}_{A_n})$ będziemy generować N realizacji zmiennych losowych $\mathbb{1}_A$ oraz $\mathbb{1}_{A_n}$ i wyliczać ich średnie. Przykładowo otrzymaliśmy:

Gdy X_i mają rozkład jednostajny $U(-3, 5)$, $n = 10$, $\varepsilon = 15$, $N = 100$:

$$P(A) \approx \frac{3}{100}, \quad P(A_n) \approx \frac{2}{100}, \quad \frac{D^2(S_n)}{\varepsilon^2} = \frac{32}{135}.$$

Gdy X_i mają rozkład Poissona P_5 , $n = 200$, $\varepsilon = 40$, $N = 100$:

$$P(A) \approx \frac{32}{100}, \quad P(A_n) \approx \frac{16}{100}, \quad \frac{D^2(S_n)}{\varepsilon^2} = \frac{5}{8}.$$

Gdy X_i mają rozkład normalny $N(0, 2)$, $n = 200$, $\varepsilon = 70$, $N = 1000$:

$$P(A) \approx \frac{29}{1000}, \quad P(A_n) \approx \frac{17}{1000}, \quad \frac{D^2(S_n)}{\varepsilon^2} = \frac{8}{49}.$$

Mocne Prawo Wielkich Liczb ma szereg innych ważnych konsekwencji. W szczególności stanowi podstawę całej statystyki. O jego kluczowym znaczeniu w metodach Monte Carlo powiemy jeszcze więcej później. Teraz podamy pewne inne zastosowanie.

Przykład – 10.23. Definiujemy ciąg x_n ; $x_0 = 1$, x_{n+1} – liczba wylosowana zgodnie z rozkładem $U(0, 2x_n)$, $n = 0, 1, 2, 3, \dots$. Badamy jego zbieżność.

Niech X_n oznacza zmienną losową, której realizacją jest x_n .

$X_0 := 1$, $X_1 := 2X_0 \cdot U_1 = 2U_1$, $X_2 := 2X_1 \cdot U_2 = 2^2U_1 \cdot U_2, \dots, X_n = 2^n U_1 \cdot \dots \cdot U_n$, gdzie U_1, \dots, U_n są ciągiem i.i.d. zmiennych o rozkładzie $U(0, 1)$, a więc:

$$\ln X_n = n \ln 2 + \sum_{i=1}^n \ln U_i = \sum_{i=1}^n (\ln 2 + \ln U_i).$$

Łatwo obliczyć (ćwiczenie), że $m = E(\ln 2 + \ln U_i) = \ln 2 - 1 < 0$.

Stąd, oznaczając $Y_i = \ln 2 + \ln U_i$, mamy: $X_n = e^{\sum_{i=1}^n Y_i} = e^{n \frac{1}{n} \sum_{i=1}^n Y_i}$.

MPWL zapewnia, że $\frac{1}{n} \sum_{i=1}^n Y_i \rightarrow m < 0$ w prawie każdym punkcie, a więc $X_n \xrightarrow{1} 0$.

Komentarz. Ze zbieżności $X_n \xrightarrow{1} X$ nie zawsze wynika zbieżność $E(X_n) \rightarrow E(X)$. Taką zbieżność gwarantuje twierdzenie Lebesgue'a, o ile jednak są spełnione pewne założenia. W powyższym przykładzie $X_n \xrightarrow{1} 0$, natomiast stosując indukcję widać, że $E(X_n) = E(X_0) = 1$, więc nasz ciąg nie mógł spełniać tych założeń. Zbadajmy to dokładniej.

Ciąg X_n nie jest monotoniczny. Trudno byłoby też wskazać funkcję sumowalną ograniczającą wszystkie X_n z góry, gdyż nie mogą być one ograniczone od góry przez żadną stałą:

Niech $A_i = \{\omega : U_i(\omega) > \frac{3}{4}\}$. Wtedy $X_n(\omega) > (\frac{3}{2})^n$ na zbiorze $\bigcap_{i=1}^n A_i$.

Ponieważ, jak widzimy, nie jest spełniona teza twierdzenia Lebesgue'a, to faktycznie nie istnieje funkcja sumowalna ograniczająca z góry wszystkie X_n .

Przy okazji zauważmy, że:

$$D^2(X_n) = E(X_n^2) - E(X_n)^2 = E((2^n U_1 \cdot \dots \cdot U_n)^2) - 1 = 4^n E(U_1^2)^n - 1 = \left(\frac{4}{3}\right)^n - 1.$$

10.3 Pytania

Pytanie 10.1. Wykazać uwagę 10.1. ROZWIĄZANIE

Pytanie 10.2. Wykazać, że granica stochastycznie zbieżnego ciągu zmiennych losowych jest wyznaczona jednoznacznie pw.. ROZWIĄZANIE

Pytanie 10.3. Niech X_1, X_2, X_3, \dots będzie ciągiem takich niezależnych zmiennych losowych, że $P_{X_i} = B(1, p_i)$ dla $i = 1, 2, 3, \dots$. Wykazać, że:

(1) $X_n \xrightarrow{s} 0 \iff \lim_{n \rightarrow \infty} p_n = 0$.

(2) $X_n \xrightarrow{1} 0 \iff \sum_{n=1}^{\infty} p_n < \infty$. ROZWIĄZANIE

Pytanie 10.4. Niech $X_n \xrightarrow{s} X$ i niech $f : \mathbb{R} \rightarrow \mathbb{R}$ będzie funkcją jednostajnie ciągłą. Wykazać, że $f(X_n) \xrightarrow{s} f(X)$. Jak można osłabić założenia, gdy wiemy, że $X = c \in \mathbb{R}$? ROZWIĄZANIE

Pytanie 10.5. Niech $X_n \xrightarrow{1} X$ i niech $f : \mathbb{R} \rightarrow \mathbb{R}$. Zaproponować założenie dotyczące f gwarantujące, że $f(X_n) \xrightarrow{1} f(X)$. ROZWIĄZANIE

Pytanie 10.6. Niech X_1, X_2, X_3, \dots będzie ciągiem niezależnych zmiennych losowych o takim samym rozkładzie. Wykazać, że jeżeli ciąg $\frac{S_n}{n}$ jest zbieżny z prawdopodobieństwem 1, to zmienne X_i mają skończoną nadzieję matematyczną m i $\frac{S_n}{n} \xrightarrow{1} m$. ROZWIĄZANIE

Rozdział 11

Zbieżność rozkładów i funkcje charakterystyczne

Przypominamy, że zdefiniowaliśmy już zbieżność ciągu dystrybuant, patrz definicja 10.5:

Niech F_n , $n = 1, 2, 3, \dots$, oraz F będą dystrybuantami.

Zbieżność rozkładów. $F_n \xrightarrow{d} F \iff$

$$\forall a \in \mathbb{R}, \text{ punktu ciągłości } F, \lim_{n \rightarrow \infty} F_n(a) = F(a).$$

Wtedy: $X_n \xrightarrow{d} X \iff F_{X_n} \xrightarrow{d} F_X$.

Możemy mówić o zbieżności samych rozkładów. Można postawić następującą definicję.

Niech P_n , $n = 1, 2, 3, \dots$, oraz P będą jednowymiarowymi rozkładami.

Definicja – 11.1 (Zbieżność rozkładów).

$$P_n \xrightarrow{d} P \iff F_{P_n} \xrightarrow{d} F_P.$$

11.1 Charakteryzacja zbieżności ciągu rozkładów

Okazuje się, że powyżej sformułowana definicja zbieżności ciągu rozkładów może być sformułowana równoważnie na kilka innych sposobów, co nieraz jest wygodne.

Zanim sformułujemy i udowodnimy odpowiednie twierdzenia, przytoczymy ważny rezultat pomocniczy.

Twierdzenie – 11.2 (twierdzenie o wyborze). *Niech F_1, F_2, F_3, \dots będzie dowolnym ciągiem dystrybuant. Wtedy istnieje podciąg tego ciągu $F_{k_1}, F_{k_2}, F_{k_3}, \dots$ oraz funkcja $F : \mathbb{R} \rightarrow \mathbb{R}$ taka, że dla każdego x punktu ciągłości F*

$$\lim_{n \rightarrow \infty} F_{k_n}(x) = F(x)$$

oraz

1. $0 \leq F(x) \leq 1$, dla każdego $x \in \mathbb{R}$,
2. F jest funkcją niemalejącą: $x < y \implies F(x) \leq F(y)$,
3. F jest prawostronnie ciągła: $\lim_{x \rightarrow a^+} F(x) = F(a)$, dla każdego $a \in \mathbb{R}$.

Funkcja F nie musi być dystrybuantą. Na przykład, biorąc dystrybuanty rozkładów jednopunktowych, F_{δ_n} , dla $n = 1, 2, 3, \dots$, widzimy, że funkcja $F = 0$.

Dowód. (szkic) Wybieram dowolny zbiór $D \subset \mathbb{R}$, który jest przeliczalny i gęsty w \mathbb{R} . Iloczyn kartezjański $[0, 1]^D$ jest zbiorem zwartym, bo $[0, 1]$ jest zwarty. Zacieśnienia $F_n|_D$ są elementami tego zbioru, więc istnieje podciąg k_n oraz $F|_D \in [0, 1]^D$, takie, że $\forall x \in D : F_{k_n}(x) \rightarrow F|_D(x)$, gdy $n \rightarrow \infty$. Określamy: $F : \mathbb{R} \rightarrow \mathbb{R}$ przez warunek $F(x) = \inf\{F|_D(y) : y > x, y \in D\}$. Można teraz sprawdzić, że F oraz F_{k_n} spełniają wszystkie żądane warunki. \square

Niech \mathcal{P} oznacza ustaloną rodzinę rozkładów.

Definicja – 11.3 (warunek Prochorowa). Mówimy, że \mathcal{P} spełnia warunek Prochorowa (rodzina \mathcal{P} jest ścisła) $\iff \forall \varepsilon > 0 \exists K$ zwarty $\subset \mathbb{R} \forall P \in \mathcal{P} : P(K) \geq 1 - \varepsilon$.

Rodzina jednoelementowa $\mathcal{P} = \{P\}$ spełnia warunek Prochorowa: dla $\varepsilon > 0$ dobieramy tak $a < b$, że $F_P(a) < \frac{\varepsilon}{2}$ oraz $F_P(b) > 1 - \frac{\varepsilon}{2}$. Niech $a' < a$. Wtedy $P[a', b] \geq P(a, b) = F_P(b) - F_P(a) \geq 1 - \varepsilon$.

Łatwo widać, że jeżeli dwie rodziny rozkładów spełniają warunek Prochorowa, to ich suma też spełnia ten warunek. Przez indukcję przenosi się to na skończoną liczbę rodzin rozkładów.

Rodzina rozkładów jednopunktowych $\mathcal{P} = \{\delta_n : n = 0, 1, 2, \dots\}$ nie spełnia warunku Prochorowa: dla dowolnego zbioru zwartego K , $P_{\delta_n}(K) = 0$ dla prawie wszystkich n .

Podobnie rodzina $\{N(m, 1) : m \in \mathbb{R}\}$ oraz rodzina $\{U(a, b) : a < b < 100\}$ nie spełnia warunku Prochorowa. Natomiast rodzina $\{N(0, \sigma) : 0 < \sigma < 1\}$ spełnia ten warunek (ćwiczenie).

Twierdzenie – 11.4. Niech $\{P_n\}$ będzie ciągiem rozkładów, P rozkładem, takimi, że: $\forall f : \mathbb{R} \rightarrow \mathbb{R}$ ciągłej o suporcie zwartym $\int_{\mathbb{R}} f dP_n \rightarrow \int_{\mathbb{R}} f dP$, gdy $n \rightarrow \infty$.

Wtedy rodzina $\{P_n\}$ spełnia warunek Prochorowa.

Dowód. Niech $\varepsilon > 0$. Istnieje takie, N , że $P[-N, N] \geq 1 - \frac{\varepsilon}{2}$. Biorę taką funkcję ciągłą $g : \mathbb{R} \rightarrow \mathbb{R}$, że $g(x) = 1$ dla $|x| \leq N$, $g(x) = 0$ dla $|x| \geq N + 1$ oraz g jest afiniczna na przedziałach $[-N - 1, -N]$, $[N, N + 1]$. $\int_{\mathbb{R}} g dP \geq \int_{\mathbb{R}} \mathbb{1}_{[-N, N]} dP = P[-N, N] \geq 1 - \frac{\varepsilon}{2}$. Istnieje n_0 , takie, że dla $n \geq n_0 : \int_{\mathbb{R}} g dP_n \geq 1 - \varepsilon$. Wtedy

$$P_n[-N - 1, N + 1] = \int_{\mathbb{R}} \mathbb{1}_{[-N - 1, N + 1]} dP_n \geq \int_{\mathbb{R}} g dP_n \geq 1 - \varepsilon. \quad \square$$

Twierdzenie – 11.5 (o względnej zwartości). Niech rodzina \mathcal{P} spełnia warunek Prochorowa. Wtedy dla każdego ciągu $\{P_n\} \subset \mathcal{P}$ istnieje podciąg k_n oraz rozkład P taki, że $P_{k_n} \xrightarrow{d} P$, gdy $n \rightarrow \infty$.

Dowód. Biorę ciąg k_n oraz funkcję F z twierdzenia 11.2 o wyborze zastosowanego do $F_n = F_{P_n}$. Z monotoniczności F wynika, że istnieją granice, powiedzmy $F(-\infty) = \lim_{x \rightarrow -\infty} F(x)$, $F(\infty) = \lim_{x \rightarrow \infty} F(x)$. Pytamy czy $F(\infty) - F(-\infty) = 1$. Niech $\varepsilon > 0$. Istnieje zbiór zwarty $K \subset \mathbb{R}$, taki, że dla każdego $k_n : P_{k_n}(K) \geq 1 - \varepsilon$. Biorę a, b – punkty ciągłości F , takie, że $(a, b] \supset K$ (można je znaleźć). $F_{k_n}(b) - F_{k_n}(a) = P_{k_n}(a, b] \geq P_{k_n}(K) \geq 1 - \varepsilon$.

W granicy: $F(b) - F(a) \geq 1 - \varepsilon$. Więc: $F(\infty) - F(-\infty) \geq F(b) - F(a) \geq 1 - \varepsilon$. \square

Naszym celem jest udowodnienie równoważności czterech warunków charakteryzujących na różne sposoby zbieżność ciągu rozkładów, twierdzenie 11.11. Dowód będzie przebiegał etapowo. Na początku wykażemy:

Twierdzenie – 11.6. $P_n \xrightarrow{d} P \implies \forall a < b, P(a) = P(b) = 0 : P_n(a, b] \rightarrow P(a, b]$, gdy $n \rightarrow \infty$.

Dowód. Warunek $P(a) = P(b) = 0$ oznacza, że F_P jest ciągła w tych punktach, więc $P_n(a, b] = F_{P_n}(b) - F_{P_n}(a) \rightarrow F_P(b) - F_P(a) = P(a, b]$. \square

Lemat – 11.7. Niech P_1, P_2 będą rozkładami. Jeżeli dla każdej funkcji $f : \mathbb{R} \rightarrow \mathbb{R}$ ciągłej o suporcie zwartym¹ $\int_{\mathbb{R}} f dP_1 = \int_{\mathbb{R}} f dP_2$, to $P_1 = P_2$.

Dowód. Niech $a < b$. Określamy ciąg funkcji ciągłych:

$$g_n(x) = \begin{cases} 1, & \text{dla } a + \frac{1}{n} \leq x \leq b \\ 0, & \text{dla } x \leq a \text{ lub } x \geq b + \frac{1}{n} \\ \text{afiniczna,} & \text{na pozostałych przedziałach.} \end{cases}$$

Ponieważ $\mathbb{1}_{(a, b]} = \lim_{n \rightarrow \infty} g_n$, więc z twierdzenia Lebesgue'a:

$$P_1(a, b] = \int_{\mathbb{R}} \mathbb{1}_{(a, b]} dP_1 = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} g_n dP_1 = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} g_n dP_2 = \int_{\mathbb{R}} \mathbb{1}_{(a, b]} dP_2 = P_2(a, b].$$

$F_{P_1}(b) - F_{P_1}(a) = F_{P_2}(b) - F_{P_2}(a)$ i przechodząc z a do $-\infty$ otrzymujemy równość dystrybuant, a więc i rozkładów. \square

¹Istnieje taki zbiór zwarty K , że $f(x) = 0$ dla $x \notin K$.

Twierdzenie – 11.8. Jeżeli $\forall a < b$, $P(a) = P(b) = 0 : P_n(a, b] \rightarrow P(a, b]$, gdy $n \rightarrow \infty$, to $\forall f : \mathbb{R} \rightarrow \mathbb{R}$, f ciągła i ograniczona,

$$\int_{\mathbb{R}} f dP_n \rightarrow \int_{\mathbb{R}} f dP,$$

gdy $n \rightarrow \infty$.

Lemat – 11.9. Zbiór punktów $\{a \in \mathbb{R} : P(a) > 0\}$ jest co najwyżej przeliczalny.

Jest to uogólnienie twierdzenia mówiącego, że rozkład dyskretny jest skupiony na zbiorze co najwyżej przeliczalnym, patrz uwaga 4.7. Dowód jest ten sam (ćwiczenie).

Dowód twierdzenia 11.8. Krok I. Ustalę $a < b$ takie, że $P(a) = P(b) = 0$ oraz funkcję ciągłą i ograniczoną $f : [a, b] \rightarrow \mathbb{R}$. Wykażemy, że:

$$\int_a^b f dP_n \rightarrow \int_a^b f dP.$$

f jest jednostajnie ciągła na $[a, b]$, więc korzystając z prostego lematu:

Dla każdego $\delta > 0$ istnieje podział $a = x_0 < x_1 < \dots < x_N = b$ taki, że dla każdego $k = 1, \dots, N$ oraz $x \in [x_{k-1}, x_k]$ mamy $|f(x) - f(x_k)| \leq \delta$, przy czym $P(x_0) = \dots = P(x_N) = 0$.

Ustalę $\varepsilon > 0$ i dla $\delta = \frac{\varepsilon}{3}$ rozważę powyższy podział $a = x_0 < x_1 < \dots < x_N = b$. Zdefiniujemy pomocniczą funkcję g

$$g(x) = f(x_k) \quad \text{dla } x \in (x_{k-1}, x_k], \quad k = 1, \dots, N.$$

Z nierówności trójkąta otrzymujemy oszacowanie

$$\left| \int_a^b f dP_n - \int_a^b f dP \right| \leq I_1 + I_2 + I_3,$$

gdzie

$$I_1 = \int_a^b |f - g| dP_n, \quad I_2 = \left| \int_a^b g dP_n - \int_a^b g dP \right|, \quad I_3 = \int_a^b |f - g| dP.$$

Całki I_1 oraz I_3 są szacowane z góry odpowiednio przez $\delta P_n[a, b]$ oraz $\delta P[a, b]$, a więc każda z nich jest nie większa niż δ . Natomiast, ponieważ g jest funkcją schodkową,

$$\begin{aligned} I_2 &= \left| \sum_{k=1}^N f(x_k) P_n(x_{k-1}, x_k] - \sum_{k=1}^N f(x_k) P(x_{k-1}, x_k] \right| \\ &= \left| \sum_{k=1}^N f(x_k) [P_n(x_{k-1}, x_k] - P(x_{k-1}, x_k)] \right| \end{aligned}$$

i zmierza do 0, gdy $n \rightarrow \infty$, więc jest $\leq \delta$ dla dużych n . Dla takich n : $\left| \int_a^b f dP_n - \int_a^b f dP \right| \leq \varepsilon$. \square

Krok 2. Dowodzimy teraz właściwej tezy. Niech $M = \sup_{\mathbb{R}} |f|$. Ustalę $\varepsilon > 0$ i dobierzmy liczby $a' < b'$ tak, aby $P(a') = P(b') = 0$ oraz $M P(\mathbb{R} \setminus (a', b')) < \frac{\varepsilon}{4}$.

Ponieważ $P_n(a', b'] \rightarrow P(a', b']$, dla $n \rightarrow \infty$, bierzemy n tak duże, że $M P_n(\mathbb{R} \setminus (a', b']) \leq \frac{\varepsilon}{4}$. Niech $a < a'$ oraz $b > b'$ będą punktami takimi, że $P(a) = P(b) = 0$. Mamy teraz

$$\begin{aligned} \left| \int_{\mathbb{R}} f dP_n - \int_{\mathbb{R}} f dP \right| &\leq \left| \int_a^b f dP_n - \int_a^b f dP \right| + \left| \int_{\mathbb{R} \setminus (a, b)} f dP_n - \int_{\mathbb{R} \setminus (a, b)} f dP \right| \\ &\leq \left| \int_a^b f dP_n - \int_a^b f dP \right| + M P_n(\mathbb{R} \setminus (a, b)) + M P(\mathbb{R} \setminus (a, b)) \leq \varepsilon, \end{aligned}$$

dla dużych n . \square

Twierdzenie – 11.10. Niech $\{P_n\}$ będzie ciągiem rozkładów, P rozkładem, takimi, że: $\forall f : \mathbb{R} \rightarrow \mathbb{R}$ ciągłej o suporcie zwartym $\int_{\mathbb{R}} f dP_n \rightarrow \int_{\mathbb{R}} f dP$, gdy $n \rightarrow \infty$.
Wtedy $P_n \xrightarrow{d} P$.

Dowód. Z twierdzenia 11.4 wynika, że ciąg $\{P_n\}$ spełnia warunek Prochorowa. Z twierdzenia o względnej zwartości dostajemy podciąg k_n oraz rozkład, powiedzmy Q , taki, że $P_{k_n} \xrightarrow{d} Q$. Z twierdzenia 11.8: $\int_{\mathbb{R}} f dP_{k_n} \rightarrow \int_{\mathbb{R}} f dQ$.

Ponieważ także: $\int_{\mathbb{R}} f dP_{k_n} \rightarrow \int_{\mathbb{R}} f dP$, więc:

$$\text{dla każdej funkcji ciągłej o suporcie zwartym zachodzi: } \int_{\mathbb{R}} f dQ = \int_{\mathbb{R}} f dP.$$

Z lematu 11.7, $Q = P$. Dowodzimy dalej nie wprost. Gdyby $\sim (P_n \xrightarrow{d} P)$, to istniałby a – punkt ciągłości F_P oraz podciąg l_n takie, że dla pewnego $\varepsilon > 0$: $|F_{l_n}(a) - F_P(a)| \geq \varepsilon$ dla wszystkich l_n . Stosując poprzednie rozumowanie dla ciągu P_{l_n} , otrzymamy jego podciąg, $P_{m_{l_n}} \xrightarrow{d} P$, co stanowi sprzeczność. \square

Wykazane poprzednio twierdzenia pozwalają na sformułowanie czterech warunków równoważnych charakteryzujących zbieżność.

Twierdzenie – 11.11 (o zbieżności rozkładów). Niech P_n będzie ciągiem rozkładów, a P rozkładem. Następujące warunki są równoważne:

1. $P_n \xrightarrow{d} P$ (to znaczy: $\lim_{n \rightarrow \infty} F_{P_n}(a) = F_P(a)$ w każdym punkcie a ciągłości dystrybuanty F_P).
2. $\forall a < b, P(a) = P(b) = 0 : P_n(a, b] \rightarrow P(a, b]$, gdy $n \rightarrow \infty$.
3. $\forall f : \mathbb{R} \rightarrow \mathbb{R}$ ciągłej i ograniczonej $\int_{\mathbb{R}} f dP_n \rightarrow \int_{\mathbb{R}} f dP$, gdy $n \rightarrow \infty$.
4. $\forall f : \mathbb{R} \rightarrow \mathbb{R}$ ciągłej o suporcie zwartym $\int_{\mathbb{R}} f dP_n \rightarrow \int_{\mathbb{R}} f dP$, gdy $n \rightarrow \infty$.

Dowód. (1) \implies (2) – twierdzenie 11.6,

(2) \implies (3) – twierdzenie 11.8,

(3) \implies (4) – oczywiste,

(4) \implies (1) – twierdzenie 11.10. \square

11.2 Funkcje charakterystyczne

Zamiast badać zbieżność miar, nieraz łatwiej jest badać zbieżność ich funkcji charakterystycznych zdefiniowanych poniżej. Funkcje charakterystyczne wykorzystywane są także do innych celów.

Niech P będzie rozkładem na \mathbb{R} .

Definicja – 11.12. Funkcję $h : \mathbb{R} \rightarrow \mathbb{C}$ określoną wzorem:

$$h(u) = \int_{\mathbb{R}} e^{iux} dP(x) = \int_{\mathbb{R}} \cos ux dP(x) + i \int_{\mathbb{R}} \sin ux dP(x)$$

nazywamy funkcją charakterystyczną rozkładu P . Dla podkreślenia związku często piszemy h_P zamiast h .

Jeżeli rozkład ma dystrybuantę F , to piszemy też h_F zamiast h_P . Podobnie gdy P jest rozkładem pewnej zmiennej losowej X , to mówimy o funkcji charakterystycznej zmiennej losowej, mając na myśli funkcję charakterystyczną jej rozkładu. Piszemy wtedy h_X .

W tym ostatnim przypadku mamy:

$$h_X(u) = E(e^{iuX}) = E(\cos uX) + iE(\sin uX).$$

Uwaga – 11.13. Niech h będzie funkcją charakterystyczną pewnego rozkładu. Wtedy:

1. $h(0) = 1$.
2. $|h(u)| \leq 1$, dla każdego $u \in \mathbf{R}$,
3. h jest funkcją ciągłą.

Dla dowolnej zmiennej losowej X oraz liczb rzeczywistych a, b mamy

$$4. h_{aX+b}(u) = e^{iub} h_X(au).$$

Dowód. Wynika natychmiast z odpowiednich własności całek (ćwiczenie). □

Twierdzenie – 11.14. Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi. Wtedy

$$h_{X_1+\dots+X_n}(u) = h_{X_1}(u) \cdot \dots \cdot h_{X_n}(u).$$

Dowód. X_1, X_2, \dots, X_n są niezależne, $e^{iuX_1}, \dots, e^{iuX_n}$ (traktowane jako wektory losowe o wartościach w \mathbb{R}^2), a więc: $h_{X_1+\dots+X_n}(u) = E(e^{iu(X_1+\dots+X_n)}) = E(e^{iuX_1} \cdot \dots \cdot e^{iuX_n}) = E(e^{iuX_1}) \cdot \dots \cdot E(e^{iuX_n}) = h_{X_1}(u) \cdot \dots \cdot h_{X_n}(u)$. □

Twierdzenie – 11.15 (o jednoznaczności). Niech P_1 oraz P_2 będą rozkładami takimi, że $h_{P_1} = h_{P_2}$.

Wtedy $P_1 = P_2$.

Dowód. Jeżeli mamy równość funkcji charakterystycznych $h_{P_1} = h_{P_2}$, to są sobie równe ich części rzeczywiste i urojone, a więc dla każdego $u \in \mathbb{R}$ mamy:

$$\int_{\mathbb{R}} \cos ux \, dP_1(x) = \int_{\mathbb{R}} \cos ux \, dP_2(x),$$

$$\int_{\mathbb{R}} \sin ux \, dP_1(x) = \int_{\mathbb{R}} \sin ux \, dP_2(x).$$

Z liniowości całek możemy rozszerzyć tę równość na dowolne wielomiany trygonometryczne, a następnie na funkcje ciągłe i okresowe $\mathbb{R} \rightarrow \mathbb{R}$, gdyż z analizy wiadomo, że każdą funkcję ciągłą i okresową można aproksymować jednostajnie na \mathbb{R} wielomianami trygonometrycznymi.

Niech teraz g będzie dowolną funkcją ciągłą o suporcie zwartym. Pokażemy, że równość powyższa zachodzi także dla g , czyli że

$$\int_{\mathbb{R}} g(x) \, dP_1(x) = \int_{\mathbb{R}} g(x) \, dP_2(x).$$

Ustalmy dowolne $\varepsilon > 0$ i niech $M = \sup\{|g(x)| : x \in \mathbb{R}\}$. Ponieważ zbiory $\mathbb{R} \setminus (-T, T)$ tworzą ciąg zstępujący (gdy T rośnie) o części wspólnej $= \emptyset$, więc można znaleźć takie T , że $P_i(\mathbb{R} \setminus I) \leq \frac{\varepsilon}{2M}$ dla $i = 1, 2$ oraz $\text{supp } g \subset I$, gdzie I oznacza przedział $(-T, T)$.

Zmodyfikujmy funkcję g poza przedziałem I tak, aby otrzymać funkcję \tilde{g} okresową i ciągłą, określoną na \mathbb{R} i taką, że $g|_I = \tilde{g}|_I$.

Oczywiście $|\tilde{g}(x)| \leq M$, dla każdego $x \in \mathbb{R}$, przy czym:

$$\int_{\mathbb{R}} \tilde{g}(x) \, dP_1(x) = \int_{\mathbb{R}} \tilde{g}(x) \, dP_2(x).$$

Z równości tej kolejno otrzymujemy:

$$\int_I \tilde{g}(x) \, dP_1(x) + \int_{\mathbb{R} \setminus I} \tilde{g}(x) \, dP_1(x) = \int_I \tilde{g}(x) \, dP_2(x) + \int_{\mathbb{R} \setminus I} \tilde{g}(x) \, dP_2(x).$$

$$\begin{aligned} \left| \int_I \tilde{g}(x) \, dP_1(x) - \int_I \tilde{g}(x) \, dP_2(x) \right| &= \left| \int_{\mathbb{R} \setminus I} \tilde{g}(x) \, dP_1(x) - \int_{\mathbb{R} \setminus I} \tilde{g}(x) \, dP_2(x) \right| \\ &\leq \left| \int_{\mathbb{R} \setminus I} \tilde{g}(x) \, dP_1(x) \right| + \left| \int_{\mathbb{R} \setminus I} \tilde{g}(x) \, dP_2(x) \right| \leq 2M \frac{\varepsilon}{2M} = \varepsilon. \end{aligned}$$

Mamy więc:

$$\left| \int_{\mathbb{R}} g(x) dP_1(x) - \int_{\mathbb{R}} g(x) dP_2(x) \right| = \left| \int_I \tilde{g}(x) dP_1(x) - \int_I \tilde{g}(x) dP_2(x) \right| \leq \varepsilon.$$

Ponieważ $\varepsilon > 0$ jest dowolne, więc dla funkcji ciągłych o supportie zwartym całki też są sobie równe.

Z lematu 11.7, $P_1 = P_2$. □

Istnieje iniekcja określona na zbiorze rozkładów w zbiór funkcji ciągłych i ograniczonych $\mathbb{R} \rightarrow K(0, 1) \subset \mathbb{C}$, spełniających warunek $h(0) = 1$.

Twierdzenie – 11.16. Niech $X : \Omega \rightarrow \mathbb{R}$ będzie zmienną losową, $h = h_X$ jej funkcją charakterystyczną. Jeżeli istnieje moment rzędu k i jest skończony, to funkcja h jest k -krotnie różniczkowalna. Wtedy:

$$h^{(k)}(0) = i^k E(X^k).$$

Dowód. Przeprowadzimy dla $k = 1$; dla pozostałych k dowód jest już prosty.

Ponieważ

$$\int_{\mathbb{R}} \left| \frac{d}{du} e^{iux} \right| dP_X(x) = \int_{\mathbb{R}} |ix e^{iux}| dP_X(x) = \int_{\mathbb{R}} |x| dP_X(x) = E(|X|) < \infty,$$

więc – korzystając ze znanego z kursu analizy twierdzenia – istnieje pochodna funkcji określonej przez całkę: $\frac{d}{du} \int_{\mathbb{R}} e^{iux} dP(x) = h'(u)$, i aby ją obliczyć, można różniczkować pod znakiem całki. Tak więc

$$h'(u) = \int_{\mathbb{R}} ix e^{iux} dP_X(x),$$

a kładąc $u = 0$, mamy $h'(0) = i \int_{\mathbb{R}} x dP_X(x) = i E(X)$. □

Większość używanych rozkładów ma już dawno wyznaczone swoje funkcje charakterystyczne. Warto jednak zobaczyć choćby jak to się robi.

1. Dla rozkładu dyskretnego

$$h_P(u) = \sum_{k=0}^N e^{iux_k} p_k. \quad (11.1)$$

2. Dla rozkładów ciągłych

$$h_P(u) = \int_{\mathbb{R}} e^{iux} f(x) dx. \quad (11.2)$$

Przykład, rozkład δ_c .

$$h_{\delta_c}(u) = e^{iuc} \cdot 1 = e^{iuc}.$$

Przykład, rozkład $N(0, 1)$.

Twierdzenie – 11.17. Funkcja charakterystyczna h rozkładu normalnego $N(0, 1)$ wyraża się wzorem

$$h(u) = e^{-\frac{1}{2}u^2}.$$

Dowód. Dopelniając do kwadratu, otrzymujemy

$$h(u) = \int_{\mathbb{R}} e^{iux} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(x-iu)^2} dx.$$

Wystarczy udowodnić, że $I := \int_{\mathbb{R}} e^{-\frac{1}{2}(x-iu)^2} dx = \sqrt{2\pi}$.

Wiedząc, że całka z funkcji analitycznej po drodze zamkniętej jest równa zeru, rozważmy funkcję $f(z) = e^{-\frac{1}{2}z^2}$ oraz prostokąt o wierzchołkach w punktach $-N-iu$, $N-iu$, N , $-N$, gdzie N oraz u są ustalonymi liczbami rzeczywistymi.

$$\int_{-N-iu}^{N-iu} f(z) dz + \int_{N-iu}^N f(z) dz + R_N = 0,$$

gdzie R_N jest sumą dwóch całek po odcinkach pionowych. Niech $z = x - iu$

$$\int_{-N}^N e^{-\frac{1}{2}(x-iu)^2} dx = \int_{-N}^N f(z) dz - R_N = \int_{-N}^N e^{-\frac{1}{2}x^2} dx - R_N.$$

Gdy $N \rightarrow \infty$, to $R_N \rightarrow 0$, gdyż miara pionowych odcinków jest stała, a funkcja f na tych odcinkach maleje do 0. W takim razie

$$I = \lim_{N \rightarrow \infty} \int_{-N}^N e^{-\frac{1}{2}(x-iu)^2} dx = \lim_{N \rightarrow \infty} \int_{-N}^N e^{-\frac{1}{2}x^2} dx = \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}. \quad \square$$

Przykład – 11.18 (Rozkład $N(m, \sigma)$). Jeżeli zmienna losowa X rozkład $N(0, 1)$, to zmienna losowa $Y = \sigma X + m$ ma rozkład $N(m, \sigma)$. W takim razie funkcja charakterystyczna h tego rozkładu jest równa h_Y .

$$h(u) = h_Y(u) = e^{ium} h_X(\sigma u) = e^{ium} e^{-\frac{1}{2}(\sigma u)^2} = e^{ium - \frac{1}{2}\sigma^2 u^2}.$$

Przykład – 11.19 (Rozkład sumy zmiennych losowych i.i.d. o rozkładach normalnych). Niech $P_X = N(m_1, \sigma_1)$, $P_Y = N(m_2, \sigma_2)$. Wtedy: $h_{X+Y}(u) = h_X(u)h_Y(u) = e^{ium_1 - \frac{1}{2}\sigma_1^2 u^2} e^{ium_2 - \frac{1}{2}\sigma_2^2 u^2} = e^{iu(m_1+m_2) - \frac{1}{2}(\sigma_1^2 + \sigma_2^2)u^2}$.

A to jest funkcja charakterystyczna rozkładu $N(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$. Więc $P_{X+Y} = N(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

Znajomość funkcji charakterystycznej pozwala czasami na oszacowanie prawdopodobieństwa „ogona” rozkładu:

Twierdzenie – 11.20. Niech P będzie rozkładem a h jego funkcją charakterystyczną. Niech $u > 0$. Wtedy:

$$P\left(x \in \mathbb{R} : |x| \geq \frac{2}{u}\right) \leq \frac{1}{u} \int_{-u}^u 1 - h(s) ds \in \mathbb{R}.$$

Dowód.

$$\begin{aligned} \int_{-u}^u 1 - h(s) ds &= \int_{-u}^u \int_{\mathbb{R}} dP(x) - \int_{\mathbb{R}} e^{isx} dP(x) ds = \int_{\mathbb{R}} \int_{-u}^u 1 - \cos(sx) ds dP(x) \\ &= \int_{\mathbb{R}} 2u - 2 \frac{\sin(ux)}{x} dP(x) = 2u \left(\int_{\{|x|: |ux| < 2\}} 1 - \frac{\sin(ux)}{ux} dP(x) + \int_{\{|x|: |ux| \geq 2\}} 1 - \frac{\sin(ux)}{ux} dP(x) \right) \\ &\geq 2u \left(0 + \frac{1}{2} P(x : |ux| \geq 2) \right) = u P(x : |ux| \geq 2). \quad \square \end{aligned}$$

Jednym z najważniejszych zalet funkcji charakterystycznych jest zgodność zbieżności ciągu rozkładów ze zbieżnością odpowiedniego ciągu funkcji charakterystycznych. Mówi o tym następujące twierdzenie.

Twierdzenie – 11.21 (o ciągłości). Niech $\{P_n\}$ będzie ciągiem rozkładów, a $h_n = h_{P_n}$ ciągiem funkcji charakterystycznych.

1. Jeżeli P jest rozkładem i $P_n \xrightarrow{d} P$, to $\forall u \in \mathbb{R} : \lim_{n \rightarrow \infty} h_n(u) = h_P(u)$.
2. Jeżeli $h : \mathbb{R} \rightarrow \mathbb{C}$ jest funkcją ciągłą w 0 oraz $\forall u \in \mathbb{R} : \lim_{n \rightarrow \infty} h_n(u) = h(u)$, to istnieje rozkład P taki, że $P_n \xrightarrow{d} P$. Wtedy też $h = h_P$.

Dowód. Ad 1. Dla każdego $u \in \mathbb{R}$, $\sin(ux)$ oraz $\cos(ux)$ są funkcjami ciągłymi i ograniczonymi, więc stosuje się twierdzenie 11.11 o zbieżności rozkładów. \square

Ad 2. Udowodnimy, że $\{P_n\}$ spełnia warunek Prochorowa. Niech $\varepsilon > 0$. Niech $0 < \varepsilon' < \varepsilon$. Istnieje $u > 0$, takie, że $|1 - h(s)| \leq \frac{\varepsilon'}{2}$, dla $|s| \leq u$.

Z własności całek: $\frac{1}{2u} \int_{-u}^u |1 - h(s)| ds \leq \frac{\varepsilon'}{2}$.

Z twierdzenia Lebesgue'a $\frac{1}{2u} \int_{-u}^u |1 - h_n(s)| ds \rightarrow \frac{1}{2u} \int_{-u}^u |1 - h(s)| ds$. Istnieje n_0 takie, że dla $n \geq n_0$ zachodzi $\frac{1}{2u} \int_{-u}^u (1 - h_n(s)) ds \leq \frac{\varepsilon}{2}$. Z poprzedniego twierdzenia: $P_n \left(x \in \mathbb{R} : |x| \geq \frac{2}{u} \right) \leq \frac{1}{u} \int_{-u}^u 1 - h_n(s) ds \leq \varepsilon$.

$\{P_n\}$ spełnia więc warunek Prochorowa. Istnieją podciąg $\{P_{k_n}\}$ oraz rozkład P , taki, że $P_{k_n} \xrightarrow{d} P$. Z punktu 1: $\forall u \in \mathbb{R} : \lim_{n \rightarrow \infty} h_{P_{k_n}}(u) = h_P(u)$. Ale również $\forall u \in \mathbb{R} : \lim_{n \rightarrow \infty} h_{P_{k_n}}(u) = h(u)$. Więc $h_P = h$. Gdyby P_n nie zmierzał do P , to istniałby podciąg P_{l_n} oraz rozkład Q różny od P taki, że $P_{l_n} \xrightarrow{d} Q$. Rozumując jak poprzednio widzimy, że $h_Q = h = h_P$. Z twierdzenia o jednoznaczności $P = Q$. \square

11.3 Dowód centralnego twierdzenia granicznego

W oparciu o teorię funkcji charakterystycznych możemy podać dowód Centralnego Twierdzenia Granicznego. Najpierw udowodnimy twierdzenie Lindeberga–Lévy’ego, twierdzenie 9.1, a jak się okaże wynikają z niego twierdzenie 9.2 oraz twierdzenie 9.3.

Przypominamy, że: X_1, X_2, X_3, \dots jest ciągiem niezależnych zmiennych losowych (i.i.d.) określonych na przestrzeni probabilistycznej (Ω, Σ, P) . Wszystkie zmienne losowe X_i mają taki sam rozkład, a ich wspólna nadzieja matematyczna m oraz wariancja σ^2 istnieją i są skończone, przy czym $\sigma > 0$.

$$S_n = X_1 + \dots + X_n, \quad Z_n := \frac{S_n - E(S_n)}{\sqrt{D^2(S_n)}} = \frac{S_n - nm}{\sigma\sqrt{n}}.$$

Wtedy $E(Z_n) = 0$ oraz $D^2(Z_n) = 1$ (ćwiczenie).

Mamy wykazać, że: dla każdego $x \in \mathbb{R}$ zachodzi równość:

$$\lim_{n \rightarrow \infty} P(Z_n \leq x) = \Phi(x),$$

gdzie Φ jest dystrybuantą standardowego rozkładu normalnego, czyli, że $P_{Z_n} \xrightarrow{d} N(0, 1)$.

Dowód twierdzenia 9.1. Skorzystamy z drugiej części twierdzenia o ciągłości.

Musimy więc wykazać, że ciąg funkcji charakterystycznych $h_n := h_{Z_n}$ jest zbieżny do funkcji charakterystycznej h_Φ , którą to funkcję wyznaczyliśmy.

$$h_\Phi(u) = e^{-\frac{1}{2}u^2}.$$

Ponieważ

$$Z_n = \frac{(X_1 - m) + \dots + (X_n - m)}{\sigma\sqrt{n}},$$

więc korzystając z niezależności zmiennych losowych $(X_1 - m), \dots, (X_n - m)$ oraz z własności funkcji charakterystycznych otrzymujemy:

$$h_n(u) = h\left(\frac{u}{\sigma\sqrt{n}}\right)^n,$$

gdzie przez h oznaczamy funkcję charakterystyczną rozkładu zmiennej losowej $X_i - m$.

Ponieważ zmienne losowe mają z założenia moment rzędu 2, więc wiemy, że funkcja h jest dwukrotnie różniczkowalna w zerze. Mamy przy tym

$$h(0) = 1, \quad h'(0) = iE(X_i - m) = 0, \quad h''(0) = i^2E((X_i - m)^2) = -\sigma^2.$$

Mamy więc:

$$h(u) = 1 - \frac{1}{2}\sigma^2 u^2 + o(u^2), \quad \text{gd } u \rightarrow 0.$$

Przypomnienie. Dla funkcji f, g określonych na przestrzeni topologicznej A oraz $a \in A$ definiujemy:

$$f(x) = o(g(x)), \text{ gdy } x \rightarrow a \iff \forall \varepsilon > 0 \exists U \text{ otoczenie } a \forall x \in U : |f(x)| \leq \varepsilon|g(x)|.$$

Intuicja. Gdy $g(x) \rightarrow 0$ dla $x \rightarrow a$ oraz $f(x) = o(g(x))$, dla $x \rightarrow a$, to $f(x)$ **zmierza istotnie szybciej** do 0 niż $g(x)$.

Ustalmy teraz u . Mamy wtedy

$$h\left(\frac{u}{\sigma\sqrt{n}}\right) = 1 - \frac{1}{2}\frac{u^2}{n} + o\left(\frac{1}{n}\right), \quad \text{gd } n \rightarrow \infty.$$

Oznacza to, że dla dużych n wartości $h\left(\frac{u}{\sigma\sqrt{n}}\right)$ leżą w kole o środku w punkcie $z = 1$ i promieniu $r = 1$.

Jak wiadomo z kursu funkcji analitycznych, w kole tym istnieje dokładnie jedna funkcja analityczna (nazywa się ją gałęzią logarytmu naturalnego i oznacza przez \log), taka, że $e^{\log z} = z$ oraz $\log 1 = 0$. Możemy też skorzystać z faktów, że $\log z^n = n \log z$ oraz

$$\log(1 + z) = z + o(z), \quad \text{dla } z \rightarrow 0.$$

Kolejno mamy:

$$\log h_n(u) = n \log h \left(\frac{u}{\sigma\sqrt{n}} \right) = n \log \left(1 - \frac{1}{2} \frac{u^2}{n} + o \left(\frac{1}{n} \right) \right), \quad \text{gdzie } n \rightarrow \infty,$$

$$\log h_n(u) = -\frac{1}{2} u^2 + \phi(n),$$

gdzie $\phi(n) = o \left(-\frac{1}{2} \frac{u^2}{n} + n o \left(\frac{1}{n} \right) \right)$ dla $n \rightarrow \infty$.

Widać, że $\phi(n) \rightarrow 0$ dla $n \rightarrow \infty$.

Tak więc ostatecznie

$$h_n(u) = e^{-\frac{u^2}{2}} e^{\phi(n)} \rightarrow e^{-\frac{u^2}{2}} = h_\Phi(u), \quad \text{dla } n \rightarrow \infty.$$

Z twierdzenia o ciągłości $P_{Z_n} \xrightarrow{d} N(0, 1)$. □

Komentarz. Wiadomo, że $\lim_{n \rightarrow \infty} \left(1 - \frac{1}{2} \frac{u^2}{n} \right)^n = e^{-\frac{u^2}{2}}$. Stosując elementy teorii funkcji analitycznych, pokazaliśmy, że również $\lim_{n \rightarrow \infty} \left(1 - \frac{1}{2} \frac{u^2}{n} + o \left(\frac{1}{n} \right) \right)^n = e^{-\frac{u^2}{2}}$, co jest zgodne z intuicją.

Twierdzenie – 11.22. *W tezie twierdzenia Lindeberga–Lévy’ego zachodzi zbieżność jednostajna ze względu na $x \in \mathbb{R}$.*

Jest to konsekwencja następującego lematu.

Lemat – 11.23. *Założmy, że*

$$F_n \xrightarrow{d} F,$$

gdzie F_n oraz F są dystrybuantami. Załóżmy ponadto, że F jest funkcją ciągłą. Wtedy

$$F_n \rightarrow F, \quad \text{jednostajnie na } \mathbb{R}.$$

Czyli: $\forall \varepsilon > 0 \exists n_0 \forall x \in \mathbb{R} \forall n \geq n_0 : |F_n(x) - F(x)| < \varepsilon$.

Dowód lematu. Ustalmy $\varepsilon > 0$ i bez straty ogólności założmy, że $\varepsilon < 1$. Ponieważ dystrybuanta F jest funkcją ciągłą, istnieją punkty x_1, \dots, x_r takie, że $F(x_k) = \frac{k\varepsilon}{2}$, $k = 1, \dots, r$, gdzie r jest największą liczbą taką, że $\frac{r\varepsilon}{2} < 1$. Połóżmy dodatkowo $x_0 = -\infty$, $x_{r+1} = \infty$. Oczywiście jest $F(x_{k+1}) - F(x_k) \leq \frac{\varepsilon}{2}$. Z założenia wiemy, że we wszystkich punktach x_k mamy $\lim_{n \rightarrow \infty} F_n(x_k) = F(x_k)$. Ponieważ jest ich skończenie wiele, istnieje takie N , że dla $n > N$ i dla wszystkich $k = 1, \dots, r$

$$|F_n(x_k) - F(x_k)| \leq \frac{\varepsilon}{2}.$$

Niech $x \in \mathbb{R}$ będzie dowolnie ustalone. Istnieje taki przedział $[x_k, x_{k+1}]$, który zawiera x . Mamy teraz

$$F_n(x) - F(x) \leq F_n(x_{k+1}) - F(x_k) \leq F(x_{k+1}) + \frac{\varepsilon}{2} - F(x_k) \leq \varepsilon$$

i podobnie

$$F(x) - F_n(x) \leq F(x_{k+1}) - F_n(x_k) \leq F(x_{k+1}) - \left(F(x_k) - \frac{\varepsilon}{2} \right) \leq \varepsilon,$$

co oznacza, że $|F_n(x) - F(x)| \leq \varepsilon$ dla dowolnego x oraz $n > N$. □

Dowód twierdzenia 9.2, CTG dla sum. Ponieważ z określenia zmiennej losowej Z_n mamy $S_n = \sigma\sqrt{n}Z_n + nm$, więc mamy

$$F_{S_n}(x) = F_{Z_n} \left(\frac{x - nm}{\sigma\sqrt{n}} \right).$$

Natomiast także w naszym przypadku:

$$\Phi_{nm, \sigma\sqrt{n}}(x) = \Phi \left(\frac{x - nm}{\sigma\sqrt{n}} \right).$$

Twierdzenie 11.22 mówi, że: $\forall \varepsilon > 0 \exists n_0 \forall n \geq n_0 \forall x \in \mathbb{R} :$

$$\left| F_{S_n}(x) - \Phi_{nm, \sigma\sqrt{n}}(x) \right| = \left| F_{Z_n} \left(\frac{x - nm}{\sigma\sqrt{n}} \right) - \Phi \left(\frac{x - nm}{\sigma\sqrt{n}} \right) \right| < \varepsilon. \quad \square$$

Dowód twierdzenia 9.3, CTG dla średnich. (ćwiczenie). □

11.4 Funkcje tworzące

Do badania rozkładów zmiennych losowych przyjmujących wartości całkowite nieujemne zamiast funkcji charakterystycznych można używać tak zwanych funkcji tworzących.

Niech a_i , $i = 0, 1, 2, 3, \dots$ będzie ciągiem liczb rzeczywistych nieujemnych. Funkcją tworzącą tego ciągu jest funkcja zmiennej zespolonej z :

$$\alpha(z) := \sum_{i=0}^{\infty} a_i z^i. \quad (11.3)$$

Zauważmy, że jeżeli ciąg $\{a_i\}$ jest ograniczony, to szereg powyższy jest zbieżny dla $|z| < 1$, a więc funkcja α jest analityczna w otwartym kole $K(0, 1)$. Jeżeli dodatkowo $\sum_{i=0}^{\infty} a_i = 1$, to szereg (11.3) jest zbieżny także dla $|z| = 1$.

Niech X będzie zmienną losową określoną na odpowiedniej przestrzeni probabilistycznej. Załóżmy, że X przyjmuje jedynie wartości całkowite nieujemne, czyli $P(\bigcup_{i=0}^{\infty} \{X = i\}) = 1$. Niech $p_i = P(X = i)$, $i = 0, 1, 2, \dots$. Określamy funkcję tworzącą zmiennej losowej X jako funkcję tworzącą ciągu $\{p_i\}$. Jest to więc funkcja:

$$\pi_X(z) := \pi(z) = \sum_{i=0}^{\infty} p_i z^i. \quad (11.4)$$

Zauważmy, że funkcja tworząca π_X oraz funkcja charakterystyczna h_X są ze sobą ściśle związane. Mianowicie, stosując wzór (11.1) na funkcję charakterystyczną rozkładu dyskretnego, mamy dla rzeczywistych u :

$$h_X(u) = \sum_{k=0}^{\infty} e^{iuk} p_k = \sum_{k=0}^{\infty} (e^{iu})^k p_k = \pi_X(e^{iu}).$$

Ponieważ funkcja π , jako funkcja analityczna w kole $K(0, 1)$, jest wyznaczona jednoznacznie przez swoje wartości na okręgu $|z| = 1$, powyższy związek pozwala wykazać wiele własności funkcji tworzących na podstawie odpowiednich własności funkcji charakterystycznych. Własności te mogą być otrzymane także bezpośrednio – bez odwoływania się do funkcji charakterystycznych. Na przykład, często wykorzystuje się następującą własność będącą konsekwencją (ćwiczenie) twierdzenia 11.14:

Twierdzenie – 11.24. *Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi przyjmującymi nieujemne wartości całkowite. Wtedy*

$$\pi_{X_1+\dots+X_n}(z) = \pi_{X_1}(z) \cdot \dots \cdot \pi_{X_n}(z).$$

Zachodzi także twierdzenie o jednoznaczności rozkładu, które jest szczególnym przypadkiem twierdzenia 11.15 (ćwiczenie).

Podobnie jak funkcje charakterystyczne, funkcje tworzące mogą służyć do wyznaczania momentów.

Twierdzenie – 11.25. *Niech X będzie nieujemną zmienną losową przyjmującą wartości całkowite.*

1. *Jeżeli $E(X) < \infty$, to $E(X) = \pi'_X(1)$.*
2. *Jeżeli $E(X^2) < \infty$, to $E(X^2) = \pi'_X(1) + \pi''_X(1)$.*

Dowód. Można zróżniczkować szereg dla $|z| < 1$. Ponieważ istnieje $E(X)$, to szereg określający π'_x jest zbieżny dla $z = 1$ (ćwiczenie). \square

Podamy teraz trzy proste zastosowania funkcji tworzących:

Spacery losowe po prostej

Wyobraźmy sobie ruch cząstki po osi liczbowej odbywający się według następujących zasad.

- (1) W chwili $n = 0$ cząstka znajduje się w ustalonym punkcie, powiedzmy w punkcie $A = 0$.
- (2) Jeżeli w chwili n cząstka znajduje się w punkcie x , to w chwili $n + 1$ znajduje się w punkcie $x + 1$ z prawdopodobieństwem p i w punkcie $x - 1$ z prawdopodobieństwem q . Zakładamy, że $p + q = 1$. Interesuje nas czy cząstka musi wrócić do wyjściowego stanu A .

Najpierw sprecyzujemy nasze zadanie. Niech X_i , $i = 1, 2, 3, \dots$ będą niezależnymi zmiennymi losowymi o rozkładzie:

$$P(X_i = -1) = q, \quad P(X_i = 1) = p.$$

Niech $S_n = X_1 + \dots + X_n$ i niech

$$T = \min\{n > 0 : S_n = 0\}. \quad (11.5)$$

Pytanie postawione poprzednio można teraz sformułować następująco: Czy $P(T < \infty) = 1$? Prawdopodobieństwo $P(T < \infty)$ nazywamy prawdopodobieństwem powrotu.

Rozważmy dwa ciągi liczb $\{a_n\}$ oraz $\{f_n\}$ określające odpowiednio prawdopodobieństwa pobytu cząstki w punkcie A w chwili n oraz prawdopodobieństwo pierwszego powrotu do A w chwili n . Definiujemy więc:

$$a_n = P(S_n = 0), \quad f_n = P(T = n), \quad \text{dla } n = 1, 2, 3, \dots$$

oraz dookreślamy $a_0 = 0$, $f_0 = 0$.

Ponieważ zdarzenia $\{T = n\}$ są parami rozłączne, więc prawdopodobieństwo powrotu:

$$P(T < \infty) = \sum_{n=0}^{\infty} f_n = \varphi(1),$$

gdzie φ jest funkcją tworzącą ciąg $\{f_n\}$. Oczywiście $\varphi(1) \leq 1$.

Znajdziemy związek między funkcją φ oraz α , funkcją tworzącą ciąg a_n . Ze wzoru na prawdopodobieństwo całkowite oraz z niezależności zmiennych X_k mamy dla $n = 1, 2, 3, \dots$ (ćwiczenie):

$$\begin{aligned} a_1 &= f_1, \\ a_2 &= f_1 a_1 + f_2, \\ a_3 &= f_1 a_2 + f_2 a_1 + f_3, \\ a_4 &= f_1 a_3 + f_2 a_2 + f_3 a_1 + f_4, \\ &\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \end{aligned}$$

Pomnóżmy te równości odpowiednio przez z , z^2 , z^3 , z^4 , \dots , $|z| < 1$, i zsumujmy stronami. Otrzymujemy:

$$\alpha(z) = f_1 z (1 + a_1 z + a_2 z^2 + \dots) + f_2 z^2 (1 + a_1 z + a_2 z^2 + \dots) + f_3 z^3 (1 + a_1 z + a_2 z^2 + \dots) + \dots$$

Mamy więc dla $|z| < 1$

$$\alpha(z) = \varphi(z)(1 + \alpha(z)).$$

Niech z zmierza do 1 po osi rzeczywistej w sposób rosnący. Wtedy wartości $\alpha(z)$, $\varphi(z)$ rosną i zbiegają odpowiednio do granic $\alpha(1)$ i $\varphi(1)$, czyli:

$$\alpha(1) = \varphi(1)(1 + \alpha(1)),$$

przy czym wiemy, że $\varphi(1) \leq 1$. Widzimy teraz, że $\varphi(1) = 1$, wtedy i tylko wtedy, gdy $\alpha(1) = \infty$. Co więcej, gdy $\alpha(1) < \infty$, to

$$\varphi(1) = \frac{\alpha(1)}{1 + \alpha(1)}. \quad (11.6)$$

Wykazaliśmy więc następujące:

Twierdzenie – 11.26.

1. Jeżeli $\sum_{n=1}^{\infty} a_n < \infty$, to prawdopodobieństwo powrotu jest mniejsze od jeden i wynosi:

$$P(T < \infty) = \frac{\sum_{n=1}^{\infty} a_n}{1 + \sum_{n=1}^{\infty} a_n}. \quad (11.7)$$

2. Jeżeli $\sum_{n=1}^{\infty} a_n = \infty$, to prawdopodobieństwo powrotu jest równe jeden.

Wyznamy teraz liczby a_n . Są to prawdopodobieństwa tego, że cząstka startująca z A znajdzie się po n krokach znowu w punkcie A . Cząstka musi więc wykonać tyle samo kroków w prawo co w lewo. Mamy więc:

$$a_n = \begin{cases} 0, & \text{gdy } n = 2k - 1, \\ \binom{2k}{k} p^k q^k, & \text{gdy } n = 2k, \end{cases} \quad \text{dla } k = 1, 2, 3, \dots$$

Do zbadania zbieżności szeregu $\sum_{n=1}^{\infty} a_n$ wykorzystamy słynny wzór:

Wzór Stirlinga

$$\lim_{n \rightarrow \infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n}} = 1, \quad (11.8)$$

który często używany jest w formie:

$$n! \cong n^n e^{-n} \sqrt{2\pi n}, \quad \text{dla dużych } n.$$

Wzór ten oznacza (ćwiczenie), że:

$$a_{2k} \cong \frac{(4pq)^k}{\sqrt{\pi k}}, \quad k \rightarrow \infty. \quad (11.9)$$

Ponieważ $pq \leq \frac{1}{4}$, a równość zachodzi dokładnie wtedy, gdy $p = q = \frac{1}{2}$, widzimy, że:

$$\sum_{n=1}^{\infty} a_n = \infty \iff p = q = \frac{1}{2}.$$

Tak więc przy symetrycznym spacerze losowym po prostej cząstka prędzej czy później powróci do stanu wyjściowego. Natomiast, gdy spacer nie jest symetryczny cząstka z dodatnim prawdopodobieństwem nigdy nie powróci do stanu wyjściowego.

Można też bezpośrednio wyznaczyć sumę szeregu $\sum_{n=1}^{\infty} a_n$ oraz $P(T < \infty)$, patrz ćwiczenie 11.4. Mianowicie:

$$P(T < \infty) = 1 - |2p - 1| = 1 - |p - q|.$$

Jako że dyskutowany przez nas spacer losowy jest szczególnym przypadkiem łańcucha Markowa, wrócimy do tego problemu w rozdziale 16. Wykorzystamy wtedy przybliżenie zadane wzorem (11.9).

Suma losowej liczby składników

Niech X_1, X_2, X_3, \dots oraz N będą zmiennymi losowymi przyjmującymi wartości całkowite nieujemne. Interesuje nas suma pierwszych N zmiennych X_i , czyli zmienna losowa:

$$S := X_1 + \dots + X_N.$$

Wykażemy następujące:

Twierdzenie – 11.27. *Jeżeli zmienne losowe X_1, X_2, X_3, \dots są niezależne i mają taki sam rozkład o funkcji tworzącej π , a zmienna losowa N jest niezależna od X_1, X_2, X_3, \dots i ma funkcję tworzącą ν , to zmienna losowa S ma funkcję tworzącą $\sigma = \nu \circ \pi$.*

Dowód. Niech $\pi(z) = \sum_{i=0}^{\infty} p_i z^i$, $\nu(z) = \sum_{i=0}^{\infty} n_i z^i$, $\sigma(z) = \sum_{i=0}^{\infty} s_i z^i$. Oznaczmy przez $\sigma^{(n)}$ funkcję tworzącą sumy $S_n = X_1 + X_2 + \dots + X_n$, $n = 0, 1, 2, 3, \dots$. Niech $\sigma^{(n)}(z) = \sum_{i=0}^{\infty} s_i^{(n)} z^i$. Po pierwsze, z niezależności N od X_i mamy dla $j = 0, 1, 2, 3, \dots$:

$$s_k = P(S = k) = \sum_{i=0}^{\infty} P(N = i) P(S_i = k) = \sum_{i=0}^{\infty} \nu_i s_k^{(i)}.$$

Po drugie, z twierdzenia 11.24 wiemy, że $\sigma^{(i)}(z) = \pi(z)^i$ i stąd:

$$\begin{aligned} (\nu \circ \pi)(z) &= \nu(\pi(z)) = \sum_{i=0}^{\infty} n_i \pi(z)^i = \sum_{i=0}^{\infty} n_i \sigma^{(i)}(z) \\ &= \sum_{i=0}^{\infty} n_i \sum_{k=0}^{\infty} s_k^{(i)} z^k = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{\infty} \nu_i s_k^{(i)} \right) z^k = \sum_{k=0}^{\infty} s_k z^k, \end{aligned}$$

co oznacza tezę. □

Przykład – 11.28. Owad (powiedzmy mucha) składa dużo jajeczek z których mogą wykluwać się nowe owady. Zakładając, że liczba jajeczek ma rozkład Poissona o parametrze λ , oraz, że owady wykluwają się z jajeczek niezależnie od siebie z tym samym prawdopodobieństwem p , wyznaczyć rozkład oraz oczekiwaną wartość liczby potomków jednego owada.

Niech N oznacza liczbę jajeczek złożonych przez muchę w ciągu całego życia i niech $X_i = 1$ lub $X_i = 0$ zależnie od tego czy z i -tego jajeczka rozwinie się dorosły owad czy nie. Wtedy $S = X_1 + \dots + X_N$ oznacza liczbę dorosłych potomków muchy. Załóżmy, że N, X_1, X_2, X_3, \dots są niezależne. Wiemy, że X_i mają taki sam rozkład $B(1, p)$ oraz, że N ma rozkład Poissona o parametrze λ . Jak łatwo sprawdzić odpowiednie funkcje tworzące są równe: $\pi(z) = 1 - p + pz$ oraz $\nu(z) = e^{-\lambda + \lambda z}$. W takim razie $\sigma(z) = (\nu \circ \pi)(z) = e^{-\lambda p + \lambda p z}$. Oznacza to, że liczba dorosłych potomków S ma rozkład Poissona o parametrze λp .

Przykład ten przeanalizujemy jeszcze raz w oparciu o teorię warunkowych wartości oczekiwanych, przykład 14.5.

Proces gałązkowy

Przykład – 11.29. Osoba, powiedzmy O_0 , która jest nosicielem wirusa może przekazywać go k innym osobom z prawdopodobieństwami p_k , $k = 0, 1, 2, \dots$, a te osoby mogą przekazywać go dalej według tego samego schematu. Niech X_n oznacza liczbę osób, które otrzymały wirusa od O_0 dokładnie po n krokach. Niech π oznacza funkcję tworzącą ciągu (p_k) .

Wyznamy kolejno funkcję tworzącą π_{X_n} (zakładając niezależność), $E(X_n)$ oraz $P(X_n = 0)$.

Niech N będzie zmienną losową o rozkładzie $P(N = k) = p_k$.

- o O_0 zarazi X_1 osób, przy czym $P_{X_1} = P_N$. Tworzą one pierwszą generację.
- o Każda osoba i z pierwszej generacji zarazi $X_{1,i}$ osób, przy czym $P_{X_{1,i}} = P_N$.
- o X_2 – liczba osób w drugiej generacji jest więc sumą: $X_2 = X_{1,1} + \dots + X_{1,X_1}$.
- o Podobnie $X_n = X_{n-1,1} + \dots + X_{n-1,X_{n-1}}$.

W początkowym stadium epidemii można założyć, że występująca w sumach zmienne losowe są niezależne. Przy takim założeniu można skorzystać z twierdzenia 11.27, z którego wynika, że π_{X_n} jest n -tym złożeniem funkcji π , czyli $\pi_{X_1} = \pi$, $\pi_{X_2} = \pi^2 = \pi \circ \pi$, $\pi_{X_3} = \pi^3$ itd.

Z określenia funkcji tworzącej całkowitoliczbowej zmiennej losowej X wynika, że $P(X = 0) = \pi_X(0)$. Natomiast twierdzenie 11.25 pozwala wyliczyć wartość oczekiwaną: $E(X) = \pi'_X(1)$. Tak więc w rozważanym przypadku:

$$P(X_n = 0) = \pi^n(0), \quad E(X_n) = (\pi^n)'(1).$$

Zakładając, że $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$, czyli rozkład Poissona, można jak poprzednio wyznaczyć funkcję tworzącą π i obliczyć (np. Maple) powyższe wielkości dla $n = 10$, gdy: (a) $\lambda = 0.9$, (b) $\lambda = 1.1$. Wskazówka: n -krotne złożenie funkcji f ze sobą uzyskuje się w Maple za pomocą polecenia `f@@n`.

Otrzymujemy:

Ad (a) $E(X_{10}) \approx 0.3486784401$, $P(X_{10} = 0) \approx 0.9150828404$.

Ad (b) $E(X_{10}) \approx 2.593742460$, $P(X_{10} = 0) \approx 0.7507181832$.

11.5 Pytania

Pytanie 11.1. Czy następujące rodziny rozkładów spełniają warunek Prochorowa?

1. $\{N(m, 1) : m \in \mathbb{R}\}$,
2. $\{N(0, \sigma) : 0 < \sigma < 1\}$,
3. $\{U(a, b) : a < b < 100\}$.

Odpowiedź uzasadnić. ROZWIĄZANIE

Pytanie 11.2. Nie korzystając z MPWL, wykazać następujące twierdzenie (Chinczyna).

Jeżeli X_1, X_2, X_3, \dots są *i.i.d.* i mają skończoną nadzieję matematyczną m , to $\frac{S_n}{n} \xrightarrow{s} m$. ROZWIĄZANIE

Pytanie 11.3. Przeprowadzić dowód CTG dla średnich. ROZWIĄZANIE

Pytanie 11.4. Niech S_n oznacza sumę orłów uzyskanych w trakcie n rzutów monetą symetryczną. Niech $\varepsilon > 0$ będzie dowolną liczbą.

Obliczyć:

1. $\lim_{n \rightarrow \infty} P \left(\left| S_n - \frac{n}{2} \right| \geq \varepsilon \right),$
2. $\lim_{n \rightarrow \infty} P \left(\left| S_n - \frac{n}{2} \right| \geq \varepsilon n \right),$
3. $\lim_{n \rightarrow \infty} P \left(\left| S_n - \frac{n}{2} \right| \geq \varepsilon \sqrt{n} \right).$

ROZWIĄZANIE

Pytanie 11.5. Niech S_n oznacza sumę orłów uzyskanych w trakcie n rzutów monetą symetryczną. Sformułować jako twierdzenie następujące spostrzeżenia: Gdy wykona się dostatecznie dużo rzutów, to różnica między liczbą uzyskanych orłów i reszek będzie tak wielka jak chcemy, natomiast ich iloraz będzie coraz bliższy 1. ROZWIĄZANIE

Pytanie 11.6. Wykazać twierdzenie o jednoznaczności rozkładu w przypadku funkcji tworzących. ROZWIĄZANIE

Rozdział 12

Metody Monte Carlo

Metody Monte Carlo są rozumiane na różne sposoby. Można powiedzieć, że stanowią one narzędzia numeryczne oparte na teorii rachunku prawdopodobieństwa służące do rozwiązywania problemów, w tym problemów o podłożu deterministycznym, najczęściej wtedy, gdy zawodzą metody analityczne czy klasyczne metody numeryczne. Jednym z twórców współczesnych metod Monte Carlo był polski matematyk Stanisław Ulam pracujący w latach 40-tych XX wieku w Stanach Zjednoczonych nad konstrukcją broni jądrowej. Wraz z rozwojem technologii komputerowych metody Monte Carlo osiągnęły zawrotny rozwój i są wykorzystywane obecnie w wielu dziedzinach nauki i techniki.

Już przykład 10.21 oraz przykład 10.22 podane przez nas przy okazji omawiania znaczenia mocnego prawa wielkich liczb ilustrowały podstawowe idee metod Monte Carlo. Sygnalizujemy teraz inny klasyczny przykład. Różni się on od tych dwóch, że używa narzędzia probabilistyczne do rozwiązania problemu o charakterze deterministycznym.

Przykład – 12.1. Należy podać przybliżoną wartość liczby π .

Schemat rozwiązania: Losujemy dużo punktów, powiedzmy n , $n \geq 1000$, z kwadratu $[-1, 1]^2$ i wyznaczamy liczbę, k , punktów, które spełniają nierówność $x^2 + y^2 \leq 1$. Liczbę $4\frac{k}{n}$ uważamy za przybliżenie π .

Uzasadnimy później sensowność powyższego postępowania. Pokażemy istotne uogólnienie. Zaczniemy od wyjaśnienie kwestii losowania.

Metody Monte Carlo bazują na możliwości generowania przez komputer w bardzo krótkim czasie bardzo wielu liczb, które można uważać za wielkości losowe niezależnie od tego, że najczęściej są one generowane przez algorytmy deterministyczne. Takie liczby nazywamy liczbami pseudolosowymi. Po omówieniu sposobów pozyskiwania liczb pseudolosowych powiemy jak mogą być one użyte do przybliżonego wyznaczania całek, a także w problemach optymalizacyjnych.

12.1 Liczby pseudolosowe

Problem pozyskiwania liczb pseudolosowych o dobrych własnościach jest kluczowy dla wielu obliczeń metodami Monte Carlo oraz w symulacjach i nadal pozostaje w centrum uwagi specjalistów.

Jeszcze do niedawna programy komputerowe używały najczęściej następującego algorytmu: dla ustalonych liczb naturalnych a , b i p wybieramy dowolną liczbę naturalną X_0 , zwaną ziarnem (ang. *seed*), a następnie określamy rekurencyjnie ciąg:

$$X_{n+1} = aX_n + b \pmod{p}.$$

Mówiąc inaczej, za każdym razem obliczamy $X'_{n+1} = aX_n + b$, a jako X_{n+1} bierzemy resztę z dzielenia X'_{n+1} przez p – tak więc wszystkie wyrazy rozważanego ciągu są liczbami naturalnymi mniejszymi od p .

Jeżeli parametry a , b , p i X_0 są **odpowiednio dobrane**, to okazuje się, że liczby:

$$U_n = X_n/p$$

mają własności niemal takie same, jak liczby wylosowane niezależnie jedna od drugiej z rozkładu jednostajnego na przedziale $(0, 1)$.

Przykład – 12.2. Niech $a = 5$, $b = 0$, $p = 7$. $X_0 = 4$. Wtedy otrzymujemy kolejno:

4, 6, 2, 3, 1, 5, 4, 6, ... i po podzieleniu przez 7 odpowiednio: .5714285714, .8571428571, .2857142857, .4285714286, .1428571429, .7142857143, .5714285714, .8571428571, ...

Istnieją pewne zasady wybierania parametrów. W szczególności, p powinno być bardzo duże, aby jak najbardziej ograniczyć zjawisko okresowości. Z podobnych względów także a powinno być dużą liczbą, najlepiej względnie pierwszą z p . Natomiast wybór b ma mniejsze znaczenie – często przyjmuje się $b = 0$.

Okazuje się, że przy odpowiednio wybranych parametrach oraz przy zastosowaniu dodatkowych procedur liczb pseudolosowe i ich zestawy mają bardzo dobre własności – potwierdzają to także odpowiednie testy statystyczne.

Przykładowo, program Maple (już w wersji 5) używał generatora liczb pseudolosowych z bardzo dużymi parametrami a oraz p , mianowicie:

$$a = 427419669081, \quad p = 99999999989.$$

Wartość ziarna X_0 można w każdej chwili zadać zgodnie z aktualnymi potrzebami. Może nam na przykład zależeć, aby przy powtórzeniach danego losowania zawsze otrzymywać te same liczby pseudolosowe – zadajemy wtedy taką samą (stałą) wartość X_0 . Z drugiej strony, możemy żądać, aby w każdym losowaniu otrzymywać inne liczby pseudolosowe – można to na przykład osiągnąć, zadając wartość ziarna w zależności od upływu czasu zużytego przez procesor od rozpoczęcia aktualnej sesji.

Mając liczby pseudolosowe z rozkładu jednostajnego na odcinku $[0, 1]$, można, stosując odpowiednią metodę, uzyskać liczby pseudolosowe z innego zadanego rozkładu.

Przykład – 12.3 (liczby pseudolosowe z rozkładu dyskretnego).

Niech rozkład Q będzie zadany przez ciągi: $x_1, x_2, \dots, p_1, p_2, \dots$, czyli $Q(x_i) = p_i$. Definiujemy liczby $F_0 = 0$, $F_i = F_{i-1} + p_i$, $i = 1, 2, 3, \dots$. Aby wylosować liczbę x z rozkładu Q losujemy najpierw liczbę u z rozkładu $U(0, 1)$ i wyznaczamy

$$x = x_k, \text{ gdzie } k \text{ jest takie, że: } u \in (F_{k-1}, F_k].$$

Uzasadnić poprawność tej procedury (ćwiczenie).

Przykład – 12.4 (liczby pseudolosowe z rozkładu ciągłego).

Założmy, że zmienna losowa X ma rozkład Q oraz dystrybuantę ciągłą, F . Dla uproszczenia założmy, że F jest silnie rosnąca. Zauważmy, że zmienna losowa $U = F(X) = F \circ X$ ma rozkład $U(0, 1)$.

Rzeczywiście, dla każdego $0 < u < 1$: $F_U(u) = P(U \leq u) = P(F \circ X \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u$, a to oznacza, że F_U jest dystrybuantą rozkładu $U(0, 1)$.

Aby więc wylosować liczbę x z rozkładu Q , można wylosować liczbę z rozkładu $U(0, 1)$ i wziąć $x = F^{-1}(u)$.

Ponieważ odwracanie dystrybuanty może być trudne (czasochłonne), istnieją również inne metody pozyskiwania liczb pseudolosowych z określonych typów rozkładów.

Obecnie chyba najczęściej stosowaną metodę pozyskiwania liczb pseudolosowych z rozkładu $U(0, 1)$ stanowią różne wersje: [Mersenne Twister Algorithm](#).

12.2 Estymatory i przedziały ufności

Definicja – 12.5 (Próbka prosta). Dana jest rozkład Q . *Próbką prostą* z rozkładu Q nazywamy ciąg niezależnych zmiennych losowych X_1, X_2, \dots, X_n takich, że dla każdego $i = 1, 2, \dots, n$, $P_{X_i} = Q$.

Definicja – 12.6 (Estymator). Jeżeli X_1, X_2, \dots, X_n jest próbką prostą, a $g : \mathbb{R}^n \rightarrow \mathbb{R}$ jest funkcją borelowską, to zmienną losową $g(X_1, X_2, \dots, X_n)$ nazywamy *estymatorem*.

Najczęściej mówiąc o estymatorze, mamy na myśli, że powinien on w jakimś określonym sensie przybliżać pewną wielkość, na przykład nadzieję matematyczną lub inny parametr rozkładu.

Definicja – 12.7 (Estymator silnie zgodny). Estymator $g(X_1, X_2, \dots, X_n)$ parametru a jest *zgodny silnie* $\iff g(X_1, X_2, \dots, X_n) \xrightarrow{1} a$, gdy $n \rightarrow \infty$.

Definicja – 12.8 (Estymator słabo zgodny). Estymator $g(X_1, X_2, \dots, X_n)$ parametru a jest *zgodny słabo* $\iff g(X_1, X_2, \dots, X_n) \xrightarrow{s} a$, gdy $n \rightarrow \infty$.

Powyższe definicje są formalnie błędne, niemniej często używane. Powinno się oczywiście mówić o ciągu o wyrazach $g_n(X_1, X_2, \dots, X_n)$.

Definicja – 12.9 (Estymator nieobciążony). Estymator $g(X_1, X_2, \dots, X_n)$ parametru a jest *nieobciążony* $\iff E(g(X_1, X_2, \dots, X_n)) = a$.

Przykład – 12.10 (Estymator wartości oczekiwanej). Dana jest zmienna losowa X oraz próbka prosta X_1, X_2, \dots, X_n z rozkładu P_X . Niech $m = E(X)$. Definiujemy $\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$. Jak wiemy, jest to nieobciążony estymator parametru m . Natomiast Mocne Prawo Wielkich Liczb dla i.i.d. (10.16) gwarantuje, że jest to estymator silnie zgodny (a więc także słabo zgodny).

Przykład – 12.11. Niech $Y_1 = (Y_{11}, Y_{12}), \dots, Y_n = (Y_{n1}, Y_{n2})$ będą niezależnymi wektorami losowymi o rozkładzie jednostajnym na kwadracie $[-1, 1]^2$. Określamy zmienne losowe:

$$X_i = \begin{cases} 1, & \text{gdy } Y_{i1}^2 + Y_{i2}^2 \leq 1, \\ 0, & \text{gdy } Y_{i1}^2 + Y_{i2}^2 > 1. \end{cases}$$

Rozważamy estymator $4\bar{X}_n$. Zauważmy, że jest to estymator nieobciążony i silnie zgodny liczby π .

Rzeczywiście: Ponieważ Y_i mają rozkład jednostajny, to $P(X_i = 1) = \frac{\pi}{4}$. Mamy więc też $m = E(X_i) = \frac{\pi}{4}$. Co więcej zmienne losowe X_i są niezależne jako funkcje wektorów niezależnych. Z poprzedniego przykładu wynika, że \bar{X}_n jest estymatorem nieobciążonym i silnie zgodnym liczby m . A więc $4\bar{X}_n$ jest estymatorem nieobciążonym i silnie zgodnym liczby $4m = \pi$.

Przykład – 12.12 (Estymatory wariancji). Dana jest zmienna losowa X oraz próbka prosta X_1, X_2, \dots, X_n z rozkładu P_X . Niech $\sigma^2 = D^2(X)$. Definiujemy:

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Można sprawdzić (ćwiczenie), że $E(\hat{\sigma}_n^2) = \frac{n-1}{n}\sigma^2$. Tak więc $\hat{\sigma}_n^2$ nie jest estymatorem nieobciążonym parametru σ^2 (jest *asymptotycznie nieobciążony*, to znaczy: $\lim_n E(\hat{\sigma}_n^2) = \sigma^2$). Jest jednak zgodny silnie. Zauważmy, że:

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2. \quad (12.1)$$

Stosujemy Mocne Prawo Wielkich Liczb do ciągu X_i oraz do ciągu X_i^2 . $\bar{X}_n \xrightarrow{1} E(X)$, $\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{1} E(X^2)$. Stąd: $\hat{\sigma}_n^2 \xrightarrow{1} E(X^2) - E(X)^2 = D^2(X)$. Zauważmy teraz, że:

$$\frac{n}{n-1} \hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

jest nieobciążonym i zgodnym silnie estymatorem σ^2 .

Przykład – 12.13 (kontynuacja przykładu 12.10). Wiemy już, że \bar{X}_n jest „dobrym” estymatorem wartości oczekiwanej m , ale ważnym pytaniem pozostaje: jak bardzo otrzymane wartości tego estymatora różnią się od prawdziwej wartości m ? Oczywiście nie potrafimy odpowiedzieć na to pytanie z całkowitą pewnością! Możemy jednak podać odpowiedź obarczoną niewielkim błędem, co więcej, możemy z góry określić wielkość tego błędu. Ustalamy małą liczbę $\alpha > 0$ (często bierze się 0.05; 0.01) i szukamy takiego przedziału, zwanego przedziałem ufności, (a, b) , że:

$$P(m \in (a, b)) = 1 - \alpha.$$

Oczywiście liczby a, b muszą mieć charakter losowy zależny od wartości próbki. Co więcej, sie mogą one być wyznaczone jednoznacznie. Najczęściej szuka się ich w trzech następujących postaciach:

- (1) $a = \bar{X}_n - \varepsilon$, $b = \bar{X}_n + \varepsilon$.
- (2) $a = -\infty$, $b = \bar{X}_n + \varepsilon$.
- (3) $a = \bar{X}_n - \varepsilon$, $b = \infty$.

Zadanie rozwiązujemy w oparciu o Centralne twierdzenie graniczne, twierdzenie 9.1. Musimy jednak wtedy założyć, że zmienna losowa X , a więc wszystkie zmienne losowe X_1, \dots, X_n , mają skończoną wariancję σ^2 . Pamiętając, że n jest duże możemy przyjąć, że \bar{X}_n ma rozkład normalny $N(m, \frac{\sigma}{\sqrt{n}})$.

Rozwiążmy problem (1). Poprzednia nierówność przyjmuje postać:

$$P(m \in (\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon)) = 1 - \alpha.$$

Otrzymujemy kolejno:

$$\begin{aligned} P(\bar{X}_n \in (m - \varepsilon, m + \varepsilon)) &= 1 - \alpha. \\ \Phi_{m, \frac{\sigma}{\sqrt{n}}}(m + \varepsilon) - \Phi_{m, \frac{\sigma}{\sqrt{n}}}(m - \varepsilon) &= 1 - \alpha. \\ 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma} - 1\right) &= 1 - \alpha, \quad \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) = 1 - \frac{\alpha}{2}, \\ \frac{\varepsilon\sqrt{n}}{\sigma} &= \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \\ \varepsilon &= \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \end{aligned}$$

Problem (2) oraz (3) rozwiązuje się podobnie (ćwiczenie), otrzymując w obydwóch przypadkach wzór:

$$\varepsilon = \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \alpha).$$

W zagadnieniach praktycznych często jednak nie znamy σ . W takiej sytuacji możemy użyć jego estymatora $\sqrt{\frac{n}{n-1}}\hat{\sigma}_n^2$, lub $\sqrt{\hat{\sigma}_n^2} = \hat{\sigma}_n$. W tym drugim przypadku wzory na ε przyjmują postać:

$$\varepsilon = \frac{\hat{\sigma}_n}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

oraz

$$\varepsilon = \frac{\hat{\sigma}_n}{\sqrt{n}}\Phi^{-1}(1 - \alpha).$$

Przykład – 12.14 (kontynuacja przykładu 10.21). Przeprowadzono następane $n = 10000$ symulacji m – średniego czasu oczekiwania Leona na dotarcie do okienka – otrzymując tym razem: realizację \hat{X}_n , $\hat{x} = 7.5402$ oraz $\hat{\sigma} = 1.0532$. Na poziomie ufności $\alpha = 0.95$ przedział ufności dla m wynosi $[7.5196, 7.5609]$. Na poziomie ufności 0.9 przedział ten jest nieco krótszy i wynosi $[7.5229, 7.5575]$.

Przykład – 12.15 (kontynuacja przykładu 10.22). Chcemy obecnie wyznaczyć przybliżoną wartość prawdopodobieństwa zdarzenia

$$A = \left\{ \max_{1 \leq k \leq n} |S_k - E(S_k)| \geq \varepsilon \right\},$$

przy czym $S_k = X_1 + \dots + X_k$, gdzie X_1, \dots, X_n są niezależnymi zmiennymi losowymi o rozkładzie Poissona P_5 . Bierzemy $\varepsilon = 20$, $n = 30$.

Pamiętamy, że $p = P(A) = m = E(\mathbb{1}_A)$. Generujemy więc realizację y_1, \dots, y_N próbki prostej, powiedzmy Y_1, \dots, Y_N z rozkładu $B(1, p)$, $N = 10000$. Mianowicie dla każdego $i = 1, \dots, 10000$ generujemy ciąg liczb x_1, \dots, x_{30} z rozkładu Poissona P_5 , wyliczamy $s_k = y_1 + \dots + y_k$ oraz $E(S_k) = 5k$ i bierzemy

$$y_i = \begin{cases} 1, & \text{gdy } \max_{1 \leq k \leq 30} |s_k - 5k| \geq 20, \\ 0, & \text{gdy } \max_{1 \leq k \leq 30} |s_k - 5k| < 20. \end{cases}$$

Wyliczamy $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$ oraz $\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N y_i^2 - \bar{y}^2$. Ponieważ $y_i^2 = y_i$, więc $\hat{\sigma}^2 = \bar{y} - \bar{y}^2$. Za pomocą Maple otrzymujemy $\bar{y} \approx 0.1518$ i dalej $p \approx 0.1518$, $\hat{\sigma} \approx 0.3588$ oraz dwustronny przedział ufności dla p na poziomie ufności 0.95: $[0.1448, 0.1588]$.

12.3 Całkowanie metodami Monte Carlo

Chcemy obliczyć całkę $J = \int_a^b f(x) dx$.

Metoda I (uogólnienie metody obliczania π), przykład 12.1.

Załóżmy, że f jest ograniczona, dla uproszczenia, że $0 \leq f(x) \leq c$, gdzie $c \in \mathbb{R}$.

Niech $U_1 = (U_{11}, U_{12}), \dots, U_n = (U_{n1}, U_{n2})$ będą niezależnymi wektorami losowymi o rozkładzie jednostajnym na prostokącie $[a, b] \times [0, c]$. Określamy zmienne losowe:

$$X_i = \begin{cases} 1, & \text{gdy } f(U_{i1}) \geq U_{i2}, \\ 0, & \text{gdy } f(U_{i1}) < U_{i2}. \end{cases}$$

Oczywiście zmienne losowe X_i są niezależne i mają rozkład $(0, 1, p)$, gdzie $p = \frac{J}{(b-a)c}$. Mamy więc $m = E(X_i) = p$, $\sigma^2 = D^2(X_i) = p(1-p)$. Ponieważ \bar{X}_n jest nieobciążonym i silnie zgodnym estymatorem m , to $\bar{J}_n = (b-a)c\bar{X}_n$ jest nieobciążonym i silnie zgodnym estymatorem całki J .

Znajdziemy dwustronny przedział ufności dla J na poziomie $1 - \alpha$. Ponieważ dla m takim przedziałem jest $(\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon)$, gdzie $\varepsilon = \frac{\sigma_X}{\sqrt{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2}\right)$, to dla $J = (b-a)c m$ jest przedział: $(\bar{J}_n - \delta_1, \bar{J}_n + \delta_1)$, gdzie:

$$\delta_1 = (b-a)c\varepsilon = (b-a)c \frac{\sqrt{\frac{J}{(b-a)c} \left(1 - \frac{J}{(b-a)c}\right)}}{\sqrt{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) = \sqrt{J(b-a)c - J^2} \frac{1}{\sqrt{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2}\right).$$

Metoda II. Zauważmy, że

$$J = \int_a^b f(x) dx = (b-a) \int_{\mathbb{R}} \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) f(x) dx = (b-a)E(f(X)),$$

gdzie X jest zmienną losową o rozkładzie $U(a, b)$.

Stosując przykład 12.10 do zmiennej $Y = f(X)$, mamy wzór na estymator nadziei $m_Y = E(Y)$ oraz umiemy znaleźć przedziały ufności dla m_Y .

Estymatorem \hat{J} dla całki J jest więc $\hat{J} = (b-a)\bar{Y}_n = \frac{b-a}{n} \sum_{i=1}^n Y_i$, gdzie $Y_i = f(X_i)$, a X_1, X_2, \dots, X_n jest próbka prostą z rozkładu $P_X = U(a, b)$.

Przedziałem ufności dla J jest więc: $((b-a)\bar{Y}_n - \delta_2, (b-a)\bar{Y}_n + \delta_2)$, gdzie

$$\delta_2 = (b-a) \frac{\sigma_Y}{\sqrt{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2}\right),$$

gdzie $\sigma_Y^2 = D^2(Y) = E(f(X)^2) - E(f(X))^2 = \frac{1}{b-a} \int_a^b f(x)^2 dx - \frac{J^2}{(b-a)^2}$.

Tak więc:

$$\delta_2 = \sqrt{(b-a) \int_a^b f(x)^2 dx - J^2} \frac{1}{\sqrt{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2}\right).$$

Wzory na przedział ufności otrzymane w powyższych metodach zawierają wielkość J , której nie znamy, więc nie mogą być bezpośrednio stosowane.

Zauważmy jednak, że metoda II dostarcza potencjalnie lepszy (na ogół mniejszy) przedział ufności niż metoda I. Mianowicie: $(b-a) \int_a^b f(x)^2 dx \leq J(b-a)c$. Czyli $\delta_2 \leq \delta_1$. Przy nieuważnym doborze c ta nierówność może być bardzo istotna.

Faktyczny przedział ufności można wyznaczyć dopiero po przeprowadzeniu eksperymentu.

W metodzie I jest to przedział: $((b-a)c\bar{X}_n - \hat{\delta}_1, (b-a)c\bar{X}_n + \hat{\delta}_1)$, $\hat{\delta}_1 = (b-a)c \frac{\hat{\sigma}_X}{\sqrt{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2}\right)$, gdzie

$$\hat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X}_n)^2.$$

W metodzie II jest to przedział: $((b-a)\bar{Y}_n - \hat{\delta}_2, (b-a)\bar{Y}_n + \hat{\delta}_2)$, $\hat{\delta}_2 = (b-a)\frac{\hat{\sigma}_Y}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$, gdzie $\hat{\sigma}_Y^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2 = \frac{1}{n}\sum_{i=1}^n Y_i^2 - (\bar{Y}_n)^2$.

Przykład – 12.16. Policzmy całkę $J = \int_0^3 x^2 dx$ stosując obie metody.

Losujemy (Maple) $n = 10000$ punktów (z prostokąta $[0, 3] \times [0, 9]$ w metodzie I, z odcinka $[0, 3]$ w metodzie II), ustalamy $\alpha := 0.05$. W tym przypadku umiemy oczywiście wyliczyć bezpośrednio: $J = 9$, a także $\delta_1 = 0.249462688356251$ oraz $\delta_2 = 0.157774057303750$.

W metodzie I otrzymujemy $\bar{J}_n = 8.934300000$ oraz $\hat{\delta}_1 = 0.249003673187557$.

W metodzie II otrzymujemy $\bar{J}_n = 8.91313296682201$ oraz $\hat{\delta}_2 = 0.157905405151449$.

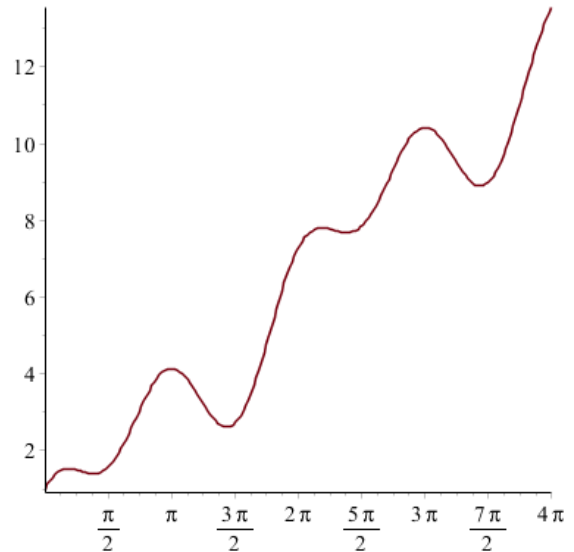
Przykład – 12.17. Porównajmy obydwie metody dla bardziej skomplikowanej funkcji.

Niech $f(x) = x + \sin x + \cos 2x$.

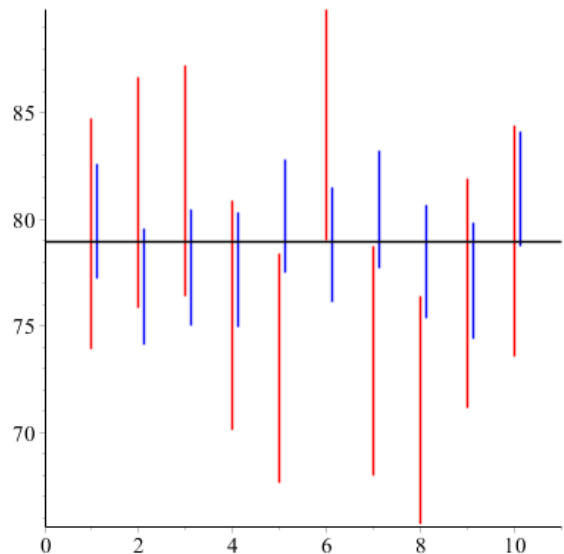
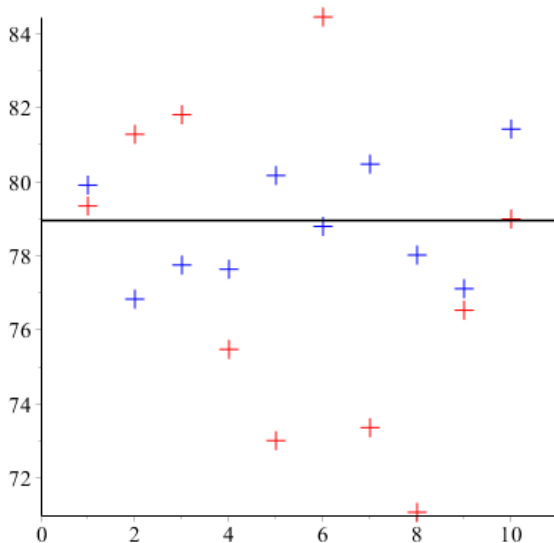
Interesuje nas całka

$$J = \int_0^{4\pi} f(x) dx.$$

Łatwo bezpośrednio obliczyć, że $J = 8\pi^2$.



Uruchamiamy 10 razy metodą I i 10 razy metodą II. Najpierw porównujemy same przybliżenia J , a później odpowiednie przedziały ufności ($\alpha = 0.05$). W każdej z metod przeprowadzamy 2000 losowań z rozkładu jednostajnego. Metoda I – kolor czerwony, metoda II – kolor niebieski.



- Opisanymi dwiema metodami oraz innymi metodami Monte Carlo można obliczać całki wielowymiarowe:

$$\int_D f d\mu, \text{ gdzie } D \subset \mathbb{R}^n, \mu - \text{dana miara.}$$

- W wielu przypadkach metody Monte Carlo są jedynymi metodami obliczania takich całek!

3. Oprócz dwóch omówionych powyżej istnieje więcej metod Monte Carlo obliczania całek. Istotnym problemem jest to, aby do danego problemu dobrać metodę, które dają szansę na mały przedział ufności.
4. Wiele problemów obliczeniowych można sprowadzić do obliczania całek.
5. **Metody Monte Carlo stosowane są też w innych zagadnieniach. W szczególności metodami Monte Carlo rozwiązywane są złożone zadania optymalizacyjne.**

12.4 Optymalizacja

Dana jest funkcja $f : A \rightarrow \mathbb{R}$, gdzie $A \subset \mathbb{R}^n$.

Należy wyznaczyć efektywnie minimum globalne, to znaczy punkt $a \in A$ taki, że

$$\forall x \in A : f(a) \leq f(x),$$

lub jego możliwie najlepsze przybliżenie.

Oznaczamy zbiór rozwiązań:

$$A^* := \{a \in A \mid \text{dla każdego } x \in A : f(a) \leq f(x)\}.$$

Szukanie maximum globalnego funkcji $f \iff$ Szukanie minimum globalnego funkcji $-f$.

Omówimy najprostszy algorytm stochastycznej optymalizacji, Pure Random Search (PRS).

Zakładamy, że $f : A \rightarrow \mathbb{R}$ jest funkcją ciągłą, A jest zwarty. Dla uproszczenia założymy, że $A = [0, 1]^n$.

Algorytm PRS.

1. Losujemy punkt ze zbioru A zgodnie z rozkładem jednostajnym na A . Oznaczamy go jako x_0 . Kolejne punkty x_1, x_2, x_3, \dots , tworzymy w następujący sposób.
Dla $t = 0, 1, 2, 3, \dots$:
2. Losujemy punkt $\mathbf{y}_t \in A$ według rozkładu jednostajnego.
3. Jeżeli $f(\mathbf{y}_t) < f(x_t)$, kładziemy $x_{t+1} = \mathbf{y}_t$. W przeciwnym przypadku kładziemy $x_{t+1} = x_t$.

Okazuje się, że punkty x_t określone powyższym algorytmem zbiegają do zbioru A^* z prawdopodobieństwem 1, to znaczy $\text{dist}(x_t, A^*) \rightarrow 0$, gdy $t \rightarrow \infty$. Bardziej formalnie formułujemy to tak.

Twierdzenie – 12.18. Niech X_t będzie ciągiem zmiennych losowych, których realizacje x_t określa algorytm. Wtedy

$$P(\{\omega : X_t(\omega) \rightarrow A^*, \text{ gdy } t \rightarrow \infty\}) = 1.$$

Dowód. Przypominamy drugą część lematu 10.17.

Lemat Borela–Cantellego Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną.

Niech $C_1, C_2, C_3, \dots \in \Sigma$ będzie ciągiem zdarzeń niezależnych:

$$\sum_{i=1}^{\infty} P(C_i) = \infty \implies P\left(\bigcap_{t=1}^{\infty} \bigcup_{i=t}^{\infty} C_i\right) = 1.$$

Oznaczmy $m = \min_A f$. Wtedy $A^* = \{x \in A : f(x) = m\}$.

Ciąg $f(X_t)$ jest nierosnący i ograniczony z dołu przez m , więc jest zbieżny do pewnej zmiennej losowej $\eta \geq m$.

Niech \mathbf{Y}_t będą niezależnymi zmiennymi losowymi o rozkładzie jednostajnym na A , których realizacjami są punkty \mathbf{y}_i .

Ustalmy $\varepsilon > m$.

Ponieważ funkcja f jest ciągła, zbiór $B = \{x \in A : f(x) < \varepsilon\}$ jest niepusty i otwarty, więc jego miara Lebesgue'a, $\nu(B) = \alpha > 0$.

Określmy:

$$C_t = \{\omega \in \Omega : f(\mathbf{Y}_t(\omega)) < \varepsilon\} = \mathbf{Y}_t^{-1}(B), \text{ dla } t = 1, 2, 3, \dots$$

C_t są zdarzeniami niezależnymi.

Ponieważ zmienne losowe \mathbf{Y}_t mają rozkład ν , więc $P(C_t) = P_{\mathbf{Y}_t}(B) = \nu(B) = \alpha$ dla $t \geq 1$. Jest więc spełnione założenie lematu Borela–Cantellego.

Na podstawie algorytmu PRS zachodzi implikacja:

$$\eta \geq \varepsilon \implies \forall t \geq 1 : f(\mathbf{Y}_t) \geq \eta \geq \varepsilon,$$

czyli

$$\{\eta \geq \varepsilon\} \subset \bigcap_{t=1}^{\infty} (\Omega \setminus C_t) = \Omega \setminus \bigcup_{t=1}^{\infty} C_t,$$

$$\bigcap_{t=1}^{\infty} \bigcup_{i=t}^{\infty} C_i \subset \bigcup_{t=1}^{\infty} C_t \subset \{\eta < \varepsilon\}.$$

Z lematu Borela–Cantellego: $P(\{\eta < \varepsilon\}) = 1$. Ponieważ, $\varepsilon > m$ było ustalone dowolnie więc $\eta = m$. Tak więc: $P(f(X_t) \rightarrow m, \text{ gdy } t \rightarrow \infty) = 1$.

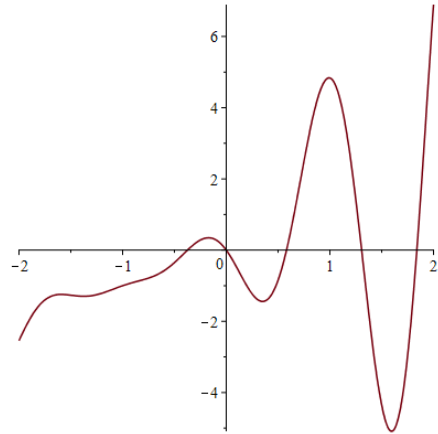
Z ciągłości f oraz zwartości A , mamy $P(X_t \rightarrow A^*, \text{ gdy } t \rightarrow \infty) = 1$. □

Przykład – 12.19. Szukamy minimum i maksimum globalnego funkcji $f(x) = x - (x + 1)^2 \sin(5x)$ na przedziale $A = [-2, 2]$.

Stosując metodę PRS i wykonując 1000 kroków otrzymujemy:

$$a_{min} = 1.5937765, \quad f(a_{min}) = -5.0895387,$$

$$a_{max} = 1.9991702, \quad f(a_{max}) = 6.86129796.$$



Rozwiązywanie równań, bądź układów równań, można sprowadzić do problemu poszukiwania minimum globalnego. Rzeczywiście, zamiast szukać rozwiązania układu $f_i(x) = 0, i = 1, \dots, n$ można szukać minimum globalnego funkcji $F(x) = \sum_{i=1}^n f_i(x)^2$.

Przykład – 12.20. Znajdziemy przybliżenie rozwiązania równania:

$$\sqrt{x^2 + e^{-x}} = x - 2 \cos(2x)$$

w przedziale $[0, 1]$.

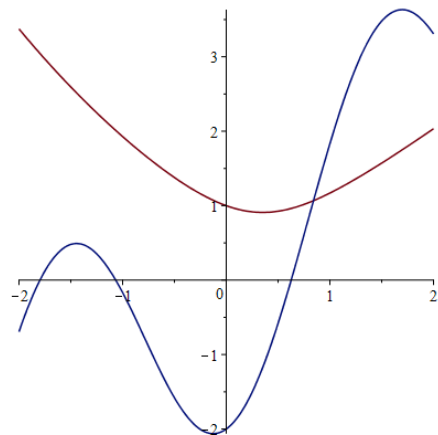
Stosujemy metodę PRS do funkcji $(f(x) - g(x))^2$, gdzie $f(x)$ oznacza lewą, a $g(x)$ prawą stronę równania. Po wykonaniu 1000 kroków otrzymamy:

$$x^* = 0.84191450,$$

przy czym

$$f(x^*) = 1.067569590,$$

$$g(x^*) = 1.067498775.$$



12.5 Pytania

Pytanie 12.1. Liczby 0.657, 0.773, 0.801, 0.501, 0.123, 0.202 wylosowano z rozkładu $U(0, 1)$. Na tej podstawie wygenerować liczby z rozkładu $B(3, 1/2)$. ROZWIĄZANIE

Pytanie 12.2. Niech X_1, X_2, X_3, \dots będą i.i.d. z rozkładu $U(0, a)$, $a > 0$. Zbadać własności estymatora parametru a : $M_n = \max(X_1, \dots, X_n)$. ROZWIĄZANIE

Pytanie 12.3. Dana jest funkcja $f : [a, b] \rightarrow \mathbb{R}$, o której wiadomo, że spełnia warunek $|f(x)| \leq M$ dla każdego $x \in [a, b]$. Jak można estymować $\int_a^b f(x) dx$, używając metody I z wykładu? ROZWIĄZANIE

Pytanie 12.4. Aby obliczyć całkę $J = \int_{-1}^1 f(x) dx$, gdzie $0 < f(x) < 1$ zastosowano metodę I. Wylosowano 1000 punktów z prostokąta $[-1, 1] \times [0, 1]$ i okazało się, że 360 z nich leży pod wykresem funkcji. Na poziomie ufności 0.95 wskazać przedział ufności dla J . ROZWIĄZANIE

Pytanie 12.5. W celu estymacji całki $J = \int_a^b f(x) dx$ wylosowano n punktów x_i według rozkładu jednostajnego $U([a, b])$ i wyznaczono sumy $s = f(x_1) + \dots + f(x_n)$ oraz $s_{kw} = f(x_1)^2 + \dots + f(x_n)^2$. Wskazać przedział ufności dla J postaci (c, ∞) na poziomie ufności $1 - \alpha$. ROZWIĄZANIE

Pytanie 12.6. Uzasadnić ostatecznie zdanie w dowodzie twierdzenia 12.18. ROZWIĄZANIE

Rozdział 13

Warunkowa wartość oczekiwana

13.1 Wartości oczekiwana rozkładów warunkowych

Zdefiniowaliśmy poprzednio rozkłady warunkowe w przypadku dwuwymiarowego wektora losowego o rozkładzie dyskretnym lub ciągłym. Można to powtórzyć dla wyżej wymiarowych wektorów losowych:

Niech (X, Y) będzie $n \times m$ -wymiarowym wektorem losowym o dyskretnym rozkładzie danym przez $\{(x_i, y_j)\}, \{p_{ij}\}$, czyli $P(X = x_i, Y = y_j) = p_{ij}$. Niech:

$$p_{j|i} = P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i\cdot}} = \frac{p_{ij}}{\sum_k p_{ik}}. \quad (13.1)$$

Rozkład dany przez ciągi $\{y_j\}, \{p_{j|i}\}$ nazywamy rozkładem warunkowym Y pod warunkiem $X = x_i$. Oznaczamy go jako $P_{Y|X=x_i}$.

Jeżeli $m = 1$, czyli gdy Y jest zmienną losową, $P_{Y|X=x_i}$ jest rozkładem jednowymiarowym i wtedy można mówić o jego nadziei matematycznej, patrz uwaga 6.10. Jeżeli istnieje, to dla powyższego rozkładu będzie to liczba oznaczana przez $E(Y|X = x_i)$, a więc:

$$E(Y|X = x_i) = \sum_j y_j p_{j|i}.$$

Przykład – 13.1. Przypomnijmy przykład dotyczący wektora losowego (X, Y) o rozkładzie podanym w tabelce:

| $X \setminus Y$ | 1 | 2 | 3 | 4 | 5 | 6 | X |
|-----------------|------|------|------|------|------|-------|------|
| 0 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/6 |
| 1 | 0 | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 5/12 |
| 2 | 0 | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 5/12 |
| Y | 1/36 | 3/36 | 5/36 | 7/36 | 9/36 | 11/36 | |

X oraz Y były określone w kontekście rzutu dwiema kostkami:

X – numer kostki na której wypadła większa liczba, lub 0, gdy liczby były równe.

Y – maksimum oczek na dwóch kostkach.

Rozkładem warunkowym $P_{Y|X=0}$ jest rozkład jednostajny w punktach 1, 2, 3, 4, 5, 6, a więc $E(Y|X = 0) = 3.5$.

Rozkład warunkowy $P_{Y|X=2}$ jest określony przez ciągi 2, 3, 4, 5, 6 oraz 1/15, 2/15, 3/15, 4/15, 5/15, a więc $E(Y|X = 2) = 70/15 = 14/3$.

Rozkład warunkowy $P_{X|Y=4}$ jest określony przez ciągi 0, 1, 2 oraz $\frac{1}{7}, \frac{3}{7}, \frac{3}{7}$, a więc $E(X|Y = 4) = \frac{10}{7}$.

Niech (X, Y) będzie $n \times m$ -wymiarowym wektorem losowym o ciągłym rozkładzie danym przez gęstość f .

$$f_{Y|X=x}(y) = f(y|x) = \begin{cases} \frac{f(x,y)}{\int_{\mathbb{R}} f(x,y) dy} = \frac{f(x,y)}{f_X(x)}, & \text{gdy } f_X(x) > 0, \\ 0, & \text{gdy } f_X(x) = 0. \end{cases} \quad (13.2)$$

Tutaj f_X oznacza gęstość wektora losowego X . Przy ustalonym x , takim, że $f_X(x) > 0$ funkcja $y \rightarrow f(y|x)$ jest gęstością. Zakładając, że $m = 1$ możemy więc, podobnie jak w przypadku dyskretnym mówić o nadziei matematycznej $E(Y|X = x)$ określonej (o ile istnieje) jako:

$$E(Y|X = x) = \int_{\mathbb{R}} y f(y|x) dy.$$

Oczywiście można w przypadku dyskretnym i w przypadku ciągłym zdefiniować także $E(X|Y = y)$ dla zmiennej losowej X i dowolnego wektora losowego Y .

$$E(X|Y = y_j) = \sum_i x_i p_{i|j}, \quad E(X|Y = y) = \int_{\mathbb{R}} x f(x|y) dx.$$

Przykład – 13.2 (c.d. przykładu 5.21). Losujemy według rozkładu jednostajnego liczbę a z odcinka $[0, 1]$ a następnie według rozkładu jednostajnego liczbę b z odcinka $[0, a]$. X oraz Y są zmiennymi losowymi odpowiadającymi powyższymi losowaniami. Pamiętamy, że gęstość warunkowa $f(y|x)$ była dana jako $f(y|x) = \frac{1}{x} \mathbb{1}_{[0,x]}(y)$, dla $0 < x \leq 1$ oraz 0 w przeciwnym przypadku. W takim razie: $E(Y|X = x) = \int_{\mathbb{R}} y f(y|x) dy = \int_0^x \frac{y}{x} dy = \frac{x}{2}$ dla $0 < x \leq 1$.

Pamiętamy, że gęstość wektora losowego f można otrzymać jako iloczyn: $f(x, y) = f(y|x)f_X(x) = \frac{1}{x}$ dla $0 < y \leq x \leq 1$ oraz 0 w przeciwnym przypadku.

Można więc wyliczyć gęstość warunkową dla $0 < y \leq x \leq 1$:

$$f(x|y) = \frac{f(x, y)}{\int_{\mathbb{R}} f(x, y) dx} = \frac{\frac{1}{x}}{\int_y^1 \frac{1}{x} dx} = \frac{1}{x} \frac{1}{-\ln y}$$

i dalej

$$E(X|Y = y) = \int_{\mathbb{R}} x f(x|y) dx = \int_y^1 x \frac{1}{x} \frac{1}{-\ln y} dx = \frac{1 - y}{-\ln y}.$$

Naszym celem będzie podanie definicji oraz własności i zastosowań warunkowej nadziei matematycznej zmiennej losowej Y względem wektora losowego X , $E(Y|X)$, a także pojęcia bardziej ogólnego, nadziei matematycznej zmiennej losowej Y względem σ -algebry \mathcal{A} , $E(Y|\mathcal{A})$. Aby jednak nawiązać do wspomnianych właśnie sytuacji szczególnych zrobimy kilka wstępnych uwag.

Uwaga – 13.3. *Każdy wektor losowy określony na przestrzeni probabilistycznej (Ω, Σ, P) , $X : \Omega \rightarrow \mathbb{R}^n$ generuje pewną σ -algebrę $\sigma(X) \subset \Sigma$. Mianowicie:*

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^n)\}.$$

Jest to oczywiście najmniejsza σ -algebra przy której X jest odwzorowaniem mierzalnym.

Przykład – 13.4. Gdy X ma rozkład dyskretny wyznaczony przez ciągi $\{x_i\}$, $\{p_i\}$, $i = 1, \dots, N$, $N \leq \infty$, to $\sigma(X)$ jest generowana przez rozkład zbioru Ω na przeciwobrazy $X^{-1}(x_i)$, czyli jest rodziną wszystkich możliwych sum przeciwobrazów $X^{-1}(x_i)$.

Z poprzedniego semestru znamy już podstawowe twierdzenie 6.8:

Twierdzenie. Niech $X : \Omega \rightarrow \mathbb{R}^n$ będzie wektorem losowym, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ funkcją borelowską. Wtedy:

$$E(g(X)) = \int_{\mathbb{R}^n} g dP_X,$$

przy czym obydwie strony istnieją jednocześnie.

Możemy teraz nieco uogólnić ten wzór:

Twierdzenie – 13.5. *Niech $X : \Omega \rightarrow \mathbb{R}^n$ będzie wektorem losowym, $h : \mathbb{R}^n \rightarrow \mathbb{R}$ funkcją borelowską, $B \in \mathcal{B}(\mathbb{R}^n)$. Wtedy:*

$$\int_{X^{-1}(B)} h(X) dP = \int_B h dP_X,$$

przy czym obydwie strony istnieją jednocześnie.

Dowód. W poprzednim twierdzeniu wystarczy wziąć: $g(x) = \mathbb{1}_B(x) \cdot h(x)$, dla $x \in \mathbb{R}^n$, gdzie $\mathbb{1}_B$ jest funkcją charakterystyczną zbioru B . \square

Podobnie można uogólnić wzór na nadzieję matematyczną dla rozkładów dyskretnych i ciągłych.

Twierdzenie – 13.6. Niech wektor X ma rozkład ciągły zadany przez gęstość $f : \mathbb{R}^n \rightarrow \mathbb{R}$. $h : \mathbb{R}^n \rightarrow \mathbb{R}$ jest funkcją borelowską, $B \in \mathcal{B}(\mathbb{R}^n)$. Wtedy:

$$\int_{X^{-1}(B)} h(X) dP = \int_B h(x) f(x) dx,$$

przy czym obydwie strony istnieją jednocześnie. Całkowanie odbywa się według miary Lebesgue'a.

Dowód. (Ćwiczenie). \square

Sformułować odpowiednią wersję w przypadku rozkładów dyskretnych (ćwiczenie).

Wróćmy do nadziei rozkładu warunkowego $E(Y|X = x)$, w przypadku wektorów losowych (X, Y) o rozkładach dyskretnych i ciągłych. Możemy teraz rozważać następujące odwzorowanie:

$$\varphi : \Omega \ni \omega \longrightarrow E(Y|X = X(\omega)).$$

Twierdzenie – 13.7. W przypadku, gdy wektor losowy (X, Y) ma rozkład dyskretny, lub rozkład ciągły, odwzorowanie φ , o ile jest dobrze określone, spełnia dwa warunki:

(M) φ jest $\sigma(X)$ mierzalne.

(C) Dla każdego $A \in \sigma(X) : \int_A \varphi dP = \int_A Y dP$.

To znaczy: φ jest zmienną losową na przestrzeni probabilistycznej $(\Omega, \sigma(X), P)$, taką, że na wszystkich zbiorach z $\sigma(X)$ ma takie same całki (można mówić o średnich) co zmienna losowa Y .

Dowód. Przypadek dyskretny. Ustalmy punkt x_i . Wtedy φ jest funkcją stałą na zbiorze $X^{-1}(x_i)$ równą $E(Y|X = x_i)$, a to oznacza mierzalność względem σ -algebry generowanej przez te zbiory. Niech $A \in \sigma(X)$. Wtedy A jest sumą co najwyżej przeliczalną zbiorów postaci $X^{-1}(x_i)$. Całka po A jest więc sumą całek po tych zbiorach. Natomiast

$$\begin{aligned} \int_{X^{-1}(x_i)} \varphi dP &= E(Y|X = x_i)P(X^{-1}(x_i)) = \sum_j y_j p_{j|i} p_i = \frac{\sum_j y_j p_{ij}}{p_i} p_i \\ &= \sum_j y_j p_{ij} = \sum_j y_j P(X = x_i, Y = y_j) = \sum_j \int_{\{X=x_i, Y=y_j\}} Y dP = \int_{X^{-1}(x_i)} Y dP. \end{aligned}$$

\square

Przypadek ciągły. Funkcja φ wyraża się wzorem:

$$\begin{aligned} \varphi(\omega) &= \int_{\mathbb{R}} y(f(y|X(\omega))) dy, \quad \text{gdzie} \\ f(y|x) &= \begin{cases} \frac{f(x,y)}{\int_{\mathbb{R}} f(x,y) dy} = \frac{f(x,y)}{f_X(x)}, & \text{gdyn } f_X(x) > 0, \\ 0, & \text{gdyn } f_X(x) = 0. \end{cases} \end{aligned}$$

Widać, że φ jest złożeniem funkcji mierzalnych względem $\sigma(X)$, a więc jest $\sigma(X)$ -mierzalne. Niech $A \in \sigma(X)$. Oznacza to, że $A = X^{-1}(B)$, gdzie $B \in \mathcal{B}(\mathbb{R}^n)$. $B = B_1 \cup B_2$, gdzie $B_1 = \{x \in B : f_X(x) > 0\}$, $B_2 = \{x \in B : f_X(x) = 0\}$. Wtedy $A = X^{-1}(B_1) \cup X^{-1}(B_2) = A_1 \cup A_2$.

Zauważmy najpierw, że $P(A_2) = 0$. Rzeczywiście: $P(A_2) = P_X(B_2) = \int_{B_2} f_X(x) dx = 0$. Tak więc $\int_{A_2} \varphi dP = \int_{A_2} Y dP = 0$.

Natomiast stosując dwukrotnie twierdzenie 13.6 dotyczące zmiany całkowania względem miary P na całkowanie przy użyciu gęstości f_X oraz f , z twierdzenia Fubniego mamy:

$$\begin{aligned} \int_{A_1} \varphi dP &= \int_{B_1} \frac{\int_{\mathbb{R}} y f(x,y) dy}{f_X(x)} f_X(x) dx = \int_{B_1} \int_{\mathbb{R}} y f(x,y) dy dx \\ &= \int_{B_1 \times \mathbb{R}} y f(x,y) d(x,y) = (\text{bo } A_1 = (X, Y)^{-1}(B_1 \times \mathbb{R})) \int_{A_1} Y dP. \end{aligned}$$

Czyli $\int_A \varphi dP = \int_A Y dP$. \square

13.2 Twierdzenie Radona–Nikodyma

Przypomnijmy twierdzenie Radona–Nikodyma, gdyż stanowi ono klucz w kolejnych rozważaniach dotyczących nadziei warunkowych.

Niech \mathcal{A} będzie σ -algebrą na zbiorze Ω .

Funkcję $\lambda : \mathcal{A} \rightarrow \mathbb{R}$ nazywamy przeliczalnie addytywną, jeżeli dla każdego ciągu parami rozłącznych zbiorów $A_1, A_2, A_3, \dots \in \mathcal{A}$

$$\lambda\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \lambda(A_i).$$

Każda miara skończona, w szczególności każda miara probabilistyczna jest przeliczalnie addytywna.

Mówimy, że przeliczalnie addytywna funkcja λ jest absolutnie ciągła względem miary $\mu : \mathcal{A} \rightarrow [0, \infty]$, piszemy często $\lambda \ll \mu$, jeżeli dla każdego $A \in \mathcal{A}$ zachodzi implikacja:

$$\mu(A) = 0 \implies \lambda(A) = 0.$$

Przykład – 13.8. Jeżeli $\mu : \mathcal{A} \rightarrow [0, \infty]$ jest miarą, a $g : \Omega \rightarrow \mathbb{R}$ taką \mathcal{A} -mierzalną funkcją, że $\int_{\Omega} g d\mu \in \mathbb{R}$, to λ zdefiniowana jako:

$$\lambda(A) = \int_A g d\mu \tag{13.3}$$

jest przeliczalnie addytywną funkcją, absolutnie ciągłą względem miary μ .

Przy pewnym założeniu powyższą implikację można odwrócić. Mówi o tym twierdzenie Radona–Nikodyma.

Mówimy, że miara μ jest σ -skończona, jeżeli istnieją takie zbiory $A_i \in \mathcal{A}$, że $\Omega = \bigcup_{i=1}^{\infty} A_i$ oraz $\mu(A_i) < \infty$.

Każda miara probabilistyczna jest σ -skończona. Miara Lebesgue'a jest σ -skończona.

Twierdzenie – 13.9 (Radon–Nikodym). *Jeżeli λ jest przeliczalnie addytywną funkcją, absolutnie ciągłą względem σ -skończonej miary μ , to istnieje \mathcal{A} -mierzalna funkcja $g : \Omega \rightarrow \mathbb{R}$, że dla każdego $A \in \mathcal{A}$*

$$\lambda(A) = \int_A g d\mu.$$

Jeżeli \mathcal{A} -mierzalna funkcja $h : \Omega \rightarrow \mathbb{R}$ spełnia dla każdego $A \in \mathcal{A}$ ten sam warunek, to g, h są równe μ -prawie wszędzie, to znaczy to $\mu(\{\omega : g(\omega) \neq h(\omega)\}) = 0$

Dowód. Pomijamy. □

Uwaga – 13.10. *Poznaliśmy wcześniej definicję rozkładu ciągłego. Był to rozkład Q , który ma gęstość, powiedzmy f , czyli dla każdego borelowskiego zbioru A : $Q(A) = \int_A f dx$, gdzie całkowanie odbywa się względem miary Lebesgue'a, μ_L . Powyższe jednak oznacza, że $Q \ll \mu_L$. Z twierdzenia Radona–Nikodyma wynika więc, że rozkład Q jest ciągły, wtedy i tylko wtedy, gdy miara Q jest absolutnie ciągła względem miary μ_L . Dlatego też część autorów używa terminologii „rozkład absolutnie ciągły” rezerwując termin „rozkład ciągły” do opisanie sytuacji w której $Q(\{a\}) = 0$ dla wszystkich a .*

W tym kursie pozostajemy przy częściej stosowanej terminologii.

13.3 Warunkowa wartość oczekiwana – sytuacja ogólna

Nadzieja warunkowa jest jednym z najważniejszych pojęć rachunku prawdopodobieństwa. Jest kilka obiektów, które określa się tym pojęciem i warto zrozumieć różnice i związki między nimi. Jak dotychczas wspomnieliśmy o wielkości $E(Y|X = x)$ i była ona określona jako „zwykła” nadzieja rozkładu warunkowego $P_{Y|X=x}$, ale zakładaliśmy, że (X, Y) ma rozkład dyskretny albo rozkład ciągły. A gdy tak nie jest to co? To właśnie teraz zobaczymy. Najważniejszym będzie zdefiniowanie pewnej zmiennej losowej, którą też nazwiemy nadzieją warunkową. Nie będzie to definicja konstruktywna tylko poprzez wymienienie własności, które ta zmienna losowa ma spełniać. Niedawno mówiliśmy już o tych własnościach, nazwaliśmy je (M) oraz (C) i pokazaliśmy, że istnieje obiekt, który je posiada, patrz twierdzenie 13.7. Więc nasza definicja nie będzie dotyczyć nieistniejących obiektów! Jednak pokażemy coś więcej, mianowicie, że w każdych okolicznościach istnieje zmienna losowa, która spełnia warunki (M), (C). I to tylko jedna

z dokładnością do zbiorów miary zero! I to jest niezwykle ważne (i piękne), gdyż dzięki temu będzie można uzyskać szereg ważnych wyników. Na przykład już wkrótce powiemy, że warunkowanie obniża wariancje, co jest kolosalnie ważne w statystyce oraz w metodach Monte Carlo. A nawet wcześniej przedstawimy sposoby obliczania „zwykłej” nadziei poprzez warunkowania.

Prostą konsekwencją twierdzenia Radona–Nikodyma jest następujące:

Twierdzenie – 13.11. Niech Y będzie zmienną losową określoną na przestrzeni probabilistycznej (Ω, Σ, P) , $\mathcal{A} \subset \Sigma$ σ -algebrą. Zakładamy, że $E(Y) \in \mathbb{R}$. Wtedy:

Istnieje odwzorowanie $\varphi : \Omega \rightarrow \mathbb{R}$ spełniająca warunki:

(M) φ jest \mathcal{A} mierzalna.

(C) Dla każdego $A \in \mathcal{A} : \int_A \varphi dP = \int_A Y dP$.

Jeżeli odwzorowanie $\psi : \Omega \rightarrow \mathbb{R}$ spełnia warunki (M) oraz (C), to $\varphi = \psi$ prawie wszędzie (skrót pw.), to znaczy $P(\{\omega \in \Omega : \varphi(\omega) = \psi(\omega)\}) = 1$.

Dowód. Można skorzystać z twierdzenia Radona–Nikodyma zastosowanego do funkcji λ określonej jako $\lambda(A) = \int_A Y dP$ dla $A \in \mathcal{A}$ (ponieważ $E(Y) \in \mathbb{R}$ jest ona przeliczalnie addytywna) oraz miary P . \square

Powyższe twierdzenie powoduje, że następująca definicja ma sens.

Definicja – 13.12 (Nadzieja warunkowa względem σ -algebry.).

$$E(Y|\mathcal{A}) = \{\varphi : \Omega \rightarrow \mathbb{R} : \varphi \text{ spełnia warunki (M) oraz (C)}\}.$$

Poprzednie twierdzenie zapewnia, że $E(Y|\mathcal{A})$ jest zbiorem niepustym, a każde dwa jego elementy są sobie równe prawie wszędzie. Najczęściej (nieformalnie) nie rozróżnia się $E(Y|\mathcal{A})$ od jego elementów, czyli traktujemy $E(Y|\mathcal{A})$ jako odwzorowanie spełniające (M) oraz (C).

Przykład – 13.13 (zupełny brak informacji). Niech $\mathcal{A} = \{\emptyset, \Omega\}$. Wtedy każda funkcja stała spełnia (M). Gdy stała ta równa się $E(Y)$, spełniony jest także warunek (C). Tak więc:

$$E(Y|\mathcal{A}) = E(Y).$$

Przykład – 13.14 (pełna informacja). Niech $\mathcal{A} = \Sigma$. Wtedy sama zmienna losowa Y spełnia warunki (M) oraz (C).

$$E(Y|\mathcal{A}) = Y.$$

Przykład – 13.15 (częściowa informacja). Niech $A_i \in \Sigma$, $i = 1, 2, 3, \dots, N$, $N \leq \infty$, będzie rozkładem Ω : $\Omega = \bigcup_{i=1}^N A_i$, $A_i \cap A_j = \emptyset$ dla $i \neq j$. Zakładamy, że $P(A_i) > 0$ dla wszystkich i . Niech $\mathcal{A} = \sigma(A_i : i = 1, 2, 3, \dots, N)$. Wtedy:

$$E(Y|\mathcal{A})(\omega) = \frac{\int_{A_i} Y dP}{P(A_i)}, \quad \text{dla } \omega \in A_i.$$

Wyraźnie widać, że powyższa funkcja jest stała na każdym zbiorze A_i , jest więc \mathcal{A} -mierzalna. Ponieważ każdy zbiór $A \in \mathcal{A}$ jest pewną sumą rozłącznych zbiorów A_i , więc warunek (C) wystarczy sprawdzić na każdym A_i , co jest oczywiste (ćwiczenie).

Przykład – 13.16. Przypuśćmy, że wektor losowy X ma rozkład dyskretny skupiony w punktach x_i , $i = 1, 2, 3, \dots, N$, $N \leq \infty$. Biorąc $A_i = X^{-1}(x_i)$, mamy sytuację taką jak w poprzednim przykładzie; teraz $\mathcal{A} = \sigma(X)$. W takim razie:

$$E(Y|\sigma(X))(\omega) = \frac{\int_{X=x_i} Y dP}{P(X=x_i)}, \quad \text{gdy } X(\omega) = x_i.$$

W sytuacji, gdy wektor (X, Y) ma rozkład dyskretny określony przez $(\{(x_i, y_j)\}, \{p_{ij}\})$, mamy:

$$\int_{X=x_i} Y dP = \sum_j y_j p_{ij}, \quad P(X=x_i) = \sum_j p_{ij},$$

więc

$$E(Y|\sigma(X))(\omega) = \frac{\sum_j y_j p_{ij}}{\sum_j p_{ij}} = E(Y|X=x_i), \quad \text{gdy } X(\omega) = x_i,$$

gdzie $E(Y|X=x_i)$ oznacza nadzieję matematyczną rozkładu warunkowego $P_{Y|X=x_i}$.

Nadzieja warunkowa względem zdarzenia Czasami używa się określenia: nadzieja matematyczna warunkowa Y pod warunkiem $W \in \Sigma$ i definiuje się ją jako

$$E(Y|W) = \frac{\int_W Y dP}{P(W)},$$

zakładając jednak, że $P(W) > 0$.

W sytuacji opisanej w przykładzie 13.15 widzimy, że $E(Y|A_i) = E(Y|\mathcal{A})(\omega)$ dla $\omega \in A_i$.

UWAGA. Mamy tutaj niestety pewną kolizję oznaczeń. $E(Y|X = x)$ nie zawsze oznacza $E(Y|\{\omega : X(\omega) = x\})$. Chociaż, gdy X ma rozkład dyskretny oraz $P(X = x_i) > 0$, to te dwie wielkości są sobie równe.

Definicja – 13.17 (Nadzieja warunkowa względem wektora losowego). Niech $Y : \Omega \rightarrow \mathbb{R}$ będzie zmienną losową, $E(Y) \in \mathbb{R}$. Niech $X : \Omega \rightarrow \mathbb{R}^k$ będzie wektorem losowym. Definiujemy:

$$E(Y|X) = E(Y|\sigma(X)).$$

Z twierdzenia 13.7 wynika następująca uwaga.

Uwaga – 13.18. Gdy (X, Y) jest wektorem losowym określonym na przestrzeni probabilistycznej (Ω, Σ, P) o rozkładzie dyskretnym albo ciągłym, $E(Y) \in \mathbb{R}$, to

$$E(Y|X)(\omega) = E(Y|X = X(\omega)).$$

Inaczej:

$$E(Y|X) = \alpha(X) = \alpha \circ X,$$

gdzie $\alpha(x) = E(Y|X = x)$ dla tych x dla których w tych przypadkach została zdefiniowana $E(Y|X = x)$.

$E(Y|X)$ jest zawsze pewną funkcją X .

Twierdzenie – 13.19. Niech (X, Y) będzie takim wektorem losowym, że $X : \Omega \rightarrow \mathbb{R}^k$, $Y : \Omega \rightarrow \mathbb{R}$, $E(Y) \in \mathbb{R}$. Wtedy istnieje funkcja borelowska $\alpha : \mathbb{R}^k \rightarrow \mathbb{R}$, taka, że $E(Y|X) = \alpha(X)$.

Twierdzenie to jest w istocie wnioskiem z bardziej ogólnego twierdzenia.

Twierdzenie – 13.20. Niech $X : \Omega \rightarrow \mathbb{R}^k$ będzie wektorem losowym oraz $Z : \Omega \rightarrow \mathbb{R}$. Wtedy: Z jest odwzorowaniem $\sigma(X)$ mierzalnym \iff Istnieje taka funkcja borelowska $\alpha : \mathbb{R}^k \rightarrow \mathbb{R}$, że $Z = \alpha \circ X$.

Dowód. „ \implies ” Rozważamy przypadki:

- I. $Z = \mathbb{1}_A$, gdzie $A = X^{-1}(B)$, B jest zbiorem borelowskim w \mathbb{R}^k . Wtedy wystarczy wziąć: $\alpha = \mathbb{1}_B$.
- II. Z jest funkcją prostą postaci $Z = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$, gdzie $A_i \in \sigma(X)$. Wtedy bierzemy: $\alpha = \sum_{i=1}^n c_i \alpha_i$, gdzie α_i są wybrane jak w punkcie I.
- III. Z jest dowolną funkcją $\sigma(X)$ mierzalną. Istnieje wtedy ciąg funkcji prostych $\sigma(X)$ mierzalnych taki, że dla każdego $\omega \in \Omega$ $\lim_{n \rightarrow \infty} Z_n(\omega) = Z(\omega)$. Na podstawie II istnieją funkcje borelowskie α_n takie, że dla wszystkich n $Z_n = \alpha_n \circ X$.

Definiujemy funkcję $\alpha : \mathbb{R}^k \rightarrow \mathbb{R}$ jako:

$$\alpha(x) = \begin{cases} \lim_{n \rightarrow \infty} \alpha_n(x), & \text{gdy } x \in X(\Omega), \\ 0, & \text{gdy } x \notin X(\Omega). \end{cases} \quad (13.4)$$

Oczywiście α jest borelowska (dlaczego?). Dla $\omega \in \Omega$ zachodzi wzór:

$$Z_n(\omega) = \lim_{n \rightarrow \infty} \alpha_n(X(\omega)),$$

więc $X(\omega)$ jest punktem x , w którym istnieje $\lim_{n \rightarrow \infty} \alpha_n(x)$ i jest ona równa $Z(\omega)$. Czyli:

$$Z = \alpha \circ X.$$

„ \impliedby ” Dla dowolnego $B \in \mathcal{B}(\mathbb{R})$ mamy $Z^{-1}(B) = (\alpha \circ X)^{-1}(B) = X^{-1}(\alpha^{-1}(B)) \in \sigma(X)$. □

Zauważmy, że we wzorze (13.4) można by zadać wartość $\alpha(x)$ dla $x \notin X(\Omega)$ na wiele różnych sposobów i nie zmieniłoby to dalszego rozumowania. Tak więc funkcja α nie jest wyznaczona jednoznacznie na zbiorze $\mathbb{R}^k \setminus X(\Omega)$.

Gdy wektor losowy (X, Y) ma rozkład dyskretny lub rozkład ciągły, możemy w sposób naturalny mówić o nadziejach warunkowych $E(Y|X = x)$ – tak postąpiliśmy na początku tego rozdziału. Możemy jednak rozszerzyć określenie $E(Y|X = x)$ nie zakładając nic o rozkładach. Mianowicie możemy postawić następującą definicję:

Definicja – 13.21. Warunkową nadzieją matematyczną Y pod warunkiem $X = x$ (gdzie $x \in \mathbb{R}^k$) nazywamy liczbę

$$E(Y|X = x) := \alpha(x),$$

gdzie α jest funkcją określoną w twierdzeniu 13.19.

Ze względu na możliwą niejednoznaczność funkcji α wielkość powyższa nie jest jednoznacznie określona dla wszystkich $z \in \mathbb{R}^k$. Nie ma to jednak istotnego znaczenia. Na przykład, gdy X ma rozkład dyskretny skupiony na zbiorze K , to wartości funkcji α poza tym zbiorem są nieistotne. Tak więc, gdy $x \notin K$, wartości $E(Y|X = x)$ są niejednoznacznie określone, ale nie ma to dla nas żadnego znaczenia.

Zawsze można mówić o nadziei warunkowej $E(Y|X)$; jest to pewna **zmienna losowa**. Natomiast twierdzenie 13.19 gwarantuje, że zawsze też można mówić o nadziei warunkowej $E(Y|X = x)$; **jest to liczba**. W przypadku, gdy wektor losowy ma rozkład dyskretny lub rozkład ciągły pokazaliśmy, że powyższa definicja $E(Y|X = x)$ pokrywa się z naturalną definicją postawioną w tamtych przypadkach. Także w wielu innych przypadkach można w sposób naturalny zinterpretować $E(Y|X = x)$.

Przykład – 13.22 (c.d. przykładu 13.1). $E(Y|X)$ jest zmienną losową przyjmującą wartości $\frac{7}{2}, \frac{14}{3}, \frac{14}{3}$ z prawdopodobieństwami $\frac{1}{6}, \frac{5}{12}, \frac{5}{12}$. Czyli $\alpha(0) = \frac{7}{2}, \alpha(1) = \frac{14}{3}, \alpha(2) = \frac{14}{3}$. Poza tymi trzema punktami możemy określać wartości α jak tylko chcemy.

Przykład – 13.23 (c.d. przykładu 13.2). W tamtym przykładzie wyznaczyliśmy nadzieję warunkową $E(Y|X = x)$ dla $0 < x \leq 1$. Mianowicie: $E(Y|X = x) = \frac{x}{2}$. W takim razie nadzieja warunkowa $E(Y|X) = \frac{X}{2}$. Tutaj $\alpha(x) = \frac{x}{2}$ dla $0 < x \leq 1$ oraz 0 dla pozostałych x (zamiast 0 mogło być na przykład 27 i nie ma to znaczenia, gdyż zmienna losowa X nie przyjmuje wartości poza $(0, 1]$, końce odcinka jako zbiory miary zero mogą być uwzględniane lub nie).

Podobnie $E(X|Y = y) = \frac{1-y}{-\ln y}$ dla $0 < y < 1$, więc $E(X|Y) = \frac{1-Y}{-\ln Y}$.

Rozważa się też prawdopodobieństwo warunkowe zdarzenia względem σ -algebry, a więc także względem wektora losowego, jako szczególny przypadek nadziei warunkowej.

Definicja – 13.24. Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną $\mathcal{A} \subset \Sigma$ σ -algebrą, $C \in \Sigma$. Określamy prawdopodobieństwo warunkowe zbioru C względem σ -algebry \mathcal{A} jako:

$$P(C|\mathcal{A}) = E(\mathbb{1}_C|\mathcal{A}).$$

Gdy $X : \Omega \rightarrow \mathbb{R}^k$ jest wektorem można więc mówić o $P(C|X)$ oraz o $P(C|X = x)$:

$$P(C|X) = P(C|\sigma(X)) = E(\mathbb{1}_C|X) \quad \text{oraz} \quad P(C|X = x) = E(\mathbb{1}_C|X = x).$$

Twierdzenie – 13.25 (Własności nadziei warunkowych). Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną, $\mathcal{A} \subset \Sigma$ – σ -algebrą, $Y : \Omega \rightarrow \mathbb{R}$, $E(Y) \in \mathbb{R}$. Wtedy

1. $E(c|\mathcal{A}) = c$, dla $c \in \mathbb{R}$.
2. $E(E(Y|\mathcal{A})) = E(Y)$.
3. $Y \geq 0$ pw. $\implies E(Y|\mathcal{A}) \geq 0$ pw.
4. $E(Y_1 + Y_2|\mathcal{A}) = E(Y_1|\mathcal{A}) + E(Y_2|\mathcal{A})$, o ile prawa strona istnieje.
5. $E(cY|\mathcal{A}) = cE(Y|\mathcal{A})$, dla $c \in \mathbb{R}$.
6. $Y_1 \leq Y_2$ pw. $\implies E(Y_1|\mathcal{A}) \leq E(Y_2|\mathcal{A})$ pw.

7. $\mathcal{B} \subset \mathcal{A}$ – σ -algebra $\implies E(E(Y|\mathcal{A})|\mathcal{B}) = E(Y|\mathcal{B})$.
8. $\mathcal{B} \subset \mathcal{A}$ – σ -algebra $\implies E(E(Y|\mathcal{B})|\mathcal{A}) = E(Y|\mathcal{B})$.
9. $Y_1 \geq 0$ pw., $Y_n \nearrow Y$ pw. $\implies E(Y_n|\mathcal{A}) \nearrow E(Y|\mathcal{A})$ pw.
10. $|Y_n| \leq Z$, $E(Z) \in \mathbb{R}$, $Y_n \xrightarrow{1} Y \implies E(Y_n|\mathcal{A}) \xrightarrow{1} E(Y|\mathcal{A})$.

Dowód. Dowody wszystkich tych własności są standardowe i opierają się na definicji nadziei warunkowej i na klasycznych własnościach całek. Dla przykładu udowodnimy dwie własności.

Własność 2. Korzystając z warunku (C) oraz tego, że $\Omega \in \mathcal{A}$ mamy: $E(E(Y|\mathcal{A})) = \int_{\Omega} E(Y|\mathcal{A}) dP = \int_{\Omega} Y dP = E(Y)$.

Własność 8. Pokażemy, że prawa strona spełnia warunki (C) oraz (M) ze względu na zmienną losową $E(Y|\mathcal{B})$ oraz σ -algebrę \mathcal{A} . Zmienna losowa $E(Y|\mathcal{B})$ jest \mathcal{B} mierzalna, a więc też \mathcal{A} -mierzalna. Niech $A \in \mathcal{A}$. Wtedy $\int_A E(E(Y|\mathcal{B})|\mathcal{A}) dP = \int_A E(Y|\mathcal{B}) dP$, ale to oznacza żądany warunek (C). Prawa strona jest więc równa $E(E(Y|\mathcal{B})|\mathcal{A})$. \square

Powyższe oraz następujące własności można sformułować dla nadziei warunkowych postaci $E(Y|X = x)$.

Twierdzenie – 13.26. Niech Y będzie wektorem losowym, $E(Y) \in \mathbb{R}$.

1. Jeżeli $X : \Omega \rightarrow \mathbb{R}^k$ jest wektorem losowym takim, że X, Y są niezależne, to $E(Y|X) = E(Y)$.
2. Jeżeli Z jest \mathcal{A} -mierzalna oraz $E(ZY) \in \mathbb{R}$, to $E(ZY|\mathcal{A}) = ZE(Y|\mathcal{A})$.
3. Jeżeli $g : \mathbb{R}^k \rightarrow \mathbb{R}$ jest funkcją borelowską, $E(g(X)) \in \mathbb{R}$, to $E(g(X)|X) = g(X)$.
4. Jeżeli X jest zmienną losową, $E(X) \in \mathbb{R}$, to $E(X|X) = X$.

Dowód. Własność 1. Załóżmy najpierw, że $Y = \mathbb{1}_A$, gdzie $A \in \Sigma$. Wtedy $E(\mathbb{1}_A) = P(A)$ jest funkcją stałą i w związku z tym jest mierzalna względem $\sigma(X)$. Wtedy też zachodzi warunek (C): dla $B = X^{-1}(D) \in \sigma(X)$ mamy:

$$\begin{aligned} \int_B Y dP &= \int_A \mathbb{1}_B dP = \int_{A \cap B} dP = P(Y^{-1}(\{1\}) \cap X^{-1}(D)) \\ &= P(Y^{-1}(\{1\})) \cdot P(X^{-1}(D)) = P(A)P(B) = \int_B E(Y) dP. \end{aligned}$$

Zachodzi więc własność 1. dla funkcji charakterystycznych Y . Z liniowości zachodzi dla funkcji prostych Y , a poprzez standardowe przejście graniczne dla dowolnych Y .

Własność 2. Dowodzi się jak poprzednio, zaczynając od przypadku $Z = \mathbb{1}_A$, gdzie $A \in \mathcal{A}$ (ćwiczenie).

Własność 3. Wystarczy wziąć $Z = g(X)$, $Y = 1$ oraz $\mathcal{A} = \sigma(X)$ i skorzystać z własności 2.

Własność 4. Wystarczy we własności 3. wziąć $g(x) = x$. \square

13.4 Rozkład nadziei warunkowej

Nadzieja warunkowa $E(Y|X)$ jest zmienną losową, więc warto się pytać o jej rozkład. Nieraz odpowiedź jest prosta. Jak pamiętamy $E(Y|X) = \alpha(X)$, $\alpha(x) = E(Y|X = x)$. Jeżeli więc znamy rozkład X oraz α możemy na tej podstawie starać się wyznaczyć rozkład $E(Y|X)$. Na przykład widać, że w przypadku, gdy X ma rozkład dyskretny wyznaczony przez ciągi $\{x_i\}, \{p_i\}$ to $E(Y|X)$ ma rozkład dyskretny wyznaczony przez ciągi $\{x_i\}, \{p_i\}$. Natomiast, gdy X ma rozkład ciągły, sytuacja jest bardziej skomplikowana.

Przykład – 13.27. Niech X, Y oznaczają liczby oczek uzyskane w rzucie parą kostek. Niech $\text{Min} = \min(X, Y)$. Wskazać rozkład $E(X|\text{Min})$.

Można wyznaczyć rozkład wektora losowego (X, Min) , następnie dla każdego $m = 1, \dots, 6$ wyznaczyć $P(\text{Min} = m)$ oraz rozkłady warunkowe $P_{X|\text{Min}=m}$ i na tej podstawie $E(X|\text{Min} = m)$. Zmienna losowa $E(X|\text{Min})$ ma rozkład skupiony w punktach $E(X|\text{Min} = m)$ z prawdopodobieństwami $P(\text{Min} = m)$. Są to więc ciągi $\frac{26}{11}, \frac{28}{9}, \frac{27}{7}, \frac{23}{5}, \frac{16}{3}, 6$ oraz $\frac{11}{36}, \frac{9}{36}, \frac{7}{36}, \frac{5}{36}, \frac{3}{36}, \frac{1}{36}$.

Przykład – 13.28. Zmienna losowa X ma rozkład $U(-\pi, \pi)$. Wskażemy rozkłady $E(\cos X|X)$ oraz $E(X|\cos X)$.
Cosinus jest funkcją borelowską, więc $\cos X$ jest \mathcal{A} -mierzalna, więc

$$E(\cos X|X) = \cos X E(1|X) = \cos X.$$

Ta zmienna losowa ma rozkład ciągły, patrz przykład 5.30.

Jeżeli $\cos X = y \in (-1, 1)$, to X przyjmuje dwie wartości, których średnią jest 0. Wydaje się więc, że $E(X|\cos X)$ może być równe 0. Sprawdzamy więc czy funkcja stałe równa 0 spełnia warunki (M) oraz (C) spełniane przez $E(X|\cos X)$. (M) jest oczywisty. Weźmy teraz dowolny zbiór borelowski $A \in \sigma(\cos \circ X)$. A jest więc postaci $X^{-1}(\cos^{-1}(B))$, gdzie B jest borelowski. Ze względu na parzystość cosinusa $\cos^{-1}(B)$ jest symetryczny względem 0 i wtedy $\int_A E(X|\cos X) dP = \int_A X dP = \int_{\cos^{-1}(B)} \frac{1}{2\pi} \mathbb{1}_{(-\pi, \pi)}(x) dx = 0 = \int_A 0 dP$. Jest więc spełniony warunek (C). Ostatecznie więc $E(X|\cos X) = 0$ ma rozkład jednopunktowy.

13.5 Pytania

Pytanie 13.1. Znaleźć rozkład warunkowy maksimum liczby oczek w rzucie parą symetrycznych kostek pod warunkiem, że na drugiej kostce wypadła liczba parzysta. Wyznacz jego nadzieję matematyczną. ROZWIĄZANIE

Pytanie 13.2. Wektor losowy (X, Y) ma rozkład jednostajny na trójkącie o wierzchołkach $(-1, 0)$, $(1, 0)$, $(0, 1)$. Wyznaczyć $E(X|Y = 1/2)$, $E(Y|X = 1/2)$. ROZWIĄZANIE

Pytanie 13.3. Dana jest przestrzeń probabilistyczna (Ω, Σ, P) oraz zmienna losowa $X : \Omega \rightarrow \mathbb{R}$. Czy i jakie zawierania zachodzą pomiędzy $\sigma(X)$, $\sigma(X^2)$ oraz $\sigma(X^3)$: wypowiedzieć i udowodnić twierdzenie, podać odpowiedni przykład jeżeli zawieranie nie zachodzi. ROZWIĄZANIE

Pytanie 13.4. Przeprowadzić dowód własności 7. w twierdzeniu 13.25. ROZWIĄZANIE

Pytanie 13.5. Przeprowadzić dowód własności 2. w twierdzeniu 13.26. ROZWIĄZANIE

Pytanie 13.6. Zakładamy, że $E(Y)$ jest skończona. Czy następujące zdania są prawdziwe? Odpowiedź uzasadnić.

- (1) Jeżeli wektor losowy X ma rozkład dyskretny, to zmienna losowa $E(Y|X)$ ma rozkład dyskretny.
- (2) Jeżeli wektor losowy X ma rozkład ciągły, to zmienna losowa $E(Y|X)$ ma rozkład ciągły.

ROZWIĄZANIE

Rozdział 14

Warunkowania

Wyznaczanie oczekiwanych wartości warunkowych może być w wielu przypadkach żmudne. Warto jednak pamiętać, że fakt, iż przy bardzo ogólnych założeniach one istnieją oraz spełniają pewne własności może być bardzo pomocny. Z drugiej strony, w pewnych zagadnieniach warunkowe wartości oczekiwane są w sposób oczywisty znane i wtedy mogą służyć do wyznaczania „zwykłej” nadziei oraz prawdopodobieństw zdarzeń (pamiętamy, że $P(C) = E(\mathbb{1}_C)$). Procedura ta jest uogólnieniem poznanej już wcześniej procedury wykorzystującej wzór na prawdopodobieństwo całkowite.

Wiedząc, że $E(Y) = E(E(Y|X))$, zauważmy, iż na podstawie twierdzenia 13.19 oraz twierdzeń 6.8 i 6.11 otrzymujemy:

Twierdzenie – 14.1. Niech $X : \Omega \rightarrow \mathbb{R}^k$ będzie wektorem losowym, $Y : \Omega \rightarrow \mathbb{R}$ zmienną losową, $E(Y) \in \mathbb{R}$. Niech $C \in \Sigma$. Wtedy:

$$E(Y) = \int_{\mathbb{R}^n} E(Y|X = x) dP_X(x), \quad P(C) = \int_{\mathbb{R}^n} P(C|X = x) dP_X(x).$$

Jeżeli X ma rozkład dyskretny zadany przez ciągi $x_1, x_2, \dots \in \mathbb{R}^n$, p_1, p_2, \dots , to

$$E(Y) = \sum_i E(Y|X = x_i)p_i, \quad P(C) = \sum_i P(C|X = x_i)p_i.$$

Jeżeli X ma rozkład ciągły o gęstości f , to

$$E(Y) = \int_{\mathbb{R}^n} E(Y|X = x)f(x) dx, \quad P(C) = \int_{\mathbb{R}^n} P(C|X = x)f(x) dx.$$

14.1 Przykłady

Przykład – 14.2. Rzucono dwa razy monetą symetryczną, a następnie tyle razy rzucono kostką symetryczną, ile wypadło orłów. Jaka jest wartość oczekiwana sumy uzyskanych oczek?

Niech X oznacza liczbę wyrzuconych orłów, a Y sumę oczek. Oczywiście zadanie można rozwiązać bezpośrednio wyznaczając rozkład zmiennej losowej Y . Pokażemy inną metodę. Mamy:

$$E(Y) = E(E(Y|X)) = E(Y|X = 0)P(X = 0) + E(Y|X = 1)P(X = 1) + E(Y|X = 2)P(X = 2)$$

Powyższe trzy nadzieje warunkowe są faktycznie znane, gdyż sprowadzają się do obliczenia 'zwykłych' nadziei. Mamy więc:

$$E(Y) = 0 \cdot \frac{1}{4} + \frac{7}{2} \cdot \frac{1}{2} + 2 \cdot \frac{7}{2} \cdot \frac{1}{4} = \frac{7}{2}.$$

Takiego wyniku można się było spodziewać ze względu na symetrię: to co tracimy w przypadku braku orła, zyskujemy, gdy wypadną dwa orły.

Przykład – 14.3. Z odcinka $[0, 1]$ losujemy liczbę X według rozkładu jednostajnego, a następnie z odcinka $[X, 1]$ liczbę Y według rozkładu jednostajnego.

(a) Znaleźć nadzieję matematyczną $E(Y)$. (b) Wyznaczyć $P(Y < \frac{1}{2})$.

Zadanie można rozwiązać w sposób klasyczny, podobnie jak w przykładach 5.21 oraz 13.2, wyznaczając najpierw rozkład wektora (X, Y) , potem rozkład Y , a w końcu $E(Y)$ i $P(Y < \frac{1}{2})$. Można jednak rozumować inaczej, opierając się na twierdzeniu 14.1.

Ad(a). Wiemy, że $E(Y|X = x) = \frac{x+1}{2}$ dla $0 < x < 1$, więc:

$$E(Y) = \int_{\mathbb{R}} E(Y|X = x)f_X(x) dx = \int_0^1 \frac{x+1}{2} dx = \frac{3}{4}.$$

Jeszcze szybciej można zrobić tak:

$$E(Y) = E(E(Y|X)) = E\left(\frac{X+1}{2}\right) = \frac{E(X)+1}{2} = \frac{\frac{1}{2}+1}{2} = \frac{3}{4}.$$

Ad(b). Widać, że $P(Y < \frac{1}{2}|X = x) = \frac{\frac{1}{2}-x}{1-x}$ dla $0 < x \leq \frac{1}{2}$ oraz 0 dla innych x .

$$\text{Więc } P(Y < \frac{1}{2}) = \int_{\mathbb{R}} P(Y < \frac{1}{2}|X = x)\mathbb{1}_{[0,1]}(x) dx = \int_0^{\frac{1}{2}} \frac{\frac{1}{2}-x}{1-x} dx = \frac{1}{2} - \frac{\ln 2}{2} \cong 0.153.$$

Przykład – 14.4. Losujemy liczbę p zgodnie z rozkładem $U(0, 1)$, a następnie wykonujemy n doświadczeń Bernoulliego z prawdopodobieństwem sukcesów w każdym doświadczeniu równym p . Ilu możemy oczekiwać sukcesów?

Niech X będzie zmienną losową o rozkładzie $U(0, 1)$. Niech Y_1, \dots, Y_n będą i.i.d o rozkładzie $(0, 1, X)$ każda. $S_n = Y_1 + \dots + Y_n$. Wtedy $E(S_n|X = p) = np$. Czyli $E(S_n|X) = nX$. A więc $E(S_n) = E(E(S_n|X)) = E(nX) = nE(X) = n \frac{1}{2} = \frac{n}{2}$.

Przykład – 14.5. W przykładzie 11.28 analizowaliśmy następujące zadanie, stosując aparat funkcji tworzących.

Owad (powiedzmy mucha) składa dużo jajeczek z których mogą wykluwać się nowe owady. Zakładając, że liczba jajeczek ma rozkład Poissona o parametrze λ , oraz, że owady wykluwają się z jajeczek niezależnie od siebie z tym samym prawdopodobieństwem p , wyznaczyć rozkład oraz oczekiwaną wartość liczby potomków jednego owada.

Niech N oznacza liczbę jajeczek, natomiast S liczbę potomków jednego owada. Chcemy wyznaczyć $P(S = k)$ dla $k = 0, 1, 2, 3, \dots$. Z treści zadania wynika, że $P(S = k | N = n) = 0$ dla $n < k$ oraz $\binom{n}{k} p^k (1-p)^{n-k}$ dla $n \geq k$. Mamy kolejno:

$$\begin{aligned} P(S = k) &= \sum_{n=0}^{\infty} P(S = k | N = n)P(N = n) \\ &= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \frac{1}{k!} p^k (1-p)^{-k} \sum_{n=k}^{\infty} \frac{(\lambda(1-p))^n}{(n-k)!} \\ &= e^{-\lambda} \frac{1}{k!} p^k \lambda^k \sum_{j=0}^{\infty} \frac{(\lambda(1-p))^j}{j!} = e^{-\lambda} \frac{1}{k!} (p\lambda)^k e^{\lambda(1-p)} = e^{-p\lambda} \frac{(p\lambda)^k}{k!}. \end{aligned}$$

Więc S ma rozkład Poissona $P_{p\lambda}$. W takim razie $E(S) = p\lambda$. Ten ostatni wynik można było otrzymać także bezpośrednio. Mianowicie zauważmy, że $E(S | N = n) = np$, czyli $E(S | N) = Xp$. Stąd:

$$E(S) = E(E(S | Np)) = E(Np) = pE(N) = p\lambda.$$

Przykład – 14.6. Kaja i Leon losują na chybił-trafił i niezależnie od siebie po jednej liczbie od 0 do 100. Jeżeli liczby różnią się nie więcej niż o 10, uważają, że warto pójść razem do kina. Jakie jest prawdopodobieństwo, że Kaja i Leon pójdą razem do kina?

Zadanie to rozwiązaliśmy już stosując prawdopodobieństwo geometryczne, przykład 2.8. Pokażemy teraz inny sposób rozwiązania.

Niech KL oznacza zdarzenie: że Kaja i Leon pójdą razem do kina. Niech L oznacza liczbę wylosowaną przez Leona. Łatwo wyznaczyć prawdopodobieństwo warunkowe $P(KL | L = x)$. Mianowicie, gdy Leon wylosował liczbę x , to aby zaszło zdarzenie KL , liczba wylosowana przez Kaję musi się różnić od x o mniej niż o 10. Licząc prawdopodobieństwo geometryczne na odcinku $(0, 100)$, otrzymujemy:

$$P(KL | L = x) = \begin{cases} \frac{x+10}{100} & \text{dla } 0 < x < 10, \\ \frac{20}{100} & \text{dla } 10 < x < 90, \\ \frac{100-(x-10)}{100} & \text{dla } 90 < x < 100. \end{cases}$$

Ponieważ zmienna losowa L ma rozkład $U(0, 100)$ otrzymujemy:

$$P(KL) = \int_0^{100} P(KL | L = x) \frac{1}{100} dx = \frac{19}{100}.$$

Przykład – 14.7. Niech X_1, \dots, X_n będzie próbką prostą z rozkładu $B(1, p)$. Niech $S = X_1 + \dots + X_n$. Wyznaczyć $E(X_1|S)$.

Zauważmy, że dla $i = 2, \dots, n$, $E(X_i|S_n) = E(X_1|S_n)$. Korzystając z własności nadziei warunkowych kolejno mamy: $S_n = E(S_n|S_n) = E(X_1 + \dots + X_n|S_n) = E(X_1|S_n) + \dots + E(X_n|S_n) = nE(X_1|S_n)$. Stąd: $E(X_1|S_n) = \frac{S_n}{n}$.

Przykład – 14.8. Punkt (x, y) wylosowano z kwadratu $[0, 1]^2$ zgodnie z rozkładem jednostajnym. Jaka jest oczekiwana wartość x , jeżeli wiadomo, że: (a) $x + y \leq \frac{1}{2}$, (b) $x + y = \frac{1}{2}$?

Ad (a). Sposób 1. Wyznaczamy rozkład zmiennej losowej X , gdy wektor losowy (X, Y) ma rozkład $U(W)$ (jednostajny) na zbiorze W określonym warunkami $x > 0$, $y > 0$, $x + y \leq \frac{1}{2}$ i wyznaczamy $E(X)$. Mamy kolejno $f_{(X,Y)} = 8\mathbb{1}_W$, $f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy = \int_0^{\frac{1}{2}-x} 8 dx = 4 - 8x$ dla $0 < x < \frac{1}{2}$. $E(X) = \int_{\mathbb{R}} x f_X(x) dx = \int_0^{\frac{1}{2}} x(4 - 8x) dx = \frac{1}{6}$.

Sposób 2. Korzystamy z określenia $E(X|W)$: $E(X|W) = \frac{\int_W X dP}{P(W)}$. Tutaj P jest miarą Lebesgue'a na całym kwadracie. W związku z tym $P(W) = \frac{1}{8}$. Natomiast $\int_W X dP = \iint_W x d(x, y) = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}-x} x dy dx = \int_0^{\frac{1}{2}} x(\frac{1}{2}-x) dx = \frac{1}{48}$. $E(X|W) = \frac{1}{6}$.

Ad (b). Sposób 1. Niech X, Y oznaczają odpowiednie zmienne losowe. Ze względu na symetrię mamy

$$E(X|X+Y) = E(Y|X+Y).$$

Z jednej strony $E(X+Y|X+Y) = X+Y$ (własność 4. w twierdzeniu 13.26). Z drugiej strony $E(X+Y|X+Y) = E(X|X+Y) + E(Y|X+Y) = 2E(X|X+Y)$. Więc $E(X|X+Y) = \frac{1}{2}(X+Y)$. $E(X|X+Y = \frac{1}{2}) = \frac{1}{4}$.

Sposób 2. Wyznaczamy gęstość wektora losowego $(X, X+Y)$ i następnie obliczamy wartość oczekiwaną odpowiedniego rozkładu warunkowego. Ponieważ $f_X = \mathbb{1}_{(0,1)}$ oraz dla każdego $x \in (0, 1)$: $f_{X+Y|X=x} = \mathbb{1}_{(x, x+1)}$, to ich iloczyn: $f_{(X, X+Y)}(x, s) = 1$ dla $0 < x < 1$, $x < s < x+1$ oraz 0 w pozostałych przypadkach. Mamy więc:

$$E(X|X+Y = s) = \frac{\int_{\mathbb{R}} x f_{(X, X+Y)|X}(x, s) dx}{\int_{\mathbb{R}} f_{(X, X+Y)|X}(x, s) dx} \Big|_{s=\frac{1}{2}} = \frac{\int_0^{\frac{1}{2}} x dx}{\int_0^{\frac{1}{2}} dx} = \frac{1}{4}.$$

Przykład – 14.9. Z przedziału $(0, 1)$ losujemy punkt x oraz rzucamy niesymetryczną monetą, gdzie prawdopodobieństwo wypadnięcia orła wynosi p . Jeżeli wypadła reszka, to punkt y losujemy z przedziału $(0, x)$ według rozkładu jednostajnego, gdy wypadnie orzeł punkt y losujemy według rozkładu jednostajnego z przedziału $(x, 1)$. Znaleźć średnie położenie punktu y .

Niech X będzie zmienną losową odpowiadającą pierwszemu losowaniu, Z zmienną losową odpowiadającą rzutowi monetą, a Y pozycją drugiego wylosowanego punktu. Aby policzyć $E(Y)$ zauważmy, że

$$E(Y) = \int_{\mathbb{R} \times \{0,1\}} E(Y|X = x, Z = z) dQ(x, z),$$

gdzie Q jest rozkładem wektora losowego (X, Z) . Z treści zadania wynika, że X, Z są niezależne, a więc Q jest iloczynem kartezjańskim ich rozkładów: $Q = U(0, 1) \times B(1, p)$. Zauważmy, że:

$$E(Y|X = x, Z = z) = \begin{cases} \frac{x}{2} & \text{dla } 0 < x < 1, z = 0 \\ \frac{x+1}{2} & \text{dla } 0 < x < 1, z = 1 \end{cases}$$

Stosujemy twierdzenie Fubiniego:

$$\begin{aligned} E(Y) &= \int_{\mathbb{R}} \left(\int_{\{0,1\}} E(Y|X = x, Z = z) dB(1, p)(z) \right) dU(0, 1)(x) \\ &= \int_0^1 \frac{x}{2}(1-p) + \frac{x+1}{2}p dx = \int_0^1 \frac{x+p}{2} dx = \frac{1}{4} + \frac{p}{2}. \end{aligned}$$

14.2 Nierówność Jensena i obniżanie wariancji

Niech $\Delta \subset \mathbb{R}$ będzie przedziałem. Funkcję $g : \Delta \rightarrow \mathbb{R}$ nazywamy *wypukłą*, jeżeli dla każdych $x, y \in \Delta$ oraz $0 \leq p, q \leq 1, p + q = 1$ zachodzi:

$$g(px + qy) \leq pg(x) + qg(y).$$

Stosując indukcję, łatwo widać (ćwiczenie), że warunek ten jest równoważny warunkowi:

$$g\left(\sum_{i=1}^n x_i p_i\right) \leq \sum_{i=1}^n g(x_i) p_i,$$

dla dowolnych skończonych ciągów $\{x_i\}, \{p_i\}$, takich, że $p_i \geq 0, \sum_{i=1}^n p_i = 1$. W języku probabilistyki mamy więc natychmiast:

Twierdzenie. Dla dowolnej zmiennej losowej o dyskretnym, skończonym rozkładzie prawdopodobieństwa takiej, że $P(X \in \Delta) = 1$ i dowolnej funkcji wypukłej $g : \Delta \rightarrow \mathbb{R}$ zachodzi:

$$g(E(X)) \leq E(g(X)).$$

Rezultat ten można znacznie wzmocnić.

Twierdzenie – 14.10 (Nierówność Jensena). Dla dowolnej zmiennej losowej o wartościach w otwartym przedziale $\Delta, E(X) \in \mathbb{R}$, dowolnej σ -algebry $\mathcal{A} \subset \Sigma$ i dowolnej funkcji wypukłej $g : \Delta \rightarrow \mathbb{R}$:

$$g(E(X|\mathcal{A})) \leq E(g(X)|\mathcal{A}).$$

W szczególności, przy powyższych założeniach:

$$g(E(X)) \leq E(g(X)).$$

Dowód oparty jest na następującej własności funkcji wypukłej.

Dla dowolnej funkcji wypukłej $g : \Delta \rightarrow \mathbb{R}$ określonej na przedziale otwartym istnieją ciągi liczb a_n oraz b_n , takie, że dla każdego $y \in \Delta$:

$$g(y) = \sup_n \{a_n y + b_n\}.$$

Dowód nierówności Jensena.

Łatwo pokazać (ćwiczenie), $E(Y|\mathcal{A}) \in \Delta$ pw.

Ponieważ $g(y) = \sup_n \{a_n y + b_n\}$, to dla każdego n

$$g(Y) \geq a_n Y + b_n.$$

Z własności nadziei warunkowej (sformułuj tę własność), dla każdego n zachodzi:

$$E(g(Y)|\mathcal{A}) \geq E(a_n Y + b_n|\mathcal{A}) = a_n E(Y|\mathcal{A}) + b_n.$$

Więc

$$E(g(Y)|\mathcal{A}) \geq \sup_n \{a_n E(Y|\mathcal{A}) + b_n\} = g(E(Y|\mathcal{A})). \quad \square$$

W statystyce ważną rolę odgrywa następująca nierówność:

Twierdzenie – 14.11. Dla dowolnej zmiennej losowej o skończonej nadziei matematycznej i skończonej wariancji oraz dowolnej σ -algebry $\mathcal{A} \subset \Sigma$ zachodzi nierówność

$$D^2(E(Y|\mathcal{A})) \leq D^2(Y).$$

Dowód. Biorąc funkcję wypukłą $g : \mathbb{R} \ni x \rightarrow x^2 \in \mathbb{R}$ i stosując nierówność Jensena, dostajemy:

$$E(Y|\mathcal{A})^2 \leq E(Y^2|\mathcal{A}).$$

Teraz:

$$\begin{aligned} D^2(E(Y|\mathcal{A})) &= E(E(Y|\mathcal{A})^2) - E(E(Y|\mathcal{A}))^2 \\ &\leq E(E(Y^2|\mathcal{A})) - E(Y)^2 = E(Y^2) - E(Y)^2 = D^2(Y). \end{aligned} \quad \square$$

Przykład – 14.12.

1. $D^2(E(Y|\{\emptyset, \Sigma\})) = D^2(E(Y)) = 0 \leq D^2(Y)$,
2. Jeżeli X, Y są niezależne, to $E(Y|X) = E(Y) \in \mathbb{R}$, więc $D^2(E(Y|X)) = 0 \leq D^2(Y)$.
3. $D^2(E(Y|Y)) = D^2(Y)$.
4. Ogólniej: $D^2(E(f(Y)|Y)) = D^2(f(Y))$ dla borelowskiej funkcji f .

Metody Monte Carlo i obniżanie wariancji

Jak pamiętamy metody Monte Carlo mogą być używane do obliczania wielkości, które dają się interpretować jako wartości oczekiwane pewnych zmiennych losowych. Załóżmy, że chcemy obliczyć wielkość $m = E(Y)$ i używamy do tego estymatora \hat{m} . Czasem jednak warto i można wyrazić m jako nadzieję matematyczną nadziei warunkowej $E(Y|\mathcal{A})$: $m = E(E(Y|\mathcal{A}))$ i wtedy m estymujemy estymatorem \hat{m} . Wariancja $E(Y|\mathcal{A})$ może być istotnie mniejsza niż wariancja Y i okazuje się, że w pewnych przypadkach wariancja estymatora \hat{m} też jest mniejsza od wariancji estymatora \hat{m} . Ponieważ interesuje nas przedział ufności dla m musimy jeszcze brać pod uwagę koszt (np. liczbę losowań) rozważanych metod wymagany do otrzymania przedziału ufności dla m .

Jako przykład naszkicujemy najprostszy wariant tak zwanej metody warstwowej obliczania całek.

Zauważmy najpierw, że metoda II poznana w poprzednim rozdziale liczenia całek, patrz punkt 12.3, może być łatwo zaadaptowana do obliczania całek wielokrotnych: Dany jest zbiór borelowski C w \mathbb{R}^n o mierze skończonej i dodatniej, $\mu_L(C) > 0$, oraz funkcja $f : C \rightarrow \mathbb{R}$ sumowalna (to znaczy $J = \int_C f dx \in \mathbb{R}$).

Ponieważ $J = \mu_L(C)E(f(X))$, gdzie $X \sim U(C)$, dobrym estymatorem jest:

$$\hat{J} = \mu_L(C) \frac{1}{N} \sum_{i=1}^N f(X_i), \quad \text{gdzie } X_1, \dots, X_N \text{ i.i.d., } X_i \sim U(C).$$

Opiszemy **Metodę III**.

Niech C będzie sumą skończoną zbiorów rozłącznych C_1, \dots, C_k .

Niech $\mathcal{A} = \sigma(A_1, \dots, A_k)$, gdzie $A_i = X^{-1}(C_i)$. Oczywiście $P(A_i) = P_X(C_i) = \frac{\mu_L(C_i)}{\mu_L(C)}$. Wtedy:

$$E(f(X)) = E(E(f(X)|\mathcal{A})), \text{ ale } D^2(E(f(X)|\mathcal{A})) \leq D^2(f(X)).$$

Wskażemy estymator nadziei matematycznej zmiennej losowej $W = E(f(X)|\mathcal{A})$, który pozwala wyznaczyć przedział ufności dla J . Zachodzi:

$$E(f(X)|\mathcal{A})(\omega) = \frac{\int_{A_i} f(X) dP}{P(A_i)} = \frac{\int_{C_i} f(x) \frac{1}{\mu_L(C)} \mathbb{1}_{C_i}(x) dx}{P(A_i)} = \frac{\int_{C_i} f(x) dx}{\mu_L(C_i)} \quad \text{dla } \omega \in A_i, \quad i = 1, \dots, k.$$

Zauważmy, że: $\int_{C_i} f(x) dx = \mu_L(C_i)E(f(X_i))$, gdzie $X_i \sim U(C_i)$.

Tak więc

$$W = \sum_{i=1}^k E(f(X_i)) \mathbb{1}_{A_i}.$$

Ponieważ $E(W) = \sum_{i=1}^k E(f(X_i))P(A_i)$ więc wskażemy estymatory $\hat{m}_1, \dots, \hat{m}_k$ wielkości $E(f(X_i))$, a jako estymator $E(W)$ weźmiemy $\hat{m} = \sum_{i=1}^k \hat{m}_i P(A_i)$.

Oczywiście weźmiemy średnie arytmetyczne: $\hat{m}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} f(X_{ij})$, gdzie X_{i1}, \dots, X_{in} jest próbką prostą z rozkładu $U(C_i)$. Więc

$$\hat{m} = \sum_{i=1}^k \frac{1}{n_i} \sum_{j=1}^{n_i} f(X_{ij}) P(A_i).$$

Estymatorem całki J jest:

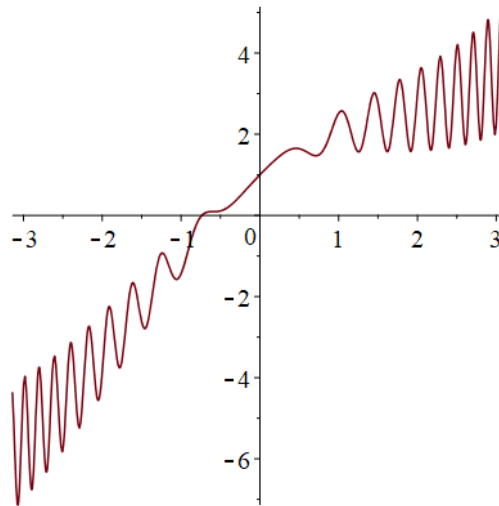
$$\hat{J} = \mu_L(C) \hat{E} = \sum_{i=1}^k \mu_L(C_i) \frac{1}{n_i} \sum_{j=1}^{n_i} f(X_{ij}).$$

Można na wiele sposobów dobierać podział zbioru C na zbiory C_i oraz wielkości n_i . Na przykład założymy, że $\mu_L(C_i) = \frac{\mu_L(C)}{k}$ oraz, że $n_1 = \dots = n_k = n$. Wtedy

$$\hat{J} = \mu_L(C) \frac{1}{kn} \sum_{i=1}^k \sum_{j=1}^n f(X_{ij})$$

Można pokazać, że przy tej samej liczbie losowań (czyli, gdy $N = kn$) przedział ufności dla J wyznaczony za pomocą \hat{J} jest węższy od analogicznego przedziału ufności wyznaczonego za pomocą \hat{J} .

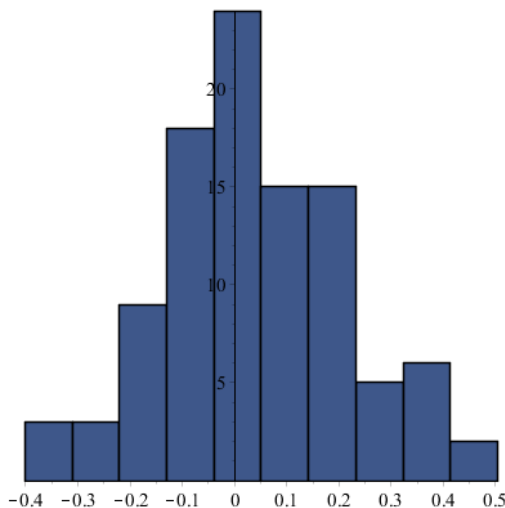
Przykład – 14.13. Porównamy metodę II i metodę III do obliczenia całki $J = \int_{-\pi}^{\pi} x + \cos x + x(\cos(3x^2))^2 dx$, o której wiadomo, że jest równa 0.



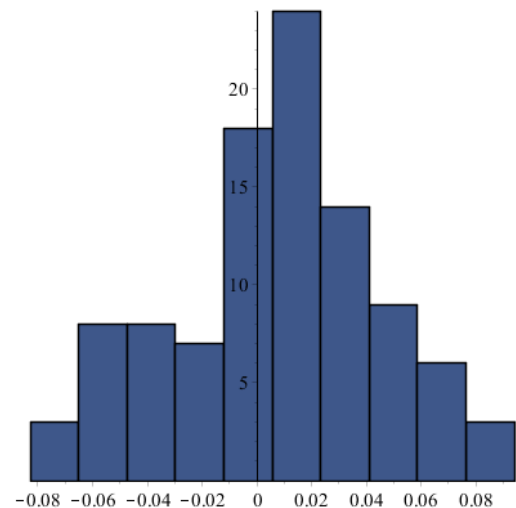
Powtarzamy 100 razy metodę II, gdzie $N = 9000$, $\alpha = 0.01$. Otrzymujemy średni błąd estymatora $\hat{J} = 0.14000$ i wartość średnią promienia przedziału ufności dla $J = \mathbf{0.49217}$.

Powtarzamy 100 razy metodę III (warstwową), gdzie przedział został podzielony na 30 jednakowych przedziałów ($k = 30$), $n = 300$, $\alpha = 0.01$. Otrzymujemy średni błąd estymatora $\hat{J} = 0.03126$ i wartość średnią promienia przedziału ufności dla $J = \mathbf{0.09484}$.

W poniższych histogramach stu wartości \hat{J} oraz \hat{J} warto zwrócić uwagę na skalę osi poziomej.



Metoda II



Metoda III (warstwową)

Istnieje wiele innych sposobów obniżania wariancji w metodach Monte Carlo, patrz [17], [18].

14.3 Pytania

Pytanie 14.1. Wektor losowy (X, Y) ma rozkład o gęstości $c x \mathbb{1}_{[0,1]^2}$, gdzie $c \in \mathbb{R}$. Wyznaczyć kolejno: $E(X|Y)$, $E(Y|X)$, $D^2(X)$, $D^2(E(X|Y))$, $D^2(Y)$, $D^2(E(Y|X))$. (ROZWIĄZANIE)

Pytanie 14.2. Zmienna losowa X ma rozkład $U(0, 3)$. Niech $A_i = X^{-1}((i-1, i])$ dla $i = 1, 2, 3$.

Obliczyć $D^2(E(X|\sigma(A_1, A_2, A_3)))$. ROZWIĄZANIE

Pytanie 14.3. Uzupełnić dowód twierdzenia 14.10.

(a) Dlaczego $E(Y|\mathcal{A}) \in \Delta$ pw.? (b) Dlaczego zachodzi nierówność (26)? ROZWIĄZANIE

Pytanie 14.4. Wektor losowy (X, Y) ma rozkład jednostajny na półkolu $x^2 + y^2 \leq 1, y \geq 0$. Znaleźć: $D^2(Y)$,

$D^2(E(Y|X)), D^2(X), D^2(E(X|Y))$. ROZWIĄZANIE

Pytanie 14.5. Wyznaczyć $D^2(Y)$ oraz $D^2(E(Y|(X, Z)))$ dla zmiennych losowych określonych w przykładzie 14.9.

ROZWIĄZANIE

Pytanie 14.6. Przy standardowych oznaczeniach i założeniu, że $E(X) \in \mathbb{R}$, porównać

$$E(X|\mathcal{A})^+, E(X^+|\mathcal{A}) \quad \text{oraz podobnie} \quad E(X|\mathcal{A})^-, E(X^-|\mathcal{A}).$$

ROZWIĄZANIE

Rozdział 15

Martynały

Wykorzystując pojęcie nadziei warunkowej, można badać procesy stochastyczne, czyli ciągi zmiennych losowych, zwane martynałami, podmartynałami (submartynałami) i nadmartynałami (supmartynałami). Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną.

15.1 Definicje i przykłady

Niech $\{\mathcal{A}_n\}$ będzie ciągiem σ -algebr: $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \mathcal{A}_3 \subset \dots \subset \Sigma$. Często nazwa – *filtracja*.

Niech $X_n : \Omega \rightarrow \mathbb{R}$ będzie ciągiem zmiennych losowych, takich, że $E(X_n) \in \mathbb{R}$ oraz dla każdego n X_n jest \mathcal{A}_n -mierzalna.

Definicja – 15.1. Parę $(\{X_n\}, \{\mathcal{A}_n\})$ nazywamy:

1. *martynałem* \iff dla każdego n $E(X_{n+1}|\mathcal{A}_n) = X_n$.
2. *submartynałem (podmartynałem)* \iff dla każdego n $E(X_{n+1}|\mathcal{A}_n) \geq X_n$.
3. *supmartynałem (nadmartynałem)* \iff dla każdego n $E(X_{n+1}|\mathcal{A}_n) \leq X_n$.

Definicja – 15.2. Ciąg $\{X_n\}$ nazywamy odpowiednio *martynałem, podmartynałem, nadmartynałem* \iff para $(\{X_n\}, \{\mathcal{A}_n\})$, gdzie $\mathcal{A}_n = \sigma(X_1, \dots, X_n)$ jest odpowiednio martynałem, podmartynałem, nadmartynałem.

Często używana interpretacja.

X_n – kapitał gracza po n grach.

\mathcal{A}_n – dostępna informacja po n grach.

$\mathcal{A}_n \subset \mathcal{A}_{n+1}$ – informacja wzrasta w trakcie gry.

$E(X_{n+1}|\mathcal{A}_n) = X_n$ – gra jest sprawiedliwa.

$E(X_{n+1}|\mathcal{A}_n) \geq X_n$ – gra jest korzystna dla gracza.

$E(X_{n+1}|\mathcal{A}_n) \leq X_n$ – gra nie jest korzystna dla gracza.

Uwaga – 15.3. Jeżeli para $(\{X_n\}, \{\mathcal{A}_n\})$ jest odpowiednio *martynałem, podmartynałem, nadmartynałem*, to ciąg $\{X_n\}$ jest odpowiednio *martynałem, podmartynałem, nadmartynałem*.

Dowód. W przypadku martynału dla każdego n zachodzi równość:

$$E(X_{n+1}|\mathcal{A}_n) = X_n,$$

a ponieważ dla $i \leq n$: X_i jest \mathcal{A}_i -mierzalna, więc jest też \mathcal{A}_n -mierzalna i stąd wektor (X_1, \dots, X_n) jest \mathcal{A}_n -mierzalny. Tak więc $\sigma(X_1, \dots, X_n) \subset \mathcal{A}_n$ i korzystając z własności 7. w twierdzeniu 13.25 oraz własności 4. w twierdzeniu 13.26 z powyższej równości otrzymujemy kolejno:

$$E(E(X_{n+1}|\mathcal{A}_n)|\sigma(X_1, \dots, X_n)) = E(X_n|\sigma(X_1, \dots, X_n)),$$

$$E(X_{n+1}|\sigma(X_1, \dots, X_n)) = X_n.$$

Pozostałe przypadki są oczywiste. □

Przykład – 15.4. X_1, X_2, X_3, \dots – niezależne zmienne losowe, $E(X_i) = 0$,

$$S_n = X_1 + X_2 + \dots + X_n.$$

Wtedy $(\{S_n\}, \{\sigma(X_1, \dots, X_n)\})$ jest martyngałem.

Rzeczywiście: Dla każdego n : S_n jest $\sigma(X_1, \dots, X_n)$ mierzalna (ćwiczenie).

$$\begin{aligned} E(S_{n+1} | \sigma(X_1, \dots, X_n)) &= E(S_n + X_{n+1} | \sigma(X_1, \dots, X_n)) \\ &= E(S_n | \sigma(X_1, \dots, X_n)) + E(X_{n+1} | \sigma(X_1, \dots, X_n)) = S_n + E(X_{n+1}) = S_n. \end{aligned}$$

Interpretacja. Jeżeli oczekiwany zysk w pojedynczej grze równa się zeru, a gry od siebie są niezależne, to gra jest sprawiedliwa.

Przykład – 15.5. W przykładzie 10.23 rozpatrywaliśmy ciąg: $x_0 = 1$, x_{n+1} – liczba wylosowana zgodnie z rozkładem $U(0, 2x_n)$, $n = 0, 1, 2, 3, \dots$

Niech X_n oznacza zmienną losową, której realizacją jest x_n .

Ciąg X_n jest martyngałem, gdyż X_n jest $\sigma(X_1, \dots, X_n)$ mierzalne, a także z określenia x_{n+1} widać, że:

$$E(X_{n+1} | \sigma(X_1, \dots, X_n)) = \frac{0 + 2X_n}{2} = X_n.$$

Przykład – 15.6. X_1, X_2, X_3, \dots i.i.d. o dyskretnym rozkładzie jednostajnym skupionym na zbiorze $\{-1, 1\}$. $S_n = X_1 + \dots + X_n$.

Wtedy $(\{S_n^2 - n\}, \{\sigma(X_1, \dots, X_n)\})$ jest martyngałem.

Rzeczywiście: Dla każdego n : $S_n^2 - n$ jest $\sigma(X_1, \dots, X_n)$ mierzalna (ćwiczenie).

$$\begin{aligned} &E(S_{n+1}^2 - (n+1) | \sigma(X_1, \dots, X_n)) \\ &= E((S_n + X_{n+1})^2 | \sigma(X_1, \dots, X_n)) - E(n+1 | \sigma(X_1, \dots, X_n)) \\ &= E(S_n^2 + 2S_n X_{n+1} + X_{n+1}^2 | \sigma(X_1, \dots, X_n)) - n - 1 \\ &= S_n^2 + 2S_n E(X_{n+1}) + E(X_{n+1}^2) - n - 1 = S_n^2 + 2S_n \cdot 0 + 1 - n - 1 = S_n^2 - n. \end{aligned}$$

Przykład – 15.7. Niech \mathcal{A}_n będzie ciągiem σ -algebr: $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \mathcal{A}_3 \subset \dots \subset \Sigma$. X zmienna losowa, $E(X) \in \mathbb{R}$. Wtedy $\{E(X | \mathcal{A}_n)\}$ jest martyngałem (ćwiczenie).

Nierówność Jensena implikuje ważne twierdzenie:

Twierdzenie – 15.8. Jeżeli $(\{X_n\}, \{\mathcal{A}_n\})$ jest podmartyngałem, $g: \mathbb{R} \rightarrow \mathbb{R}$ funkcją rosnącą i wypukłą, to $(\{g(X_n)\}, \{\mathcal{A}_n\})$ jest podmartyngałem.

Dowód. (ćwiczenie). □

Wniosek – 15.9. Jeżeli $(\{X_n\}, \{\mathcal{A}_n\})$ jest podmartyngałem, to:

- (1) $(\{X_n^+\}, \{\mathcal{A}_n\})$ jest podmartyngałem
- (2) $(\{X_n^2\}, \{\mathcal{A}_n\})$ jest podmartyngałem

15.2 Wybór strategii w grze

Następujące twierdzenie jest jednym z wielu wyników dotyczących martyngałów i mających interpretację w języku teorii gier. Wyobraźmy sobie, że po każdej grze gracz na podstawie znajomości dotychczasowych rezultatów może podjąć decyzję: gra w kolejnej, lub ją opuszcza. Twierdzenie orzeka, że gdy gra jest sprawiedliwa, lub korzystna, to jakakolwiek strategia tego typu nie zmieni jej charakteru.

Twierdzenie – 15.10 (Halmos). Niech $(\{X_n\}, \{\mathcal{A}_n\})$ będzie martyngałem (podmartyngałem), $B_n \in \mathcal{B}(\mathbb{R}^n)$, dla $n = 1, 2, 3, \dots$

$$\varepsilon_n = \begin{cases} 1, & \text{gdy } (X_1, \dots, X_n) \in B_n, \\ 0, & \text{gdy } (X_1, \dots, X_n) \notin B_n. \end{cases}$$

Określamy:

$$Y_1 = X_1, \quad Y_{n+1} = Y_n + \varepsilon_n(X_{n+1} - X_n), \quad \text{dla } n \geq 1.$$

Wtedy:

- (1) $(\{Y_n\}, \{\mathcal{A}_n\})$ jest martyngałem (podmartyngałem).
 (2) $E(Y_n) = E(X_n)$, $(E(Y_n) \leq E(X_n))$, dla $n \geq 1$.

Przypomnienie twierdzenia 13.26.

1. Jeżeli $X : \Omega \rightarrow \mathbb{R}^k$ jest wektorem losowym takim, że X, Y są niezależne, to $E(Y|X) = E(Y)$.
2. Jeżeli Z jest \mathcal{A} -mierzalna oraz $E(ZY) \in \mathbb{R}$, to $E(ZY|\mathcal{A}) = ZE(Y|\mathcal{A})$.
3. Jeżeli $g : \mathbb{R}^k \rightarrow \mathbb{R}$ jest funkcją borelowską, $E(g(X)) \in \mathbb{R}$, to $E(g(X)|X) = g(X)$.
4. Jeżeli X jest zmienną losową, $E(X) \in \mathbb{R}$, to $E(X|X) = X$.

Dowód. Ad (1). Zmienna losowa $\varepsilon_n = \mathbb{1}_{(X_1, \dots, X_n) \in B_n}$ jest $\sigma(X_1, \dots, X_n)$ -mierzalna, a więc \mathcal{A}_n -mierzalna. W takim razie zmienna losowa Y_n jest \mathcal{A}_n mierzalna.

$$\begin{aligned} E(Y_{n+1}|\mathcal{A}_n) &= E(Y_n + \varepsilon_n(X_{n+1} - X_n)|\mathcal{A}_n) = E(Y_n|\mathcal{A}_n) + \varepsilon_n E(X_{n+1} - X_n|\mathcal{A}_n) \\ &= (\geq) Y_n + \varepsilon_n(X_n - X_n) = Y_n. \end{aligned}$$

Stosujemy zapis $= (\geq)$ w sensie alternatywnym, aby jednocześnie dowodzić punkt wykazać dwa warunki. □

Ad (2) Oczywiście $E(Y_1) = E(X_1)$, czyli $E(X_1 - Y_1) = 0$. Dalej mamy:

$$X_{n+1} - Y_{n+1} = X_{n+1} - Y_n - \varepsilon_n(X_{n+1} - X_n) = (1 - \varepsilon_n)(X_{n+1} - X_n) + (X_n - Y_n).$$

Funkcja $1 - \varepsilon_n$ oraz funkcja $X_n - Y_n$ są \mathcal{A}_n -mierzalne, więc z własności 2. w przypomnianym u góry twierdzeniu 13.26 otrzymujemy:

$$E(X_{n+1} - Y_{n+1}|\mathcal{A}_n) = (1 - \varepsilon_n)(E(X_{n+1}|\mathcal{A}_n) - E(X_n|\mathcal{A}_n)) + X_n - Y_n.$$

Z założenia $E(X_{n+1}|\mathcal{A}_n) = (\geq)X_n$, więc

$$E(X_{n+1} - Y_{n+1}|\mathcal{A}_n) = (\geq)(1 - \varepsilon_n)(X_n - X_n) + (X_n - Y_n) = X_n - Y_n.$$

Gdy weźmiemy nadzieję matematyczną obydwu stron i skorzystamy z założenia indukcyjnego, widzimy, że $E(X_{n+1} - Y_{n+1}) = (\geq)E(X_n - Y_n)$. □

Przykład – 15.11. Adam i Bolek zawierają następującą umowę. Bolek będzie co minutę rzucał symetryczną kostką, a przed każdym rzutem Adam będzie decydował, czy podejmuje następujące wyzwanie: wpłaca Bolkowi 3.5 złotych i otrzymuje x złotych, gdzie x jest uzyskaną liczbą oczek. Adam jednak postanawia, że jeżeli w kolejnych dwóch rzutach pojawi się „6”, wtedy nie obstawi kolejnego rzutu (Bolek jednak wykona rzut), a w każdym innym przypadku podejmuje wyzwanie. Zinterpretujemy tę grę w języku twierdzenia Halmosa.

Niech Z_1, Z_2, Z_3, \dots oznaczają liczby oczek w kolejnych rzutach wykonywanych przez Bolka. Niech $X_n = Z_1 + \dots + Z_n - 3.5n$. Jest to martyngał ze względu na filtrację $\sigma(Z_1, \dots, Z_n)$ (ćwiczenie). Określamy zbiory borelowskie B_i w sposób następujący. $B_1 = \mathbb{R}$, $B_2 = \{(x_1, x_2) : \sim (x_1 = 3.5, x_2 = 7)\}$, $B_3 = \{(x_1, x_2, x_3) : \sim (x_2 = x_1 + 3.5, x_3 = x_2 + 3.5)\}, \dots$. Wtedy Y_n określone w twierdzeniu Halmosa jest ciągiem wygranych Adama.

Twierdzenie Halmosa można uogólnić. Poniżej wypowiedź dla przypadku martyngałów.

Twierdzenie – 15.12. Niech $(\{X_n\}, \{\mathcal{A}_n\})$ będzie martyngałem, niech v_n będą \mathcal{A}_n -mierzalne i ograniczone, dla $n = 0, 1, 2, 3, \dots$

Określamy:

$$Y_1 = v_0 X_1, \quad Y_{n+1} = Y_n + v_n(X_{n+1} - X_n), \quad \text{dla } n \geq 1.$$

Wtedy: $(\{Y_n\}, \{\mathcal{A}_n\})_{n=1}^{\infty}$ jest martyngałem. $E(Y_n) = E(v_0 X_1)$ dla $n \geq 1$.

Dowód. Powtórzenie dowodu pierwszej części twierdzenia Halmosa. Uwaga. Założenie o ograniczoność v_n oznacza, że można skorzystać z punkt 2. twierdzenia 13.26 i otrzymać równość: $E(v_n(X_{n+1} - X_n)|\mathcal{A}_n) = v_n E(X_{n+1} - X_n|\mathcal{A}_n)$ (ćwiczenie). □

Gdy obserwujemy wyniki X_n kolejnych gier pewnego gracza, możemy na tej podstawie robić zakłady dotyczące następnej gry i v_n może być interpretowane jako wysokość zakładu w kolejnej grze, natomiast Y_n oznacza nasz zysk (stratę). Jeżeli gra jest sprawiedliwa, nasz średni zysk (strata) nie zmienia się.

15.3 Momenty stopu

Dana jest filtracja $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \mathcal{A}_3 \subset \dots \subset \Sigma$ oraz funkcja $\tau : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$.

Definicja – 15.13. τ jest *momentem stopu* względem filtracji $\{\mathcal{A}_n\} \iff \{\tau \leq n\} \in \mathcal{A}_n$ dla każdego $n \in \mathbb{N}$.

Przykład – 15.14. $\tau = n_0$, $n_0 \in \mathbb{N}$ jest momentem stopu dla każdej filtracji, $n_0 \in \mathbb{N}$ ustalone.

Niech X_1, X_2, X_3, \dots będzie ciągiem zmiennych losowych, Niech $B \subset \mathbb{R}$ będzie zbiorem borelowskim.

$\tau = \min\{k : X_k \in B\}$ jest momentem stopu dla filtracji $\sigma(X_1, \dots, X_n)$,

bo $\{\tau \leq n\} = \bigcup_{k=1}^n \{X_k \in B\} \in \sigma(X_1, \dots, X_n)$.

$\tau = \max\{k : X_k \in B\}$ na ogół **nie jest momentem stopu** dla filtracji $\{\sigma(X_1, \dots, X_n)\}$,

bo $\{\tau \leq n\} = \bigcap_{k=n+1}^{\infty} \{X_k \notin B\}$ na ogół nie należy do $\sigma(X_1, \dots, X_n)$.

Uwaga – 15.15. Przy poprzednich oznaczeniach następujące warunki są równoważne:

1. dla każdego $n \in \mathbb{N}$: $\{\tau \leq n\} \in \mathcal{A}_n$ (τ jest momentem stopu),
2. dla każdego $n \in \mathbb{N}$: $\{\tau > n\} \in \mathcal{A}_n$,
3. dla każdego $n \in \mathbb{N}$: $\{\tau = n\} \in \mathcal{A}_n$.

Dowód. $\{\tau \leq n\} = \bigcup_{k=1}^n \{\tau = k\}$. $\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\}$. □

Uwaga – 15.16. Jeżeli τ_1, τ_2 są momentami stopu, to $\min(\tau_1, \tau_2), \max(\tau_1, \tau_2)$ są momentami stopu.

Dowód. (ćwiczenie). □

Dla danego ciągu zmiennych losowych X_n oraz momentu stopu τ względem filtracji $\{\sigma(X_1, \dots, X_n)\}$ takiego, że $P(\tau < \infty) = 1$ określamy funkcję $X_\tau : \Omega \rightarrow \mathbb{R}$ wzorem

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega).$$

X_τ jest zmienną losową, gdyż dla każdego zbioru borelowskiego B : $\{X_\tau \in B\} = \bigcup_{n=1}^{\infty} \{\tau = n\} \cap \{X_n \in B\} \in \Sigma$.

Następujące twierdzenie Walda pozwala obliczać nadzieję matematyczną sumy losowej liczby składników zmiennych losowych i.i.d.

Twierdzenie – 15.17 (Tożsamość Walda). Niech X_1, X_2, X_3, \dots i.i.d., $E(X_i) \in \mathbb{R}$. $S_n = X_1 + \dots + X_n$. Niech τ będzie takim momentem stopu dla filtracji $\{\sigma(X_1, \dots, X_n)\}$, że $E(\tau) < \infty$. Wtedy:

$$E(S_\tau) = E(\tau)E(X_1).$$

Dowód. Wykorzystamy lemat, którego dowód jest bardzo podobny do dowodu analogicznego lematu 10.18 (ćwiczenie).

Lemat – 15.18. Jeżeli X jest zmienną losową taką, że $P(X \in \mathbb{N}) = 1$, to $E(X) = \sum_{n=1}^{\infty} P(X \geq n)$. □

Zważmy, że dla ustalonego $\omega \in \Omega$: $S_{\tau(\omega)}(\omega) = X_1(\omega) + \dots + X_n(\omega)$, gdzie $n = \tau(\omega)$.

Ponieważ Ω jest sumą rozłącznych zbiorów $\{\tau = n\}$, to

$$S_\tau = X_1 \mathbb{1}_{\{\tau \geq 1\}} + X_2 \mathbb{1}_{\{\tau \geq 2\}} + X_3 \mathbb{1}_{\{\tau \geq 3\}} + \dots = \sum_{n=1}^{\infty} X_n \mathbb{1}_{\{\tau \geq n\}}.$$

Przypadek 1. $X_n \geq 0$ pw. Wtedy $\sum_{n=1}^{\infty} X_n \mathbb{1}_{\{\tau \geq n\}}$ jest dobrze określona. Ponieważ $\mathbb{1}_{\{\tau \geq n\}} = \mathbb{1}_{\{\tau > n-1\}}$ jest $\sigma(X_1, \dots, X_{n-1})$ mierzalna, to zmienne losowe X_n oraz $\mathbb{1}_{\{\tau \geq n\}}$ są niezależne. Dlatego $E(X_n \mathbb{1}_{\{\tau \geq n\}}) = E(X_n)E(\mathbb{1}_{\{\tau \geq n\}}) = E(X_1)P(\{\tau \geq n\})$. Zatem:

$$E(S_\tau) = \sum_{n=1}^{\infty} E(X_n)E(\mathbb{1}_{\{\tau \geq n\}}) = E(X_1) \sum_{n=1}^{\infty} P(\{\tau \geq n\}) = E(X_1)E(\tau).$$

Przypadek 2. Sytuacja ogólna. Stosujemy przypadek 1. do ciągów $\{X_n^+\}$ oraz $\{X_n^-\}$. Wtedy: $E(|S_\tau|) = E(S_\tau^+) + E(S_\tau^-) = E(X_1^+)E(\tau) + E(X_1^-)E(\tau) = E(|X_1|)E(\tau) < \infty$, więc $E(X_\tau) \in \mathbb{R}$ i można powtórzyć rachunki z przypadku 1. □

Twierdzenie – 15.19. Niech $(\{X_n\}, \{\mathcal{A}_n\})$ będzie martyngałem (podmartyngałem, nadmartyngałem), τ momentem stopu względem filtracji $\{\mathcal{A}_n\}$.

Określamy:

$$Y_n = X_{\min(n, \tau)}, \quad \text{dla } n = 1, 2, 3, \dots$$

Wtedy $(\{Y_n\}, \{\mathcal{A}_n\})$ jest martyngałem (podmartyngałem, nadmartyngałem)

Dowód. Ustalmy $n \geq 1$ oraz $\omega \in \Omega$. Albo $\tau(\omega) = k$ dla pewnego $k \leq n-1$, albo $\tau(\omega) \geq n$. W pierwszym przypadku $Y_n(\omega) = X_k(\omega)$, w drugim $Y_n(\omega) = X_n(\omega)$. Zachodzi więc równość:

$$Y_n = \sum_{k=1}^{n-1} X_k \mathbb{1}_{\{\tau=k\}} + X_n \mathbb{1}_{\{\tau > n-1\}}.$$

Y_n są \mathcal{A}_n -mieralne (złożenia funkcji mierzalnych).

$$Y_{n+1} - Y_n = (X_{n+1} - X_n) \mathbb{1}_{\{\tau > n\}}, \quad \text{bo } X_n \mathbb{1}_{\{\tau > n-1\}} = X_n \mathbb{1}_{\{\tau=n\}} + X_n \mathbb{1}_{\{\tau > n\}}$$

Ponieważ $\{\tau > n\} \in \mathcal{A}_n$, to $\mathbb{1}_{\{\tau > n\}}$ jest \mathcal{A}_n mierzalne, więc:

$$E(Y_{n+1} - Y_n | \mathcal{A}_n) = \mathbb{1}_{\{\tau > n\}} E(X_{n+1} - X_n | \mathcal{A}_n),$$

więc łatwo dokończyć dowód. □

Przykład – 15.20. W pewnym kasynie gracz rozpoczyna serię rzutów monetą symetryczną, przy czym uzgodnił następujące warunki: Jeżeli w pierwszym rzucie wypadnie orzeł otrzyma 1000\$ i kończy grę. Jeżeli w pierwszym rzucie wypadnie reszka zapłaci 1000\$ i wykonuje następne rzuty według tych samych zasad, podwajając jednak w każdym rzucie stawkę. Gdy więc gra się kończy (po wypadnięciu orła), gracz otrzyma $2^n \cdot 1000$ \$, gdzie n oznacza liczbę wykonanych rzutów. O ile gra się kiedyś skończy. Czy gra jest opłacalna?

Rozwiązanie bezpośrednie: Jeżeli gracz wyrzuci orła w pierwszym rzucie, to otrzyma 1000\$ i zakończy grę. Jeżeli orzeł po raz pierwszy wypadnie w n -tym rzucie, to gracz zapłaci za $n-1$ przegranych: $1000 + \dots + 2^{n-2} \cdot 1000 = 1000 \cdot \frac{2^{n-1}-2}{2-1} = 2^{n-1} \cdot 1000 - 1000$ dolarów, ale za ostatni rzut otrzyma $2^{n-1} \cdot 1000$ \$. O ile orzeł wypadnie w skończonym czasie gracz zarobi 1000\$.

Interpretacja w języku teorii martyngałów: Niech X_1, X_2, X_3, \dots będą i.i.d. o rozkładzie $P(X_i = -1) = P(X_i = 1) = \frac{1}{2}$, to znaczy 1 utożsamiamy z orłem, -1 z reszką. Niech $\mathcal{A}_n = \sigma(X_1, \dots, X_n)$, $S_n = X_1 + \dots + X_n$.

Wiemy, że $(\{S_n\}, \{\mathcal{A}_n\})$ jest martyngałem.

Niech $v_0 = 1000$. Dla $n = 1, 2, 3, \dots$ określamy $v_n = 2^n \cdot 1000$ o ile $X_1 = \dots = X_{n-1} = -1$ oraz $v_n = 0$ w pozostałych przypadkach.

Niech, podobnie jak w twierdzeniu 15.12,

$$K_1 = v_0 S_1, \quad K_{n+1} = K_n + v_n (S_{n+1} - S_n) \quad \text{dla } n \geq 1.$$

Więc $(\{K_n\}, \{\sigma(X_1, \dots, X_n)\})$ jest martyngałem.

Mamy

$$K_n = v_0 X_1 + v_1 X_2 + \dots + v_{n-1} X_n,$$

co oznacza stan konta gracza po n grach.

Niech $\tau = \min(k : X_k = 1)$. Oczywiście jest to moment stopu względem filtracji $\{\sigma(X_1, \dots, X_n)\}$. Co więcej, τ ma rozkład geometryczny $G_{\frac{1}{2}}$, co oznacza, że $P(\tau < \infty) = 1$. więc gra się kiedyś skończy.

Gdy $\tau = 1$, to $X_1 = 1$, to oznacza, że $K_1 = 1000$. Ustalmy $n > 1$. Gdy $\tau = n$, to oznacza, że $X_1 = \dots = X_{n-1} = -1$ oraz $X_n = 1$. Wtedy: $K_n = -v_0 - \dots - v_{n-2} + v_{n-1} = -1000 - \dots - 2^{n-2} \cdot 1000 + 2^{n-1} \cdot 1000 = -\frac{2^{n-1}-1}{2-1} \cdot 1000 + 2^{n-1} \cdot 1000 = 1000$.

Więc K_τ jest stałą równą 1000.

15.4 Twierdzenie o zbieżności

Jednym z ważniejszych twierdzeń w teorii martyngałów jest:

Twierdzenie – 15.21. Niech $(\{X_n\}, \{\mathcal{A}_n\})$ będzie nadmartyngałem, lub podmartyngałem spełniającym warunek:

$$\sup_n E(|X_n|) < \infty. \quad (15.1)$$

Wtedy istnieje taka zmienna losowa X , że $E(X) \in \mathbb{R}$ oraz:

$$X_n \xrightarrow{1} X.$$

Dowód. Pomijamy. □

Łatwo sprawdzić (ćwiczenie)

Uwaga – 15.22.

W przypadku nadmartyngału: warunek (15.1) $\iff \sup_n E(X_n^-) < \infty$.

W przypadku podmartyngału: warunek (15.1) $\iff \sup_n E(X_n^+) < \infty$.

Jako wniosek otrzymujemy:

Wniosek – 15.23. Niech $(\{X_n\}, \{\mathcal{A}_n\})$ będzie nadmartyngałem, $X_n \geq 0$ dla $n = 1, 2, 3, \dots$

Wtedy istnieje taka zmienna losowa X , że $E(X) \in \mathbb{R}$ oraz:

$$X_n \xrightarrow{1} X.$$

15.5 Pytania

Pytanie 15.1. Niech $a > 0$. Określamy: $X_0 = 1$. Gdy znamy już punkt X_n , określamy X_{n+1} jako liczbę wylosowaną zgodnie z rozkładem $U(0, aX_n)$. Dla jakich a powyższy ciąg jest martyngałem, podmartyngałem, nadmartyngałem.

ROZWIĄZANIE

Pytanie 15.2. Niech $X_1, X_2, X_3, \dots, X_n$ będzie ciągiem niezależnych zmiennych losowych o rozkładzie $B(1, p)$ każda. Niech $S_n = X_1 + \dots + X_n$.

- Wskażać takie rozłączne zbiory A_i oraz liczbę N , że $\sigma(X_1, \dots, X_n) = \sigma(A_1, \dots, A_N)$.
- Wykazać, że $\sigma(S_n) \subsetneq \sigma(X_1, \dots, X_n)$ dla $n = 2$.
- Czy $\sigma(X_1, \dots, X_n) = \sigma(S_1, \dots, S_n)$? Odpowiedź uzasadnić.

ROZWIĄZANIE

Pytanie 15.3. Wykazać twierdzenie wypowiedziane w przykładzie 15.7. ROZWIĄZANIE

Pytanie 15.4. Czy suma, różnica, iloczyn dwóch momentów stopu względem tej samej filtracji jest momentem stopu? Uzasadnić. ROZWIĄZANIE

Pytanie 15.5. Podać dowód uwagi 15.16 ROZWIĄZANIE

Pytanie 15.6. Wykazać uwagę 15.22 i wniosek 15.23. ROZWIĄZANIE

Rozdział 16

Definicja i przykłady łańcuchów Markowa

Przedstawimy jedną z najprostszych sytuacji, gdy rozważne zmienne losowe są zależne. Warto podkreślić, że łańcuchy Markowa, które będziemy za chwilę omawiać, stanowią bardzo interesujący przykład procesów stochastycznych. Ich teoria ma z kolei podstawowe znaczenie przy budowie probabilistycznych modeli wielu zjawisk przyrodniczych, technicznych, a także ekonomicznych. W szczególności, teoria procesów stochastycznych znajduje w ostatnich latach coraz większe znaczenie w wycenie instrumentów finansowych.

16.1 Jednorodny łańcuch Markowa

Niech $M \subset \mathbb{R}^d$ będzie zbiorem skończonym lub przeliczalnym i niech: $\mathbf{P} : M \times M \rightarrow \mathbb{R}$, $\mathbf{p} : M \rightarrow \mathbb{R}$.

Będziemy myśleć o \mathbf{P} i \mathbf{p} jako o skończonej lub przeliczalnej macierzy o wyrazach $\mathbf{P}(i, j)$ oraz wektorze (macierzy jedno kolumnowej) o współrzędnych $\mathbf{p}(i)$, gdzie $i, j \in M$.

Definicja – 16.1 (Jednorodny łańcuch Markowa). Niech $\{X_n\}$, $n = 0, 1, 2, \dots$, będzie ciągiem wektorów losowych określonych na przestrzeni probabilistycznej (Ω, Σ, P) i przyjmujących wartości w \mathbb{R}^d .

Mówimy, że $\{X_n\}$ jest jednorodnym łańcuchem Markowa, jeżeli spełnione są następujące warunki.

1. Dla każdego $i \in M : P(X_0 = i) = \mathbf{p}(i)$.

2. Dla każdego $n \geq 0$ zachodzi

$$P(X_{n+1} = i_{n+1} \mid (X_0 = i_0, \dots, X_n = i_n)) = P(X_{n+1} = i_{n+1} \mid X_n = i_n) = \mathbf{P}(i_n, i_{n+1}),$$

dla każdego ciągu $\{i_0, \dots, i_{n+1}\} \subset M$, o ile $P(X_0 = i_0, \dots, X_n = i_n) > 0$.

3. $\sum_{i \in M} \mathbf{p}(i) = 1$.

4. $\sum_{j \in M} \mathbf{P}(i, j) = 1$, dla każdego $i \in M$.

Interpretacja.

1. M – zbiór wszystkich możliwych stanów pewnego systemu.
2. X_n – stan, w którym znajduje się system w chwili czasowej n .
3. Warunek, że X_n jest zmienną losową, oznacza, że faktycznie nie znamy dokładnie tego położenia.
4. Znamy rozkład prawdopodobieństwa położenia systemu w chwili zerowej,
5. Prawdopodobieństwo przejścia układu z jednego stanu do innego stanu w jednostkowym odcinku czasu zależy jedynie od tych stanów, a nie zależy od historii układu ani od konkretnej chwili, w której to przejście następuje.
6. Układ nigdy nie opuści swojej przestrzeni stanów M , gdyż $P(X_0 \in M) = \sum_{i \in M} \mathbf{p}_i = 1$, a wzór na

prawdopodobieństwo całkowite oraz warunek 4 implikują:

$$\begin{aligned} P(X_{n+1} \in M) &= \sum_{j \in M} P(X_{n+1} = j) = \sum_{j \in M} \sum_{i \in M} P(X_{n+1} = j \mid X_n = i) P(X_n = i) \\ &= \sum_{i \in M} \sum_{j \in M} \mathbf{P}(i, j) P(X_n = i) = \sum_{i \in M} 1 P(X_n = i) = 1. \end{aligned}$$

W związku z powyższą interpretacją będziemy nazywać:

M – zbiorem stanów, lub przestrzenią stanów,

\mathbf{p} – rozkładem początkowym,

\mathbf{P} – macierzą przejścia łańcucha Markowa.

Rozważa się też niejednorodny łańcuch Markowa dopuszczając możliwość, że prawdopodobieństwo przejścia zależy od chwili w której to przejście następuje. W tym kursie nie zajmujemy się jednak takimi łańcuchami i w dalszym ciągu dla prostoty wypowiedzi opuszczamy słowo „jednorodny”.

Przykład – 16.2 (kontynuacja przykładu 3.2). Kaja i Leon umówili się w sprawie sprzątania, a ponieważ Kaja sprząta dokładniej niż Leon, ustalili następujące zasady. Jeżeli w pewnym dniu sprząta Leon, to rzuca kostką i jeżeli nie wyrzuci „6”, to sprząta także w następnym dniu, gdy wypadnie „6”, to sprząta Kaja. Jeżeli sprząta Kaja, to w następnym dniu nie sprząta nikt. Jeżeli w jakimś dniu nikt nie sprząta, to o sprzątaniu w następnym dniu decyduje rzut monetą. O sprzątaniu w pierwszym dniu umowy decyduje rzut monetą.

Modelem powyższej sytuacji może być łańcuch Markowa $\{X_n\}$, $n = 0, 1, 2, \dots$, w którym X_n są zmiennymi losowymi o wartościach w M , gdzie:

$$M = \{\text{Kaja, Leon, Nikt}\}, \quad \mathbf{p}^T = \left[\frac{1}{2}, \frac{1}{2}, 0 \right], \quad \mathbf{P} = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{6} & \frac{5}{6} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

Dyskutowaliśmy już poprzednio spacer losowy po prostej, patrz sekcja 11.4. Jak się okazuje są one łańcuchami Markowa.

Przykład – 16.3 (Spacer losowy po prostej). Wyobraźmy sobie cząsteczkę, która może się poruszać wzdłuż linii prostej według następujących reguł. W chwili zero cząsteczka znajduje się w punkcie o współrzędnej zero, natomiast w następnych momentach czasu $1, 2, 3, \dots$ może się przesuwać o jeden w lewo lub o jeden w prawo z prawdopodobieństwami odpowiednio q oraz p , przy czym $p + q = 1$. Jeżeli $p = q = \frac{1}{2}$, mówimy, że spacer losowy jest standardowy.

Spacer losowy jest rzeczywiście łańcuchem Markowa. Mianowicie, stanami są wszystkie możliwe liczby całkowite, czyli $M = \mathbb{Z} \subset \mathbb{R}$.

X_n oznacza pozycję cząsteczki w chwili n .

Zdefiniujmy:

$$\mathbf{p}(i) = \begin{cases} 1 & \text{dla } i = 0, \\ 0 & \text{dla } i \neq 0 \end{cases}$$

oraz

$$\mathbf{P}(i, j) = \begin{cases} q & \text{dla } j = i - 1, \\ p & \text{dla } j = i + 1, \\ 0 & \text{w innych przypadkach.} \end{cases}$$

Mamy więc: $P(X_0 = 0) = 1$, $P(X_0 = i) = 0$ dla $i \neq 0$,

$P(X_{n+1} = i - 1 | X_n = i) = q$, $P(X_{n+1} = i + 1 | X_n = i) = p$, $P(X_{n+1} = i | X_n = j) = 0$ dla $|i - j| \neq 1$.

Określony powyżej spacer losowy może być modyfikowany na różne sposoby. Na przykład, założmy, że cząsteczka może nie zmieniać swojego położenia z prawdopodobieństwem r . Oczywiście wtedy zakładamy, że $p + q + r = 1$. Inną modyfikacją jest założenie o istnieniu jednej lub dwóch barier (ekranów), które ograniczają możliwość ruchu cząsteczki i są usytuowane w punktach, powiedzmy, $A < 0 < B$.

Wtedy zbiór M składa się $A + B + 1$ stanów, a $(A + B + 1)$ -wymiarowa macierz \mathbf{P} może być zdefiniowana na przykład tak:

$$\mathbf{P} = \begin{bmatrix} s_A & 1 - s_A & 0 & 0 & \cdots & 0 \\ q & r & p & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & q & r & p \\ 0 & \cdots & 0 & 0 & 1 - s_B & s_B \end{bmatrix}.$$

Liczby s_A oraz s_B oznaczają prawdopodobieństwa tego, że cząsteczka jest pochłaniana przez barierę A lub B . Dwa interesujące przypadki skrajne są wtedy, gdy liczby te są albo zerami, co oznacza pełną elastyczność barier, albo są jedynkami, co oznacza pełną absorpcję cząsteczki z chwilą jej dojścia do bariery.

Przykład – 16.4. Animacja pokazuje pierwszych 500 kroków wędrówki cząstki startującej z punktu 0, gdy bariery ustawione są w punktach $A = -5$, $B = 5$, a prawdopodobieństwa wynoszą: $p = 0.2$, $q = 0.25$, $r = 0.55$, $s_A = 0.1$, $s_B = 0.7$.

Można też opisać spacer losowy, używając innego podejścia.

Załóżmy, nieco ogólniej niż poprzednio, że cząsteczka startuje w chwili zero z punktu i . Gdy nie uwzględniamy barier, mamy:

$$X_0 = i, \quad \text{oraz} \quad X_n = X_{n-1} + \xi_n, \quad \text{dla } n = 1, 2, 3, \dots,$$

gdzie $\xi_1, \xi_2, \xi_3, \dots$ są niezależnymi zmiennymi losowymi przyjmującymi wartości $-1, 0, 1$ z prawdopodobieństwami odpowiednio q, r, p .

Można także rozpatrywać spacery losowe na płaszczyźnie i ogólnie w przestrzeni wielowymiarowej.

Przykład – 16.5. Dla uproszczenia załóżmy, że $p = q = \frac{1}{2}$, czyli także $r = 0$. Dla $i \in \mathbb{Z}^d$ mamy:

$$X_0 = i, \quad \text{oraz} \quad X_n = X_{n-1} + \xi_n, \quad \text{dla } n = 1, 2, 3, \dots$$

Tym razem $\xi_1, \xi_2, \xi_3, \dots$ są niezależnymi wektorami losowymi przyjmującymi 2^d wartości $(\varepsilon_1, \dots, \varepsilon_d)$, gdzie $\varepsilon_j = \pm 1$, z jednakowym prawdopodobieństwem $\frac{1}{2^d}$ (bo każde ε_j przyjmuje wartości ± 1 z prawdopodobieństwem $\frac{1}{2}$).

Zauważmy, że współrzędnymi d -wymiarowego spaceru losowego są niezależne jednowymiarowe standardowe spacery losowe.

Prezentowany powyżej mechanizm tworzenia łańcucha Markowa można istotnie uogólnić.

Twierdzenie – 16.6. Niech $M \subset \mathbb{R}^d$ będzie zbiorem skończonym lub przeliczalnym, $B \subset \mathbb{R}^k$ zbiorem borelowskim. Załóżmy, że $T : M \times B \rightarrow M$ jest odwzorowaniem spełniającym warunek mierzalności:

$$\forall i, j \in M : \{y \in \mathbb{R}^k : T(i, y) = j\} \in \mathcal{B}(\mathbb{R}^k).$$

Niech $\eta, \xi_1, \xi_2, \xi_3, \dots$ będzie ciągiem niezależnych wektorów losowych, przy czym η ma wartości w M , a k -wymiarowe wektory $\xi_1, \xi_2, \xi_3, \dots$ mają identyczny rozkład na zbiorze B . Definiujemy:

$$X_0 = \eta, \quad X_n = T(X_{n-1}, \xi_n), \quad \text{dla } n \geq 1.$$

Wtedy ciąg $\{X_n\}$ jest łańcuchem Markowa.

Dowód. Niech $\mathbf{p}(i) = P(\eta = i)$ oraz $\mathbf{P}(i, j) = P(T(i, \xi_n) = j)$ dla $i, j \in M$. Wtedy:

$$\begin{aligned} P(X_{n+1} = i_{n+1} \mid (X_0 = i_0, \dots, X_n = i_n)) &= \frac{P(X_{n+1} = i_{n+1}, X_0 = i_0, \dots, X_n = i_n)}{P(X_0 = i_0, \dots, X_n = i_n)} \\ &= \frac{P(T(i_n, \xi_{n+1}) = i_{n+1}, X_0 = i_0, \dots, X_n = i_n)}{P(X_0 = i_0, \dots, X_n = i_n)} \\ &= \frac{P(T(i_n, \xi_{n+1}) = i_{n+1})P(X_0 = i_0, \dots, X_n = i_n)}{P(X_0 = i_0, \dots, X_n = i_n)} = P(T(i_n, \xi_{n+1}) = i_{n+1}) = \mathbf{P}(i_n, i_{n+1}). \end{aligned}$$

Podobnie $P(X_{n+1} = i_{n+1} \mid X_n = i_n) = \frac{P(T(i_n, \xi_{n+1}) = i_{n+1})P(X_n = i_n)}{P(X_n = i_n)} = P(T(i_n, \xi_{n+1}) = i_{n+1}) = \mathbf{P}(i_n, i_{n+1})$. \square

Przykład – 16.7 (Urnowy model Bernoulliego). W każdej z dwóch urn umieszczono k kul, przy czym k z nich ma kolor biały, a k ma kolor czerwony. Następnie w kolejnych momentach losujemy jednocześnie po jednej kuli z każdej urny i przekładamy je do drugiej urny. Niech X_n oznacza liczbę białych kul w pierwszej urnie (więc tym samym liczbę czerwonych kul w drugiej urnie) w chwili n . Widzimy, że zmienne X_n tworzą łańcuch Markowa na przestrzeni stanów $M = \{0, 1, 2, \dots, k\}$ z macierzą przejścia \mathbf{P} mającej zerowe wyrazy oprócz

$$\mathbf{P}(i, i-1) = \left(\frac{i}{k}\right)^2, \quad \mathbf{P}(i, i+1) = \left(\frac{k-i}{k}\right)^2, \quad \mathbf{P}(i, i) = \frac{2(k-i)i}{k^2}.$$

dla $0 < i < k$ oraz $\mathbf{P}(0, 1) = 1$, $\mathbf{P}(k, k - 1) = 1$.

Jeżeli na początku eksperymentu w pierwszej urnie było b_0 białych kul, to $\mathbf{p}(b_0) = 1$ oraz $\mathbf{p}(i) = 0$ dla $i \neq b_0$.

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/16 & 3/8 & 9/16 & 0 & 0 \\ 0 & 1/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & 9/16 & 3/8 & 1/16 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad k = 4.$$

Alternatywnie powyższy łańcuch można opisać tak:

$$X_0 = b_0, \quad X_n = T(X_{n-1}, \xi_n), \quad \text{dla } n = 1, 2, 3, \dots$$

ξ_n są niezależnymi wektorami losowymi mającymi rozkład jednostajny na zbiorze $\{1, \dots, k\}^2$, $T: M \times \{1, \dots, k\}^2 \rightarrow M$ jest określone jako:

$$T(x, y) = \begin{cases} x - 1, & \text{gdy } y_1 \leq x, y_2 > k - x \\ x, & \text{gdy } y_1 \leq x, y_2 \leq k - x \text{ lub } y_1 > x, y_2 > k - x \\ x + 1, & \text{gdy } y_1 > x, y_2 \leq k - x. \end{cases}$$

Faktycznie, przed każdym losowaniem można (można hipotetycznie) ponumerować kule w urnach w taki sposób, że białe kule w obydwóch urnach mają początkowe numery: od 1 do x w pierwszej urnie, a więc od 1 do $k - x$ w drugiej urnie, natomiast czarne kule mają pozostałe numery: od $x + 1$ do k w pierwszej oraz od $k - x + 1$ do k w drugiej urnie.

16.2 Macierz przejścia i jej potęgi

Wyznamy rozkłady zmiennych losowych X_n tworzących łańcuch Markowa.

$$\mathbf{p}_n(j) = P(X_n = j)$$

dla wszystkich $n \geq 1$ oraz $j \in M$.

Stosując wzór na prawdopodobieństwo całkowite, mamy:

$$\mathbf{p}_n(j) = P(X_n = j) = \sum_{i \in M} P(X_n = j \mid X_{n-1} = i)P(X_{n-1} = i) = \sum_{i \in M} \mathbf{P}(i, j)\mathbf{p}_{n-1}(i).$$

Czyli

$$\mathbf{p}_n = \mathbf{P}^T \mathbf{p}_{n-1},$$

gdzie \mathbf{P}^T oznacza transpozycje macierzy \mathbf{P} .

Oznaczając n -tą potęgę macierzy \mathbf{P} przez \mathbf{P}^n , otrzymujemy wreszcie poszukiwany rozkład:

$$\mathbf{p}_n = (\mathbf{P}^T)^n \mathbf{p}_0.$$

W szczególności, jeżeli wiemy, że $X_0 = i$, czyli że łańcuch znajduje się w stanie i z prawdopodobieństwem 1, powyższy wzór implikuje:

$$\mathbf{p}_n(j) = \mathbf{P}^n(i, j), \quad \text{dla wszystkich } n,$$

co wyjaśnia znaczenie współczynników $\mathbf{P}^n(i, j)$ n -tej potęgi macierzy przejścia \mathbf{P} .

Przykład – 16.8. Antoni i Bolesław, mają kapitał odpowiednio A i B złotych. Powtarzają oni tę samą grę (może grają w szachy), przy czym przegrywający płaci wygrywającemu złotówkę. Gra kończy się wtedy, gdy jednemu z graczy skończą się pieniądze. Załóżmy, że w każdej grze prawdopodobieństwo wygrania przez Antoniego wynosi p , a prawdopodobieństwo wygrania przez Bolesława q . Zakładamy, że $p + q \leq 1$ i oznaczamy przez r prawdopodobieństwo remisu, $r = 1 - p - q$. Oznaczmy kapitał Antoniego po zakończeniu n -tej gry przez X_n .

Opisana sytuacja jest faktycznie spacerem losowym startującym w punkcie o współrzędnej A i mającym bariery pochłaniające w punktach o współrzędnych 0 oraz $A + B$.

Zakładamy, że Antoni ma 8 złotych, a Bolesław 5 złotych. Bolesław gra na ogół lepiej niż Antoni: zakładamy: $p = 0.2$, $q = 0.4$, $r = 0.4$. Grają 200 razy, chyba że jednemu z nich zabraknie wcześniej pieniędzy.

X_n – kapitał Antoniego po rozegraniu n gier.

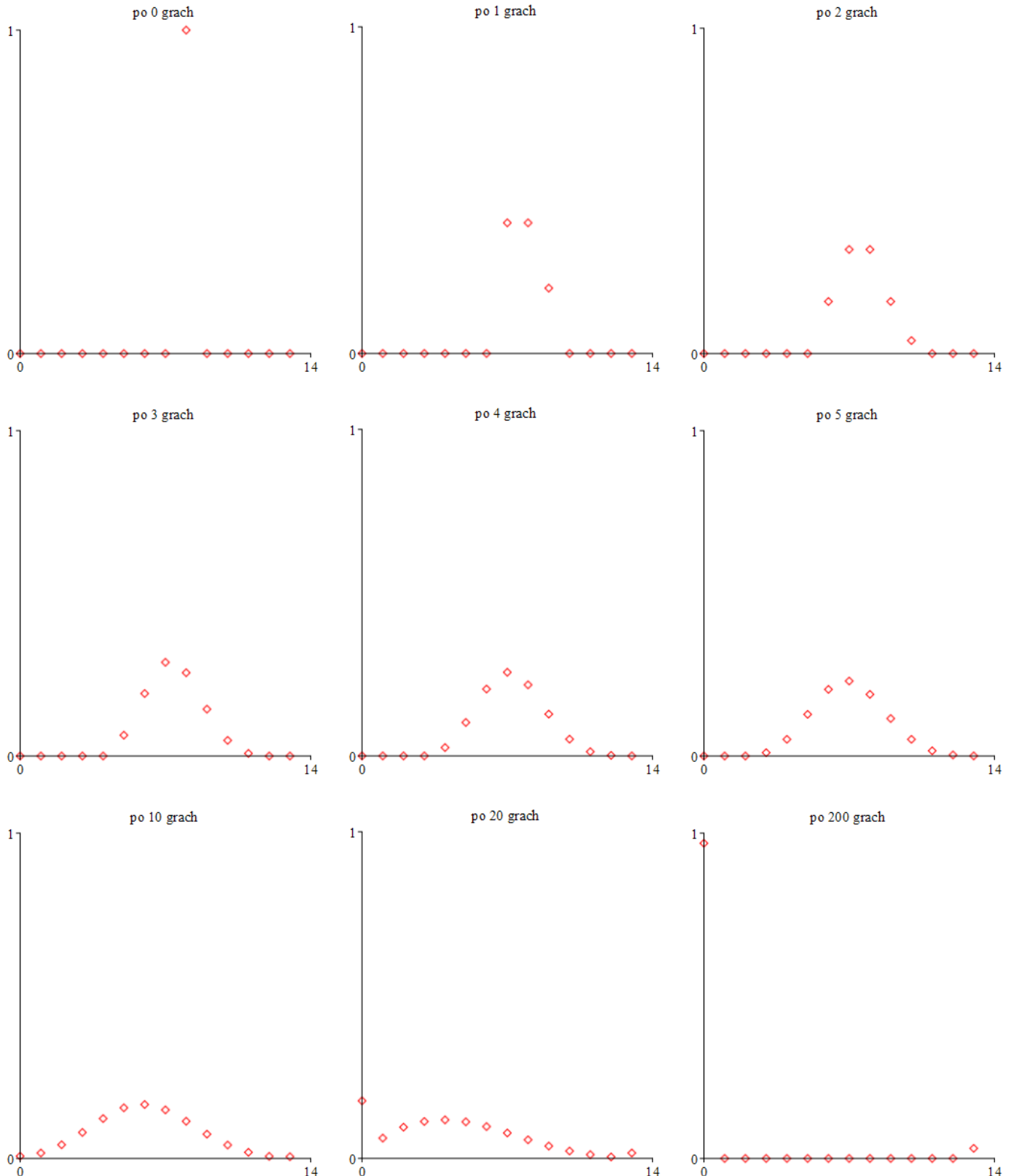
X_0 ma rozkład jednopunktowy δ_8 ,

X_1 ma rozkład dany przez: $[0, 0, 0, 0, 0, 0, 0, 0, 0, 4, 0, 4, 0, 2, 0, 0, 0, 0]$, $E(X_1) = 7.8$.

X_5 ma rozkład dany przez: $[0, 0, 0, 0.102e-1, 0.512e-1, 0.128, 0.205, 0.230, 0.189, 0.115, 0.512e-1, 0.160e-1, 0.320e-2, 0.32e-3]$, $E(X_3) = 7.000$

X_{20} ma rozkład dany przez: $[0.176, 0.620e-1, 0.954e-1, 0.113, 0.118, 0.112, 0.973e-1, 0.776e-1, 0.568e-1, 0.378e-1, 0.225e-1, 0.114e-1, 0.422e-2, 0.165e-1]$, $E(X_{20}) = 4.1579$

X_{200} ma rozkład dany przez: $[0.969, 0.117e-4, 0.160e-4, 0.161e-4, 0.142e-4, 0.114e-4, 0.854e-5, 0.604e-5, 0.402e-5, 0.25e-5, 0.1430e-5, 0.7070e-6, 0.2570e-6, 0.311e-1]$, $E(X_{200}) = .4050783639$



Niech A oznacza zbiór opisany przez zmienne losowe X_0, \dots, X_{n-1} , czyli A ma postać:

$$A = \bigcup \{X_0 = i_0, \dots, X_{n-1} = i_{n-1}\},$$

gdzie suma jest brana po pewnym zbiorze, powiedzmy B , indeksów i_0, \dots, i_{n-1} . Mamy wtedy:

Twierdzenie – 16.9. *Zachodzi równość*

$$P(X_{n+1} = j \mid (X_n = i \text{ oraz } A)) = \mathbf{P}(i, j).$$

Dowód. Zauważmy najpierw, że:

$$\begin{aligned} P(X_{n+1} = j \mid (X_n = i \text{ oraz } A)) &= \frac{P(X_{n+1} = j, X_n = i, A)}{P(X_n = i, A)} \\ &= \frac{\sum P(X_{n+1} = j, X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)}{\sum P(X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)}, \end{aligned}$$

gdzie obie sumy brane są po zbiorze B .

Z własności 2. w definicji łańcucha Markowa mamy:

$$\begin{aligned} P(X_{n+1} = j, X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= P(X_{n+1} = j \mid (X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)) \cdot P(X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= P(X_{n+1} = j \mid X_n = i) \cdot P(X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= \mathbf{P}(i, j) \cdot P(X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0), \end{aligned}$$

więc

$$P(X_{n+1} = j \mid (X_n = i \text{ oraz } A)) = \mathbf{P}(i, j). \quad \square$$

Następne twierdzenie daje inną, bardziej ogólną, interpretację współczynników $\mathbf{P}^k(i, j)$ macierzy \mathbf{P}^k jako prawdopodobieństw przejścia w k krokach ze stanu i do stanu j .

Twierdzenie – 16.10. *Dla każdego $k \geq 1$ oraz $i, j \in M$ mamy*

$$P(X_{n+k} = j \mid X_n = i) = \mathbf{P}^k(i, j).$$

Dowód. Dla $k = 1$ formuła jest konsekwencją własności 2. w definicji łańcucha Markowa.

Załóżmy dla przeprowadzenia kroku indukcyjnego, że zachodzi powyższy wzór dla pewnego k . Wykażemy go dla $k + 1$. Mamy:

$$\begin{aligned} P(X_{n+k+1} = j \mid X_n = i) &= \frac{P(X_{n+k+1} = j, X_n = i)}{P(X_n = i)} \\ &= \frac{\sum_{l \in M} P(X_{n+k+1} = j, X_{n+k} = l, X_n = i)}{P(X_n = i)} \\ &= \frac{\sum_{l \in M} P(X_{n+k+1} = j \mid X_{n+k} = l, X_n = i) P(X_{n+k} = l, X_n = i)}{P(X_n = i)}. \end{aligned}$$

Założenie indukcyjne oraz poprzednie Twierdzenie daje:

$$\begin{aligned} P(X_{n+k+1} = j \mid X_n = i) \\ &= \frac{\sum_{l \in M} P(X_{n+k+1} = j \mid X_{n+k} = l) P(X_{n+k} = l \mid X_n = i) P(X_n = i)}{P(X_n = i)} \\ &= \sum_{l \in M} \mathbf{P}(l, j) \mathbf{P}^k(i, l) = \mathbf{P}^{k+1}(i, j). \quad \square \end{aligned}$$

Tak więc $\mathbf{P}^k(i, j)$ jest prawdopodobieństwem przejścia w k krokach ze stanu i do stanu j .

Warunek $\mathbf{P}^k(i, j) > 0$ oznacza, że takie przejście jest możliwe.

Zauważmy dalej, że dla każdych trzech stanów i, j, k :

$$\mathbf{P}^{m+n}(i, j) = \sum_{l \in M} \mathbf{P}^m(i, l) \mathbf{P}^n(l, j) \geq \mathbf{P}^m(i, k) \mathbf{P}^n(k, j).$$

Odpowiada to naszej intuicji, która podpowiada, że jeżeli jest możliwe przejście ze stanu i do stanu k oraz ze stanu k do stanu j , to możliwe jest także przejście ze stanu i do j .

16.3 Pytania

Pytanie 16.1. Na początku w urnie są trzy białe kule. Co minutę losujemy z urny jedną kulę i jeżeli jest czerwona, to wrzucamy ją z powrotem, a jeżeli jest biała, to z takim samym prawdopodobieństwem wrzucamy tę kulę do urny albo zamiast niej wrzucamy kulę czerwoną. Niech C_n oznacza liczbę czerwonych kul w urnie po upływie n minut. Wskazać rozkład prawdopodobieństwa zmiennej losowej C_3 . (ROZWIĄZANIE)

Pytanie 16.2. Kontynuując przykład 16.2.

- (1) Obliczyć prawdopodobieństwo tego, że w trzecim dniu umowy: sprząta Kaja, sprząta Leon.
- (2) Używając komputera obliczyć prawdopodobieństwo tego, że Leon sprząta: w dziesiątym dniu umowy, w setnym dniu umowy. (ROZWIĄZANIE)

Pytanie 16.3. Uzupełnić lukę w dowodzie twierdzenia 16.6. (ROZWIĄZANIE)

Pytanie 16.4. Niech ciąg $(X_0, X_1, X_2, X_3, \dots)$ będzie łańcuchem Markowa, Wykazać, że podciąg (X_1, X_3, X_5, \dots) oraz podciąg (X_2, X_4, X_6, \dots) są łańcuchami Markowa o odpowiednich parametrach. (ROZWIĄZANIE)

Pytanie 16.5. Niech $\{X_t\}$ będzie ciągiem wektorów losowych określonych przez algorytm PRS, patrz punkt 12.4. Zakładając, że A jest zbiorem skończonym, wykazać, że ciąg ten jest łańcuchem Markowa: wskaż macierz przejścia. Zweryfikować uzyskany wynik. (ROZWIĄZANIE)

Pytanie 16.6. Wskazać taki ciąg zmiennych losowych X_t , $t = 0, 1, 2, 3, \dots$ o wartościach w zbiorze co najwyżej przeliczanym M , że:

- (1) $\{X_t\}$ jest martyngałem, $\{X_t\}$ jest łańcuchem Markowa.
- (2) $\{X_t\}$ jest martyngałem, $\{X_t\}$ nie jest łańcuchem Markowa.
- (3) $\{X_t\}$ nie jest martyngałem, $\{X_t\}$ łańcuchem Markowa.
- (4) $\{X_t\}$ nie jest martyngałem, $\{X_t\}$ jest nie łańcuchem Markowa.

(ROZWIĄZANIE)

Rozdział 17

Nieredukowalne łańcuchy Markowa

Definicja – 17.1. Łańcuch Markowa jest *nieredukowalny*, jeżeli dla każdych dwóch stanów i oraz j prawdopodobieństwo przejścia $\mathbf{P}^k(i, j)$ jest dodatnie dla pewnego $k > 0$ (zależne od i, j). Warunek ten oznacza, że każde dwa stany się ze sobą *komunikują*.

Większość łańcuchów Markowa spotykanych w zastosowaniach jest nieredukowalna, jakkolwiek łatwo pokazać przykłady łańcuchów, które nie spełniają tej własności.

Spacer losowy z ekranami pochłaniającymi nie jest nieredukowalny, gdyż prawdopodobieństwo przejścia z jednego do drugiego ekranu jest równe zeru.

17.1 Powracanie

Dla nieredukowalnego łańcucha Markowa oznaczmy prawdopodobieństwo pierwszego powrotu do stanu i dokładnie w n krokach przez $f_n(i)$, czyli

$$f_n(i) = P(X_n = i, X_{n-1} \neq i, \dots, X_1 \neq i \mid X_0 = i).$$

Określmy $F(i)$ jako $F(i) = \sum_{n=1}^{\infty} f_n(i)$. Jest to więc prawdopodobieństwo powrotu do stanu i w czasie skończonym.

Jako prawdopodobieństwo, $F(i)$ jest nie większe niż 1. Będziemy mówić, że stan i jest *powracający*, jeżeli $F(i) = 1$ i *niepowracający*, jeżeli $F(i) < 1$.

Następujące twierdzenie jest prostym uogólnieniem twierdzenia 11.26. Pozwala ono w wielu przypadkach stwierdzić, czy stan łańcucha Markowa jest powracający, czy niepowracający. Oznaczmy:

$$\mathbf{P}(i) = \sum_{n=1}^{\infty} \mathbf{P}^n(i, i).$$

Twierdzenie – 17.2. Niech $i \in M$ będzie ustalonym stanem nieredukowalnego łańcucha Markowa. Wtedy:

1. Stan i jest powracający, wtedy i tylko wtedy, gdy $\mathbf{P}(i) = \infty$.
2. Jeżeli i jest stanem niepowracającym, to $F(i) = \frac{\mathbf{P}(i)}{1 + \mathbf{P}(i)}$.

Dowód. Dowód polega na powtórzeniu rozumowania zastosowanego w uzasadnieniu twierdzenia 11.26, które dotyczyło szczególnego łańcucha Markowa. Zauważmy, że zdefiniowane tam prawdopodobieństwa a_n oraz f_n są szczególnymi przypadkami $\mathbf{P}^n(i, i)$ oraz $f_n(i)$. Szczegóły pozostawiamy Czytelnikowi (ćwiczenie). \square

Wykażemy, że albo wszystkie stany są powracające, albo wszystkie stany są niepowracające. W związku z tym mówimy, że łańcuch Markowa (nieredukowalny) jest odpowiednio powracający albo niepowracający.

Lemat – 17.3. Niech $i, j \in M$ będą stanami nieredukowalnego łańcucha Markowa. Wtedy:

$$\sum_{n=1}^{\infty} \mathbf{P}^n(i, i) < \infty \iff \sum_{n=1}^{\infty} \mathbf{P}^n(j, j) < \infty.$$

Dowód. Istnieją takie liczby s, t , że $\mathbf{P}^s(i, j) > 0$ oraz $\mathbf{P}^t(j, i) > 0$. Oznaczmy te dwie ostatnie wielkości odpowiednio przez c oraz d . Wybierzmy dowolną liczbę naturalną n . Wtedy

$$\mathbf{P}^{n+t+s}(i, i) \geq \mathbf{P}^s(i, j)\mathbf{P}^n(j, j)\mathbf{P}^t(j, i) = cd\mathbf{P}^n(j, j).$$

Jeżeli szereg $\sum_{n=1}^{\infty} \mathbf{P}^n(i, i)$ jest zbieżny, to oczywiście $\sum_{n=1}^{\infty} \mathbf{P}^{n+t+s}(i, i)$ jest zbieżny i stosując kryterium porównawcze zbieżności szeregów widzimy, że $\sum_{n=1}^{\infty} \mathbf{P}^n(j, j)$ też jest zbieżny. Rozumowanie symetryczne kończy dowód. \square

Liczby $\mathbf{P}(i)$ mają także nieco inną interpretację. Oznaczmy przez r_i liczbę wszystkich powrotów do stanu i .

Twierdzenie – 17.4. *Dla każdego $i \in M$, $E(r_i) = \mathbf{P}(i)$.*

Dowód. Załóżmy, że w chwili 0 system znajdował się w stanie i . W takim razie $\mathbf{p}(i) = 1$ oraz $\mathbf{p}(j) = 0$ dla $j \neq i$. Mamy więc

$$P(X_n = i) = P(X_n = i \mid X_0 = i) = \mathbf{P}^n(i, i).$$

Wiemy, że wartość oczekiwana funkcji charakterystycznej $\mathbb{1}_{\{X_n=i\}}$ wynosi $\mathbf{P}^n(i, i)$. Mamy też: $r_i = \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=i\}}$, co właśnie oznacza tezę. \square

Stosując twierdzenie 11.26 wykazaliśmy, że standardowy spacer losowy po prostej jest łańcuchem powracającym. Przypomnijmy, że efekt ten uzyskaliśmy korzystając z faktu, że dla dowolnego stanu $i \in \mathbb{Z}$: $\mathbf{P}^n(i, i) = 0$ dla nieparzystych n oraz na podstawie wzoru Stirlinga $\mathbf{P}^{2k}(i, i) \cong \frac{1}{\sqrt{\pi k}}$ dla dużych k . Ponieważ relacja \cong implikuje równoczesną zbieżność, albo rozbieżność, szeregów o wyrazach ciągów w niej występujących, wyraźnie widać, że $\sum_{n=1}^{\infty} \mathbf{P}^n(i, i) = \infty$.

Przykład – 17.5. Rozważmy spacer losowy d -wymiarowy opisany w przykładzie 16.5, $d \geq 2$.

Niech \mathbf{P}_d oznacza macierz przejścia naszego łańcucha Markowa. Ustalmy stan $i = (i_1, \dots, i_d) \in \mathbb{Z}^d$. Bez straty ogólności można założyć, że $i = (0, \dots, 0)$. Widzimy teraz, że przejście w n krokach ze stanu i z powrotem do tego stanu podczas d -wymiarowego spaceru losowego jest równoważne przejściom ze stanów i_j do i_j w n krokach podczas jednowymiarowych spacerów losowych niezależnych od siebie. Właśnie korzystając z tej niezależności, mamy:

$$\mathbf{P}_d^n(i, i) = \mathbf{P}^n(i_1, i_1) \cdot \dots \cdot \mathbf{P}^n(i_d, i_d) = (\mathbf{P}^n(0, 0))^d.$$

Mamy więc, że $\mathbf{P}_d^n(i, i) = \mathbf{P}^n(0, 0)^2 \cong \left(\frac{1}{\sqrt{\pi k}}\right)^2 = \frac{1}{\pi k}$, gdy $n = 2k$, $\mathbf{P}_d^n(i, i) = 0$, gdy $n = 2k - 1$. Tak więc szereg $\sum_{n=1}^{\infty} \mathbf{P}_d^n(i, i) = \infty$, a więc również dwuwymiarowy spacer losowy jest powracający.

Zauważmy, że dla $d \geq 3$ jest $\mathbf{P}_d^n(i, i) = \mathbf{P}^n(0, 0)^d \cong \left(\frac{1}{\sqrt{\pi k}}\right)^d = \frac{1}{\sqrt{\pi^d k^{\frac{d}{2}}}}$, dla $n = 2k$. więc szereg $\sum_{n=1}^{\infty} \mathbf{P}_d^n(i, i) < \infty$ i dlatego łańcuch nie jest powracający.

Wiedząc, że jest on zbieżny możemy obliczyć przybliżoną wartość sumy dla $d = 3, 4, 5, 6$. Wielkości te wynoszą: 0.35742, 0.11763, 0.046788, 0.020459.

Możemy więc obliczać prawdopodobieństwa powrotu $F(i)$. Wynoszą one odpowiednio: 0.263308, 0.10524, 0.044696, 0.0200488.

17.2 Okresowość i ergodyczność

Rozważmy nieredukowalny łańcuch Markowa i ustalmy pewien jego stan $i \in M$. Określamy:

$$N_i = \{n : \mathbf{P}^n(i, i) > 0\}.$$

Ponieważ i komunikuje się z samym sobą, to $N_i \neq \emptyset$.

Jeżeli $m, n \in N_i$, to także $m + n \in N_i$, gdyż $\mathbf{P}^{m+n}(i, i) \geq \mathbf{P}^m(i, i)\mathbf{P}^n(i, i) > 0$.

Definicja – 17.6. Stan i jest *okresowy*, jeżeli $\nu_i := \text{NWD}(N_i) > 1$. Wtedy liczbę ν_i nazywamy *okresem stanu i* .

Lemat – 17.7. *Dla każdego stanów $i, j \in M$ zachodzi równość $\nu_i = \nu_j$.*

Dowód. Ponieważ łańcuch jest nieredukowalny, to istnieją takie liczby n, m , że $\mathbf{P}^n(i, j) > 0$ oraz $\mathbf{P}^m(j, i) > 0$. W takim razie $\mathbf{P}^{n+m}(i, i) \geq \mathbf{P}^n(i, j)\mathbf{P}^m(j, i) > 0$, a więc $n + m \in N_i$. Podobnie $n + m \in N_j$. Niech $t \in N_i$. Wtedy $\mathbf{P}^{n+t+m}(j, j) \geq \mathbf{P}^n(i, j)\mathbf{P}^t(i, i)\mathbf{P}^m(j, i) > 0$, a więc $n + t + m \in N_j$. Więc ν_j dzieli zarówno $n + t + m$ jak i $n + m$, czyli ν_j dzieli t . Ponieważ ν_i jest największym wspólnym dzielnikiem N_i , to $\nu_j \geq \nu_i$. Podobnie dowodzimy, że $\nu_i \geq \nu_j$. \square

Wykazaliśmy więc, że w nieredukowalnym łańcuchu Markowa: albo wszystkie stany są okresowe i mają wspólny okres, albo żaden ze stanów nie jest okresowy. W pierwszym z tych przypadków mówimy, że łańcuch Markowa jest *okresowy*, a jego *okresem* jest okres każdego jego stanu. W drugim przypadku mówimy, że łańcuch jest *nieokresowy*.

Standardowy spacer losowy po prostej jest okresowy a jego okres wynosi 2. Natomiast spacer losowy, dla którego $p + q < 1$ i który nie posiada ekranów, nie jest okresowy. Nawet istnienie ekranów wraz z warunkiem $p + q = 1$ nie gwarantuje okresowości.

Nieokresowym łańcuchem jest urnowy model Bernoulliego w przypadku $k > 1$, przykład 16.7.

Twierdzenie – 17.8. *Załóżmy, że przestrzeń stanów M nieredukowalnego łańcucha Markowa jest skończona. Wtedy następujące warunki są równoważne.*

- (1) łańcuch jest nieokresowy.
- (2) istnieje takie n_0 , że dla każdego $n \geq n_0$ oraz każdych $i, j \in M$: $\mathbf{P}^n(i, j) > 0$.

Najpierw udowodnimy lemat.

Lemat – 17.9. *Niech $n_1, \dots, n_r \in \mathbb{N}$. Jeżeli $\text{NWD}(n_1, \dots, n_r) = 1$, to*

$$\exists n_0 \forall n \geq n_0 \exists x_1, \dots, x_r \in \mathbb{N} : n = \sum_{i=1}^r x_i n_i.$$

Dowód lematu. Niech $f : \mathbb{Z}^r \ni x \rightarrow \sum_{i=1}^r x_i n_i \in \mathbb{Z}$. Istnieje więc takie $x^0 \in \mathbb{Z}^r$, że $f(x^0)$ jest najmniejszą wartością w zbiorze $\{f(x) : x \in \mathbb{Z}^r, f(x) \geq 1\}$. Zauważmy, że $d = f(x^0)$ dzieli wszystkie liczby $f(x)$ (piszemy $d|f(x)$). Mianowicie, dla ustalonego x mamy $f(x) = kd + \varrho$, $0 \leq \varrho < d$, $k \in \mathbb{Z}$. Wtedy $f(x - kx^0) = f(x) - kf(x^0) = \varrho < f(x^0)$, więc zachodzi $\varrho = 0$. W szczególności $d = f(x^0)$ dzieli wszystkie liczby n_1, \dots, n_r , więc $d = 1$.

Niech $\mathbf{1} = (1, \dots, 1)$ i niech $n \geq 1$. Niech ϱ będzie resztą z dzielenia n przez $f(\mathbf{1})$. Stąd $n = kf(\mathbf{1}) + \varrho = kf(\mathbf{1}) + \varrho f(x^0) = f(k\mathbf{1} + \varrho x^0)$. Gdy n dąży do nieskończoności, także k dąży do nieskończoności, a dla dużych k wektor $k\mathbf{1} + \varrho x^0 \in \mathbb{N}^r$. \square

Dowód twierdzenia 17.8. (1) \implies (2). Ustalmy stany $i, j \in M$. Ponieważ i nie jest okresowy istnieją takie liczby $n_1, \dots, n_r \in N_i$, że $\text{NWD}(n_1, \dots, n_r) = 1$. Z udowodnionego lematu i z własności zbioru N_i wnioskujemy, że wszystkie dostatecznie duże $n \in N_i$, czyli dla takich n mamy $\mathbf{P}^n(i, i) > 0$. Niech $k = k(i, j) \geq 1$ będzie takie, że $\mathbf{P}^k(i, j) > 0$. Wtedy $\mathbf{P}^{n+k}(i, j) \geq \mathbf{P}^n(i, i)\mathbf{P}^k(i, j) > 0$ dla odpowiednio dużych n , powiedzmy dla $n \geq n(i, j)$. Ponieważ jednak zbiór stanów M jest skończony, to biorąc $n_0 = \max\{n(i, j) : i, j \in M\}$ mamy tezę.

(2) \implies (1). Oczywiście. \square

W pewnych okolicznościach możemy być zainteresowani w zachowaniu się łańcucha Markowa po upływie długiego czasu. W szczególności, warto się pytać o asymptotyczny rozkład wektorów X_n . Poniższe twierdzenie opisuje właśnie taką sytuację w najprostszym szczególnym przypadku. Znane są jednak wyniki dużo ogólniejsze.

Twierdzenie – 17.10 (Twierdzenie ergodyczne). *Rozważamy nieredukowalny łańcuch Markowa określony na skończonej przestrzeni stanów M , $\#M = k$. Jeżeli łańcuch jest nieokresowy, to istnieje wektor π o współrzędnych π_1, \dots, π_k taki, że*

1. $\pi_i > 0$ dla wszystkich $i \in M$,
2. Dla każdych $i, j \in M$,

$$\lim_{n \rightarrow \infty} \mathbf{P}^n(i, j) = \pi_j.$$

Wtedy też

3. Wektor π jest jedynym rozwiązaniem równania

$$\mathbf{P}^T x = x$$

spełniającym warunek $\sum_{i \in M} x_i = 1$.

Wektor ten nazywa się rozkładem stacjonarnym.

4. Dla każdego $i \in M$ $\lim_{n \rightarrow \infty} P(X_n = i) = \pi_i$.

Ergodyczność. Jeżeli łańcuch jest nieokresowy, to dla dużych n prawdopodobieństwo przejścia ze stanu i do stanu j w n krokach jest dodatnie i faktycznie zależy od stanu końcowego j oraz nie zależy od stanu początkowego i . Prawdopodobieństwa te można otrzymać, rozwiązując odpowiedni układ równań liniowych. Taka lub podobna własność nazywa się ergodycznością. Często wtedy mówimy, że łańcuch jest ergodyczny.

Dowód. Wykorzystamy wnioski z twierdzenia Frobeniusa–Perrona¹.

Zauważmy jednak najpierw, że macierz przejścia dowolnego łańcucha Markowa ma wartość własną $\lambda_1 = 1$, a wektorem własnym macierzy odpowiadającym λ_1 jest $\mathbf{1} = (1, \dots, 1)^T$, co łatwo wynika z faktu, że $\sum_{j \in M} \mathbf{P}(i, j) = 1$ dla każdego $i \in M$ (ćwiczenie). Oczywiście λ_1 jest także wartością własną macierzy \mathbf{P}^T , a więc istnieje taki niezerowy wektor π , $\mathbf{P}^T \pi = \pi$.

Ponieważ łańcuch jest nieokresowy, więc na podstawie twierdzenia 17.8 wiemy, że istnieje potęga \mathbf{P}^n macierzy \mathbf{P} mająca wszystkie wyrazy dodatnie. Twierdzenie Frobeniusa–Perrona gwarantuje wtedy, że wszystkie pozostałe wartości własne macierzy \mathbf{P} są na moduł mniejsze niż 1 oraz, że wektor π ma wszystkie współrzędne dodatnie. Można założyć (dzieląc przez odpowiednią stałą), że $\pi_1 + \dots + \pi_k = 1$. Wiadomo także, że wtedy:

$$\lim_{n \rightarrow \infty} \mathbf{P}^n = \frac{\mathbf{1}\pi^T}{\pi^T \mathbf{1}} = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \cdots & \pi_k \\ \pi_1 & \pi_2 & \cdots & \cdots & \pi_k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \cdots & \cdots & \pi_k \end{bmatrix},$$

co oznacza, że $\lim_{n \rightarrow \infty} \mathbf{P}^n(i, j) = \pi_j$.

$$P(X_n = i) = \mathbf{p}_n(i) = \sum_{j=1}^k \mathbf{P}^n(j, i) \mathbf{p}(j) \rightarrow \sum_{j=1}^k \pi_j \mathbf{p}(j) = \pi_i \text{ dla } n \rightarrow \infty. \quad \square$$

Przykład – 17.11. Syzyf wtacza kamień na górę wysokości 4000 m. Pod koniec każdego dnia, z równymi prawdopodobieństwami: udało mu się pokonać kolejne 1000 m, lub kamień wysunął mu się z rąk i stoczył do stóp góry. Jednakże w chwili, gdy ma już dotrzeć do szczytu, złośliwy Zeus zawsze strąca mu kamień na dół. Oszacować prawdopodobieństwo, że po 10 000 dniach pracy Syzyf będzie znajdował się dokładnie w połowie góry.

Pracę Syzyfa opisuje łańcuch Markowa X_n przyjmujący wartości w przestrzeni stanów $M = \{0, 1, 2, 3\}$. Początkowy rozkład zmiennej X_0 , \mathbf{p}_0 , jest jednopunktowy $= \delta_0$. Natomiast macierz przejścia ma postać:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Jak widać, każde dwa stany się komunikują; z dowolnego stanu można z dodatnim prawdopodobieństwem przejść do każdego innego stanu. Widać też, że stan 0 nie jest okresowy (bo można w nim pozostać), więc łańcuch nie jest okresowy.

Zachodzi więc warunek 2. w twierdzeniu ergodycznym. Należy więc teraz rozwiązać układ równań liniowych: 5 równań o 4 niewiadomych.

$$\mathbf{P}^T x = x, \quad x_1 + x_2 + x_3 + x_4 = 1,$$

mamy pewność, że ma on rozwiązanie i to dokładnie jedno. Łatwo się przekonać, że jest nim układ:

$$x_1 = \frac{8}{15}, \quad x_2 = \frac{4}{15}, \quad x_3 = \frac{2}{15}, \quad x_4 = \frac{1}{15}.$$

To oznacza, że dla dużych t , w szczególności dla $t = 10\,000$, macierz \mathbf{P}^t ma w przybliżeniu postać.

$$\mathbf{P}^t \cong \begin{bmatrix} \frac{8}{15} & \frac{4}{15} & \frac{2}{15} & \frac{1}{15} \\ \frac{8}{15} & \frac{4}{15} & \frac{2}{15} & \frac{1}{15} \\ \frac{8}{15} & \frac{4}{15} & \frac{2}{15} & \frac{1}{15} \\ \frac{8}{15} & \frac{4}{15} & \frac{2}{15} & \frac{1}{15} \end{bmatrix}.$$

¹K. Wójcik, Stosowana algebra liniowa, UJ 2018/2019, wykłady 8 – 10

Tak więc odpowiedź na pytanie brzmi: $P^{10000}(i, 2) \approx \frac{2}{15}$.

Zgodnie z twierdzeniem Frobeniusa–Perrona macierz przejścia ergodycznego łańcucha Markowa ma wartość własną równą 1, a pozostałe wartości własne mają moduły mniejsze od 1.

Na przykład, dla powyższej macierzy wynoszą one: $1, -\frac{1}{2}, \frac{i}{2}, -\frac{i}{2}$.

Dla modelu urnowego Bernoulliego, przykład 16.7:

$k = 4$: $1, -\frac{1}{8}, \frac{1}{2}, \frac{1}{8}, -\frac{1}{4}$.

$k = 10$: $1, 1/10, 1/5, 31/50, 23/50, -2/25, 1/50, -1/25, 4/5, -1/10, 8/25$.

Dla łańcucha (nie jest nieredukowalny) z przykładu 16.8:

$-.149, .949, .901, -.101, -.0234, .823, .332, .468, .199, .601, .721, .0787, 1., 1.$

Przykład – 17.12 (Kontynuacja przykładów 3.2 oraz 16.2.). Wyraźnie widać, że łańcuch określony w przykładzie 16.2 jest ergodyczny, a łatwo sprawdzić, że jego stan stacjonarny wynosi:

$$\pi = \begin{bmatrix} 1/5 \\ 3/5 \\ 1/5 \end{bmatrix}.$$

Tak więc po dłuższym okresie obowiązywania umowy Leon będzie sprzątał w danym dniu z prawdopodobieństwem bliskim $\frac{3}{5}$, zobacz też symulacje w Ćwiczeniu 3.1.

Powracanie i ergodyczność. Gdy spełnione są założenia twierdzenia ergodycznego to dla każdego $i \in M$ mamy $\mathbf{P}^n(i, i) \approx \pi_i > 0$ dla dużych n , więc $\mathbf{P}(i) = \sum_{i=1}^{\infty} \mathbf{P}^n(i, i) = \infty$, więc na podstawie twierdzenia 17.2 łańcuch jest powracający.

Oznacza to też, że zmienne losowe τ_i określające liczbę powrotów do stanu i mają nieskończone wartości oczekiwane.

17.3 Łańcuch Markowa jako graf

Łańcuch Markowa można opisać za pomocą **grafu skierowanego** (digrafu) z wagami.

Niektóre spacery losowe po **grafie (nieskierowanym)** będące łańcuchami Markowa mają ciekawą interpretację, a także znaczenie praktyczne.

Definicja – 17.13. Dany jest zbiór skończony $V = \{v_1, \dots, v_d\}$.

1. *Grafem* nazywamy parę (V, E) , przy czym E jest pewnym zbiorem dwuelementowych **podzbiorów** utworzonych z elementów zbioru V .

2. *Grafem skierowanym* nazywamy parę (V, E) , przy czym E jest pewnym zbiorem **par** utworzonych z elementów zbioru V ($E \subset V \times V$).

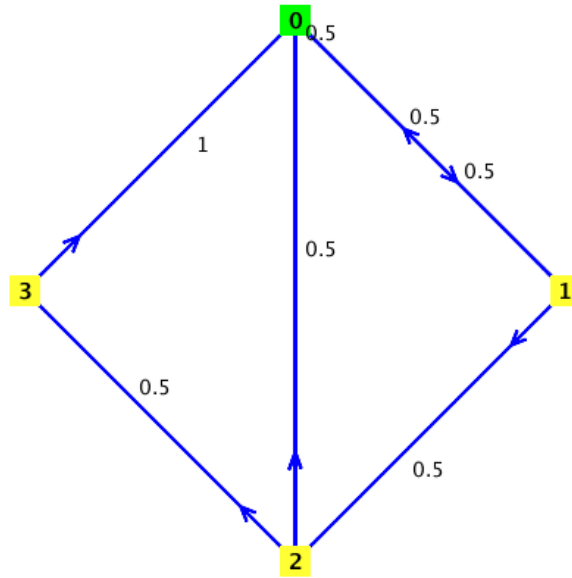
3. *Grafem z wagami* nazywamy parę (V, W) , przy czym $W : V \times V \rightarrow \mathbb{R}$ (W jest pewną macierzą indeksowaną przez elementy V).

Elementy zbioru V nazywamy *wierzchołkami*, elementy zbioru E nazywamy *krawędziami*. Każdy graf z wagami (V, W) indukuje graf skierowany (V, E) , przy czym $E = \{(v_i, v_j) : A(v_i, v_j) \neq 0\}$. Wtedy elementy $W(v_i, v_j)$ nazywamy wagami krawędzi (v_i, v_j) . Najczęściej przy zapisie krawędzi opuszczamy nawiasy. Czyli $uv = (u, v)$ w przypadku grafu skierowanego, $uv = \{u, v\}$ w przypadku grafu.

Bardzo często grafy interpretuje się w sposób geometryczny, co ma mocne uzasadnienie intuicyjne i w wielu sytuacjach ułatwia ich analizę. Niemniej jest to pojęcie mające niekiedy daleko szerszą interpretację.

Uwaga – 17.14. Jeżeli $\{X_n\}$ jest łańcuchem Markowa na przestrzeni stanów M i ma macierz przejścia \mathbf{P} , to (M, \mathbf{P}) jest grafem z wagami.

Przykład – 17.15. W przykładzie 17.11 (o Syzyfie) graf z wagami może wyglądać tak:

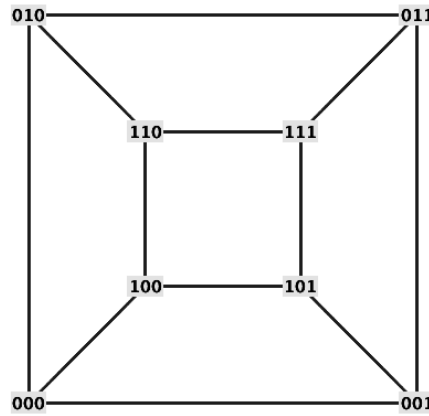


d-wymiarowa kostka dyskretna jako graf. Niech *d* będzie ustaloną liczbą naturalną. Określamy

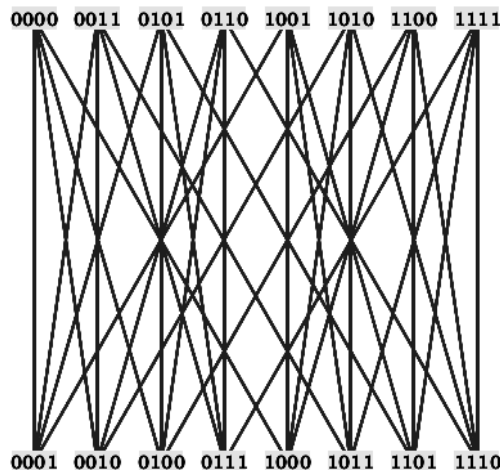
$$V = \{0, 1\}^d = \{(\varepsilon_1, \dots, \varepsilon_d) : \varepsilon_i \in \{0, 1\}, i = 1, \dots, d\}.$$

Niech *E* będzie zbiorem wszystkich takich dwuelementowych zbiorów $\{u, v\}$, $u, v \in V$, że *u*, *v* różnią się od siebie dokładnie na jednym miejscu.

Dla $d = 2$ mamy więc cztery wierzchołki i cztery krawędzie, które możemy interpretować jako boki kwadratu. Dla $d = 3$ mamy 8 wierzchołków i 12 krawędzi, które możemy interpretować jako krawędzie sześcianu.



Dla $d \geq 3$ liczba krawędzi gwałtownie wzrasta.



17.4 Spacery losowe po grafie (nieskierowanym)

Niech $G = (V, E)$ będzie grafem. Dla wierzchołka v określamy jego stopień $\deg(v)$ jako liczbę krawędzi wychodzących z v , czyli liczbę sąsiadów v .

Spacer losowy $\{X_t\}$ po grafie G .

Cząstka, która w chwili t znajduje się w wierzchołku v , czyli $X_t = v$, może przejść w jednym kroku do jednego z sąsiednich wierzchołków, X_{t+1} , z prawdopodobieństwem $\frac{1}{\deg(v)}$. Gdy wierzchołek nie ma sąsiadów, cząstka z prawdopodobieństwem 1 pozostaje w tym wierzchołku.

Następujące dwa twierdzenia są oczywiste.

Twierdzenie – 17.16. $\{X_t\}$ jest łańcuchem Markowa na przestrzeni stanów V z macierzą przejścia \mathbf{P} określoną następująco: $\mathbf{P}(u, v) = \frac{1}{\deg(u)}$, gdy $uv \in E$, $\mathbf{P}(v, v) = 1$, gdy v nie ma sąsiadów, $\mathbf{P}(u, v) = 0$ w pozostałych przypadkach.

Gdy każde dwa wierzchołki można połączyć ciągiem krawędzi (mówimy wtedy, że graf jest spójny), to $\{X_t\}$ jest nieredukowalny.

Twierdzenie – 17.17. Spacer losowy po d -wymiarowej kostce $\{0, 1\}^d$ jest nieredukowalny i ma okres 2.

Twierdzenie – 17.18. Niech $\{X_t\}$ oznacza taki spacer losowy po grafie $G = (V, E)$, $\#V = d$, że dla każdego $v \in V$ $\deg(v) > 0$ (nie ma wierzchołków izolowanych).

Określamy $\pi_v = \frac{\deg(v)}{C}$ dla $v \in V$, przy czym $C = \sum_{u \in V} \deg(u)$.

Wtedy wektor $\pi \in \mathbb{R}^d$ o współrzędnych π_v jest rozkładem stacjonarnym łańcucha Markowa X_t , to znaczy:

$$\mathbf{P}^T \pi = \pi, \quad \sum_{v \in V} \pi_v = 1, \quad \pi_v > 0 \text{ dla } v \in V.$$

Dowód. Wystarczy pokazać, że $\sum_{u \in V} \mathbf{P}(u, v) \pi_u = \pi_v$ dla każdego $v \in V$. Pamiętając, że $\deg(v)$ oznacza $\#N(v)$, przy czym $N(v)$ jest zbiorem sąsiadów wierzchołka v , i korzystając z określenia łańcucha, widzimy, że:

$$\sum_{u \in V} \mathbf{P}(u, v) \pi_u = \sum_{u \in N(v)} \mathbf{P}(u, v) \pi_u = \sum_{u \in N(v)} \frac{1}{\deg(u)} \frac{\deg(u)}{C} = \frac{\deg(v)}{C} = \pi_v. \quad \square$$

17.5 Markowskie metody Monte Carlo (MCMC)

Idea metod Monte Carlo opartych na teorii łańcuchów Markowa (*Markov Chain Monte Carlo*), w skrócie MCMC, polega na tym, że chcąc losować liczby według ustalonego rozkładu dyskretnego $(\{x_i\}, \{p_i\})$, budujemy taki ergodyczny łańcuch Markowa na przestrzeni stanów, którymi są punkty x_i , że jego stan stacjonarny π ma współrzędne p_i . Symulując zachowanie się tego łańcucha, możemy mieć niemal pewność, że po odpowiedniej liczbie kroków uzyskiwane elementy będą pochodzić z rozkładu π . Największym problemem jest tutaj dobranie odpowiedniego łańcucha. Dużą pomoc stanowi tu teoria grafów.

Jeżeli graf, po którym odbywa się spacer, jest skończony, spójny, a łańcuch Markowa jest nieokresowy, to na podstawie twierdzenia ergodycznego istnieje dokładnie jeden stan stacjonarny, a więc musi to być ten sam stan, co w twierdzeniu 17.18 udowodnionym powyżej. Jeżeli więc dla danego rozkładu $(\{x_i\}, \{p_i\})$ potrafimy zbudować taki graf o wierzchołkach x_i , że określone w twierdzeniu liczby π_i są równe liczbom p_i , to podczas symulowania spaceru losowego będziemy po pewnym czasie otrzymywać punkty z rozkładu $(\{x_i\}, \{p_i\})$. Omówimy zagadnienie, tak zwany *problem plecakowy*, gdzie taka metoda może z powodzeniem być zastosowana.

Zagadnienie plecakowe stanowi klasyczny problem optymalizacji dyskretnej, który w sensie informatycznym jest równoważny wielu innym klasycznym zagadnieniom.

Mamy d przedmiotów, powiedzmy $1, 2, \dots, d$, i znamy wartość każdego z nich, powiedzmy c_1, \dots, c_d , oraz ich wagi, powiedzmy w_1, \dots, w_d . Chcemy tak załadować plecak, aby wartość zabranych przedmiotów była największa.

Problem w tym, że waga plecaka nie może przekroczyć ustalonej liczby B .

Matematyczny model:

$$V = \left\{ v = (v_1, \dots, v_d) \in \{0, 1\}^d : \sum_{i=1}^d v_i w_i \leq B \right\}$$

oraz niech $f(v) = \sum_{i=1}^d v_i c_i$ dla $v \in V$. Szukamy $\arg \max f$.

W dalszym ciągu będziemy zakładać, że zadanie jest nietrywialne, czyli że $\sum_{i=1}^d w_i > B$.

Stosujemy algorytm PRS, patrz punkt 12.4. W tym celu chcemy generować punkty ze zbioru V według rozkładu jednostajnego. To może być bardzo duży zbiór i trudno jest nawet ustalić liczbę jego elementów.

Aby jednak wykonać to zadanie, modyfikujemy opisany poprzednio spacer po grafie o wierzchołkach z V , który jest naturalnym zacieśnieniem d -wymiarowej kostki dyskretnej.

W kolejnym kroku przechodzimy ze stanu $v = (v_1, \dots, v_d) \in V$ do stanu sąsiedniego w sposób następujący: losujemy wskaźnik j spośród liczb $1, 2, \dots, d$ zgodnie z rozkładem jednostajnym i dokonujemy zmiany j -tej współrzędnej v_j na $v'_j = 1 - v_j$, dzięki czemu otrzymujemy wierzchołek

$$v' = (v_1, \dots, v_{i-1}, v'_j, v_{j+1}, \dots, v_d).$$

Jeżeli $v' \in V$, to przenosimy się do v' .

Jeżeli $v' \notin V$ (waga plecaka jest za duża), to pozostajemy w v .

Zaczynamy w chwili 0 z pustym plecakiem $(X_0 = 0, \dots, 0) \in V$, generujemy w powyższy sposób kolejne wierzchołki X_t

Otrzymany w ten sposób spacer oznaczamy przez $\{X_t\}$.

Widać, że:

1. X_t jest łańcuchem Markowa.
2. X_t jest nieredukowalny (bo każdy stan komunikuje się ze stanem zerowym).
3. X_t jest nieokresowy, gdyż istnieją stany v , w których spacer z dodatnim prawdopodobieństwem może przystanąć (gdy $v' \notin V$).

Istnieje więc dokładnie jeden rozkład stacjonarny π .

Co więcej, $\{X_t\}$ ma własność symetrii: $\mathbf{P}(u, v) = \mathbf{P}(v, u)$ dla wszystkich $u, v \in V$.

Ale to oznacza, że wektor $\hat{\pi}$ o wszystkich współrzędnych równych $\hat{\pi}_v = \frac{1}{\#V}$ jest rozkładem stacjonarnym Rzeczywiście:

$$\sum_{u \in V} \mathbf{P}(u, v) \frac{1}{\#V} = \sum_{u \in V} \mathbf{P}(v, u) \frac{1}{\#V} = 1 \cdot \frac{1}{\#V} = \frac{1}{\#V}.$$

Z jednoznaczności gwarantowanej przez twierdzenie ergodyczne otrzymujemy, że $\pi = \hat{\pi}$.

Wniosek. Dla dużych t zmienne losowe X_t mają rozkład niemal identyczny co rozkład jednostajny na V .

Opisana metoda jest mało efektywna. Istnieją dużo lepsze (szybsze) metody MCMC rozwiązywania problemu plecakowego oraz podobnych problemów. Wiele z nich opartych jest jednak na podobnej metodologii, co opisana powyżej.

17.6 Pytania

Pytanie 17.1. Definicje stanu powracającego i stanu okresowego można bez żadnych zmian powtórzyć w przypadku dowolnego łańcucha Markowa. Podać przykład takiego łańcucha, który ma jednocześnie: (1) stan powracający i stan niepowracający, (2) stan okresowy i stan nieokresowy. ROZWIĄZANIE

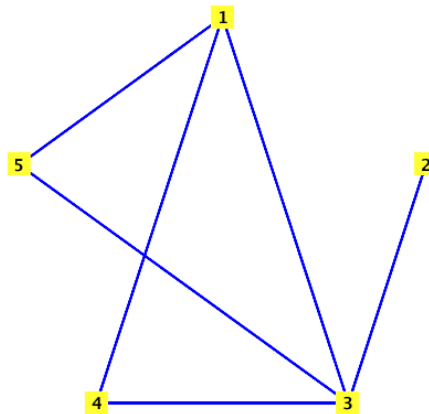
Pytanie 17.2. Podać przykład wskazujący, że twierdzenie 17.8 jest fałszywe w przypadku, gdy przestrzeń stanów nie jest skończona. ROZWIĄZANIE

Pytanie 17.3. Udowodnić, że jeżeli nieredukowalny łańcuch Markowa jest niepowracający, to dla każdego i, j : $\lim_{n \rightarrow \infty} \mathbf{P}^n(i, j) = 0$. ROZWIĄZANIE

Pytanie 17.4. Rozważmy spacer losowy po prostej z barierami w punktach $A = -1, B = 1, p = q = 1/4, s_A = 9/10, s_B = 1/10$. Znaleźć rozkład stacjonarny. ROZWIĄZANIE

Pytanie 17.5. Rozważmy urnowy model Bernoulliego, Przykład 16.7. Niech $k = 3$, $b_0 = 3$. Wyznaczyć oczekiwaną liczbę czerwonych kul w drugiej urnie po pierwszym, drugim, dziesiątym i trzydziestym losowaniu. ROZWIĄZANIE

Pytanie 17.6. Wyznaczyć macierz przejścia spaceru losowego po grafie.



- (1) Czy łańcuch jest okresowy?
- (2) Narysować graf skierowany tego łańcucha.
- (3) Wskazać rozkład stacjonarny tego łańcucha.

ROZWIĄZANIE

Rozdział 18

Rozkłady wielowymiarowe

W poprzednich rozdziałach mieliśmy już, raczej incydentalnie, do czynienia z wektorami losowymi i ich rozkładami. Obecnie zajmujemy się tą sprawą bardziej dokładnie. Głównym celem jest omówienie podstawowych własności rozkładu normalnego. Wspomniemy też o problemie regresji, który w przypadku rozkładów anormalnych sprowadza się do projektu regresji liniowej.

18.1 Macierze symetryczne dodatnio określone

Nasze rozważania poprzedzimy przypomnieniem przydatnych faktów z algebry liniowej.

Niech $M(m, n)$ oznacza zbiór macierzy rzeczywistych o m wierszach i n kolumnach.

Rozważamy macierze kwadratowe stopnia n , $M(n) = M(n, n)$.

Macierz $C \in M(n)$ jest symetryczna, gdy $C^T = C$.

Macierz symetryczna C jest nieujemnie (dodatnio) określona, gdy

$$\forall x \in \mathbb{R}^n \setminus \{0\} : x^T C x = \sum_{i,j} c_{ij} x_i x_j \geq 0 \quad (x^T C x > 0).$$

Piszemy wtedy $C \geq 0$ ($C > 0$).

$$\sigma(C) = \{\lambda \in \mathbb{C} : \exists z \in \mathbb{C}^n \setminus \{0\} : Cz = \lambda z\}.$$

jest zbiorem wszystkich wartości własnych macierzy C .

Łatwo wykazać:

Twierdzenie – 18.1. *Jeżeli $C \in M(n)$, $C = C^T$, to $\sigma(C) \subset \mathbb{R}$.*

Jednym z ważniejszych twierdzeń leżących u podstaw matematyki stosowanej jest:

Twierdzenie – 18.2 (o diagonalizacji macierzy symetrycznej). *Jeżeli $C \in M(n)$, $C = C^T$, to istnieje macierz $P \in M(n)$, $PP^T = I$, taka że $C = P \operatorname{diag}(\lambda_1, \dots, \lambda_n) P^T$, gdzie $\lambda_1, \dots, \lambda_n \in \mathbb{R}$.*

Wtedy

$$\sigma(C) = \{\lambda_1, \dots, \lambda_n\}.$$

Łatwo widać, że jeżeli $C \geq 0$ ($C > 0$), to również $\operatorname{diag}(\lambda_1, \dots, \lambda_n) \geq 0$ ($\operatorname{diag}(\lambda_1, \dots, \lambda_n) > 0$), co z kolei jest równoważne temu, że wszystkie wartości własne λ_i są nieujemne (dodatnie).

Wniosek – 18.3. *Jeżeli $C \geq 0$, to $\det C \geq 0$. Ponadto następujące warunki są równoważne:*

1. $C > 0$,
2. $\det C > 0$,
3. C jest nieosobliwa.

W takich przypadkach można rozważać macierz $B = P \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}) P^T$. Łatwo sprawdzić, że $B^2 = C$. Będziemy używać oznaczenia $C^{\frac{1}{2}} := B$.

18.2 Nadzieja matematyczna i macierz kowariancji

Niech $X : \Omega \rightarrow \mathbb{R}^n$ będzie wektorem losowym, czyli

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix},$$

gdzie X_i są zmiennymi losowymi.

Definicja – 18.4. Określamy *nadzieję matematyczną* oraz *kowariancję wektora losowego*:

$$\mu = E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{bmatrix}, \quad \Sigma = \text{cov}(X) = \begin{bmatrix} \text{cov}(X_1, X_1) & \dots & \text{cov}(X_1, X_n) \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ \text{cov}(X_n, X_1) & \dots & \text{cov}(X_n, X_n) \end{bmatrix}.$$

Przypominamy: $\text{cov}(X_i, X_j) = E((X_i - E(X_i)) \cdot (X_j - E(X_j)))$. Mamy więc natychmiast:

$$\Sigma = \Sigma^T.$$

Twierdzenie – 18.5. Jeżeli $A \in M(m, n)$, $b \in \mathbb{R}^m$, $W = AX + b$, to

$$E(W) = AE(X) + b, \quad \text{cov}(W) = A \text{cov}(X) A^T.$$

Dowód. (ćwiczenie). □

Twierdzenie – 18.6. Dla każdego $a \in \mathbb{R}^n$ mamy $a^T \text{cov}(X) a \geq 0$, czyli $\text{cov}(X) \geq 0$.

Dowód. Niech $a \in \mathbb{R}^n$. Określamy zmienną losową: $W = a^T X$. Jej wariancja $D^2(W) \geq 0$. Ale: $D^2(W) = \text{cov}(W) = a^T \text{cov}(X) a$. □

Twierdzenie – 18.7. Jeżeli X ma rozkład ciągły, to dla każdego $a \in \mathbb{R}^n \setminus \{0\}$ zachodzi nierówność $a^T \text{cov}(X) a > 0$, czyli $\text{cov}(X) > 0$.

Dowód. Nie wprost.

Niech $a \in \mathbb{R}^n \setminus \{0\}$ będzie takie, że $a^T \text{cov}(X) a = 0$. Niech jak poprzednio $W = a^T X$.

Ponieważ $D^2(W) = 0$, to W jest stała pw. Istnieje więc takie $c \in \mathbb{R}$, że $P(W = c) = 1$. Wtedy jednak $P(a^T X = c) = 1$. Stąd $P(X \in M) = 1$, gdzie $M = \{y \in \mathbb{R}^n : a^T y = c\}$ jest przestrzenią afiniczną, $\dim M = n - 1$. Ale to prawdopodobieństwo można też inaczej obliczyć, gdyż X ma gęstość, powiedzmy $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Mianowicie:

$$P(X \in M) = \int_M f dx = 0,$$

gdź miara Lebesgue'a każdej właściwej podprzestrzeni afinicznej równa się zero. □

18.3 Funkcje generujące momenty.

Pamiętamy, że dla każdego rozkładu, powiedzmy Q , istnieje funkcja charakterystyczna $h_Q : \mathbb{R} \rightarrow \mathbb{C}$:

$$h_Q(u) = \int_{\mathbb{R}} e^{iux} dQ(x).$$

Definicja – 18.8. Niech Q będzie n -wymiarowym rozkładem. Funkcję $M_Q : \mathbb{R}^n \rightarrow \mathbb{R}$ określoną wzorem

$$M_Q(t) = \int_{\mathbb{R}^n} e^{t^T y} dQ(y).$$

nazywamy funkcją generującą momenty.

Niech X będzie wektorem losowym. Określamy

$$M_X(t) = M_{P_X}(t) = E\left(e^{t^T X}\right).$$

Ponieważ dla niektórych $t \in \mathbb{R}^n$ wartość $E\left(e^{t^T X}\right)$ może być nieskończona, funkcja generująca momenty może nie istnieć! Jest to jej podstawowa wada w porównaniu do funkcji charakterystycznych.

Podajemy bez dowodu następujące twierdzenie będące odpowiednikiem twierdzenia 11.15.

Twierdzenie – 18.9. *Jeżeli $M_{Q_1}(t) = M_{Q_2}(t)$ dla $t \in G$, gdzie $G \subset \mathbb{R}^n$ jest zbiorem otwartym, to $Q_1 = Q_2$.*

Łatwo udowodnić (ćwiczenie) następujące:

Twierdzenie – 18.10. *Jeżeli $A \in M(m, n)$, $b \in \mathbb{R}^m$, $W = AX + b$, to*

$$M_W(t) = e^{b^T t} M_X(A^T t).$$

Twierdzenie – 18.11. *Niech $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$. Wtedy wektory losowe X_1, X_2 są niezależne, wtedy i tylko wtedy, gdy*

$$\forall t = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} : M_X\left(\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}\right) = M_{X_1}(t_1) \cdot M_{X_2}(t_2).$$

Dowód. „ \implies ” Z niezależności X_1, X_2 wynika niezależność $e^{t_1^T X_1}, e^{t_2^T X_2}$ dla wszystkich t_1, t_2 . Stąd:

$$M_X(t) = E\left(e^{t^T X}\right) = E\left(e^{t_1^T X_1 + t_2^T X_2}\right) = E\left(e^{t_1^T X_1} e^{t_2^T X_2}\right) = E\left(e^{t_1^T X_1}\right) E\left(e^{t_2^T X_2}\right) = M_{X_1}(t_1) M_{X_2}(t_2).$$

„ \impliedby ” Chcemy pokazać, że: $P_X = P_{X_1} \times P_{X_2}$. Wystarczy więc pokazać, że $M_{P_X} = M_{P_{X_1} \times P_{X_2}}$. Dla $t \in \mathbb{R}^n$ mamy:

$$\begin{aligned} M_{P_{X_1} \times P_{X_2}}(t) &= \int_{\mathbb{R}^n} e^{t^T y} d(P_{X_1} \times P_{X_2})(y) = \int_{\mathbb{R}^n} e^{t_1^T y_1 + t_2^T y_2} d(P_{X_1} \times P_{X_2})(y_1, y_2) \\ &= \int_{\mathbb{R}^{n_1}} e^{t_1^T y_1} dP_{X_1}(y_1) \int_{\mathbb{R}^{n_2}} e^{t_2^T y_2} dP_{X_2}(y_2) = M_{P_{X_1}}(t_1) M_{P_{X_2}}(t_2) = M_{P_X}(t). \quad \square \end{aligned}$$

Przykład – 18.12 (funkcja generująca momenty dla rozkładu $N(0, 1)$). Niech Z będzie zmienną losową o rozkładzie $N(0, 1)$.

$$\begin{aligned} M_Z(t) &= E(e^{tZ}) = \int_{\mathbb{R}} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{tz - \frac{1}{2}z^2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(z^2 - 2tz + t^2)} e^{\frac{1}{2}t^2} dz = e^{\frac{1}{2}t^2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(z-t)^2} dz = e^{\frac{1}{2}t^2}. \end{aligned}$$

Przykład – 18.13 (funkcja generująca momenty dla i.i.d. $\sim N(0, 1)$). Niech (Z_1, \dots, Z_n) będzie ciągiem i.i.d. o wspólnym rozkładzie $N(0, 1)$. Obliczymy M_Z dla:

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix},$$

Na podstawie poprzedniego twierdzenia:

$$M_Z(t) = M_{Z_1}(t_1) \cdot \dots \cdot M_{Z_n}(t_n) = e^{\frac{1}{2}t_1^2} \cdot \dots \cdot e^{\frac{1}{2}t_n^2} = e^{\frac{1}{2}(t_1^2 + \dots + t_n^2)} = e^{\frac{1}{2}\|t\|^2} = e^{\frac{1}{2}t^T t}.$$

Przy okazji zauważmy, że: $E(Z) = 0$, $\text{cov}(Z) = I_n$ – macierz identycznościowa.

Przykład – 18.14. Rozważmy wektor $X = AZ + \mu$, gdzie wektor losowy Z jest taki, jak w poprzednim przykładzie, $A \in M(m, n)$ oraz $\mu \in \mathbb{R}^m$. Policzmy jego nadzieję matematyczną, kowariancję oraz funkcję generującą momenty. Ponieważ $E(Z) = 0$, $\text{cov}(Z) = I_n$, więc $E(X) = \mu$, $\Sigma = \text{cov}(X) = AA^T$. Mamy

$$M_X(t) = e^{\mu^T t} M_Z(A^T t) = e^{\mu^T t} e^{\frac{1}{2}(A^T t)^T A^T t} = e^{\mu^T t + \frac{1}{2}t^T AA^T t} = e^{\mu^T t + \frac{1}{2}t^T \Sigma t}.$$

Powyższy przykład okaże się punktem wyjścia do definicji wielowymiarowego rozkładu normalnego.

18.4 Pytania

Pytanie 18.1. Znaleźć $\Sigma^{\frac{1}{2}}$, gdy $\Sigma = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. ROZWIĄZANIE

Pytanie 18.2. Udowodnić twierdzenie 18.5. ROZWIĄZANIE

Pytanie 18.3. Udowodnić twierdzenie 18.10. ROZWIĄZANIE

Pytanie 18.4. Wyznaczyć funkcję generującą momenty dla rozkładu Poissona P_λ . ROZWIĄZANIE

Pytanie 18.5. Znaleźć macierz kowariancji wektora losowego X mającego rozkład o gęstości:

(a) $f(x, y) = ce^{-x^2-4y^2}$, gdzie c jest odpowiednio dobraną stałą.

(b) $f(x, y) = ce^{-x^2-4y^2+2xy-2x}$, gdzie c jest odpowiednio dobraną stałą. ROZWIĄZANIE

Pytanie 18.6. Znaleźć macierz kowariancji wektora losowego X mającego rozkład o gęstości $f = c \cdot g \cdot \mathbf{1}_K$, gdzie K jest trójkątem o wierzchołkach $(0, 0)$, $(0, 1)$, $(1, 1)$, $g(x, y) = x^2$, a c jest odpowiednią stałą. ROZWIĄZANIE

Rozdział 19

Rozkład normalny wielowymiarowy

19.1 Definicja i własności

Przykład 18.14 stanowi motywacją do następującej definicji.

Definicja – 19.1 (Rozkład normalny $N_n(\mu, \Sigma)$). Dane są: $\mu \in \mathbb{R}^n$ oraz $\Sigma \in M(n)$, $\Sigma = \Sigma^T$, $\Sigma \geq 0$.

Rozkład Q nazywamy normalnym i oznaczamy $N_n(\mu, \Sigma)$, jeżeli

$$M_Q(t) = e^{\mu^T t + \frac{1}{2} t^T \Sigma t}.$$

Poprzedni przykład pokazuje, że dla dowolnych μ oraz symetrycznej $\Sigma \geq 0$ istnieje taki rozkład. Wystarczy wziąć $A = \Sigma^{\frac{1}{2}}$. Wtedy rozkład $P_X = N_n(\mu, \Sigma)$.

Uwaga – 19.2. W przypadku $n = 1$, $\Sigma = \sigma^2 > 0$. Wtedy $N_1(\mu, \sigma^2) = N(\mu, \sigma)$.

Ponadto $N_1(\mu, 0) = \delta_\mu$.

Podana poprzednio definicja rozkładu normalnego może wydawać się trochę nienaturalna, niemniej ma szereg korzyści. Jedną z nich jest następujące:

Twierdzenie – 19.3. (1) Jeżeli wektor losowy X ma rozkład normalny $N_n(\mu, \Sigma)$, to

$$E(X) = \mu, \quad \text{cov}(X) = \Sigma.$$

(2) Jeżeli ponadto $W = AX + b$, $A \in M(m, n)$, $b \in \mathbb{R}^m$, to W ma rozkład

$$N_m(A\mu + b, A\Sigma A^T).$$

Dowód. Ad (1). Jakikolwiek wektor losowy o rozkładzie $N_n(\mu, \Sigma)$ ma taki sam rozkład jak wektor X otrzymany w przykładzie 18.14, a więc ich parametry są takie same. \square

Ad (2). $M_W(t) = e^{b^T t} M_X(A^T t) = e^{b^T t} e^{\mu^T A^T t + \frac{1}{2} (A^T t)^T \Sigma A^T t} = e^{(A\mu + b)^T t + \frac{1}{2} t^T A \Sigma A^T t}$. $M_W(t)$ jest więc funkcją generującą momenty dla rozkładu $N_m(A\mu + b, A\Sigma A^T)$. \square

Twierdzenie – 19.4. Niech wektor losowy X ma rozkład normalny $N_n(\mu, \Sigma)$,

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Wtedy:

(1) X_i ma rozkład normalny $N_{n_i}(\mu_i, \Sigma_{ii})$, $i = 1, 2$

(2) X_1, X_2 są niezależne $\iff \Sigma_{12} = \Sigma_{21}^T = 0$.

Dowód. Ad (1) $X_1 = [I_{n_1}, 0]X$, $X_2 = [0, I_{n_2}, 0]X$, gdzie $I_{n_i} \in M(n_i)$ jest macierzą identycznościową. Stosujemy poprzednie twierdzenie.

Ad (2)

$$\begin{aligned} M_X(t) &= M_X \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \exp \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}^T \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} + \frac{1}{2} [t_1^T, t_2^T] \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \right) \\ &= \exp \left(\mu_1^T t_1 + \mu_2^T t_2 + \frac{1}{2} [t_1^T, t_2^T] \begin{bmatrix} \Sigma_{11} t_1 + \Sigma_{12} t_2 \\ \Sigma_{21} t_1 + \Sigma_{22} t_2 \end{bmatrix} \right) \\ &= \exp \left(\mu_1^T t_1 + \mu_2^T t_2 + \frac{1}{2} (t_1^T \Sigma_{11} t_1 + t_1^T \Sigma_{12} t_2 + t_2^T \Sigma_{21} t_1 + t_2^T \Sigma_{22} t_2) \right). \end{aligned}$$

Z drugiej strony:

$$M_{X_1}(t_1) = \exp \left(\mu_1^T t_1 + \frac{1}{2} t_1^T \Sigma_{11} t_1 \right), \quad M_{X_2}(t_2) = \exp \left(\mu_2^T t_2 + \frac{1}{2} t_2^T \Sigma_{22} t_2 \right).$$

Widać, że warunkiem równoważnym warunkowi $M_X(t) = M_{X_1}(t_1)M_{X_2}(t_2)$ jest:

$$\forall t_1, t_2 : t_1^T \Sigma_{12} t_2 + t_2^T \Sigma_{21} t_1 = 0.$$

Po lewej stronie występują dwie liczby, które są sobie równe, gdyż $\Sigma_{12}^T = \Sigma_{21}$.

A więc ostatni warunek brzmi:

$$\forall t_1, t_2 : t_2^T \Sigma_{21} t_1 = 0,$$

co z kolei jest równoważne warunkowi $\Sigma_{21} = 0$. □

Twierdzenie – 19.5 (o istnieniu gęstości). *Niech wektor losowy X ma rozkład $N_n(\mu, \Sigma)$. Zachodzi równoważność:*

$$X \text{ ma rozkład ciągły} \iff \Sigma \text{ jest nieosobliwa.}$$

Wtedy gęstość f_X wyraża się wzorem:

$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}.$$

W przeciwnym przypadku istnieje taka właściwa podprzestrzeń afiniczna $M \subset \mathbb{R}^n$, że $P(X \in M) = 1$.

Dowód. Twierdzenie udowodnimy dla szczególnie dobranego wektora X mającego rozkład $N_n(\mu, \Sigma)$, co nie zmniejsza ogólności. Niech wektor losowy Z ma współrzędne Z_1, \dots, Z_n , które są i.i.d. o rozkładzie normalnym standardowym każda. Z ma więc gęstość, która jest iloczynem gęstości zmiennych losowych Z_i .

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_1^2} \cdot \dots \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_n^2} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum_{i=1}^n z_i^2} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}z^T z}.$$

Niech $X = \Sigma^{\frac{1}{2}} Z + \mu$. W przypadku, gdy Σ jest nieosobliwa, odwzorowanie $\varphi(z) = \Sigma^{\frac{1}{2}} z + \mu$ jest dyfeomorfizmem, a więc na podstawie twierdzenia 5.33:

$$f_X(x) = \frac{1}{\det \Sigma^{\frac{1}{2}}} f_Z \left((\Sigma^{\frac{1}{2}})^{-1}(x - \mu) \right) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}.$$

Gdy Σ jest osobliwa, także $\Sigma^{\frac{1}{2}}$ jest osobliwa, a więc jej rząd k jest mniejszy niż n . Więc $\dim M = k < n$, gdzie $M = \{\Sigma^{\frac{1}{2}} z + \mu : z \in \mathbb{R}^n\}$ jest przestrzenią afiniczną zawierającą wszystkie wartości X . □

19.2 Rozkład normalny na płaszczyźnie

Rozpatrzmy teraz rozkład normalny dwuwymiarowy, bardzo ważny w zastosowaniach. Załóżmy, że $X = (\xi, \eta)$ jest wektorem o rozkładzie $N_2(\mu, \Sigma)$. Rozkład ten zależy od pięciu parametrów:

$$\mu = \begin{bmatrix} m_\xi \\ m_\eta \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_\xi^2 & \varrho \sigma_\xi \sigma_\eta \\ \varrho \sigma_\xi \sigma_\eta & \sigma_\eta^2 \end{bmatrix}, \quad (19.1)$$

gdzie ϱ jest współczynnikiem korelacji.

Widać, że: $\Sigma > 0 \iff |\varrho| \neq 1$, gdyż:

$$\det \Sigma = \sigma_\xi^2 \sigma_\eta^2 (1 - \varrho^2), \quad \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_\xi^2(1-\varrho^2)} & -\frac{\varrho}{\sigma_\xi \sigma_\eta(1-\varrho^2)} \\ -\frac{\varrho}{\sigma_\xi \sigma_\eta(1-\varrho^2)} & \frac{1}{\sigma_\eta^2(1-\varrho^2)} \end{bmatrix}.$$

Twierdzenia 19.4 (2) oraz 19.5 zyskują teraz prostą interpretację.

Wniosek – 19.6. Zmienne losowe ξ, η są niezależne, wtedy i tylko wtedy, gdy $\rho = 0$.

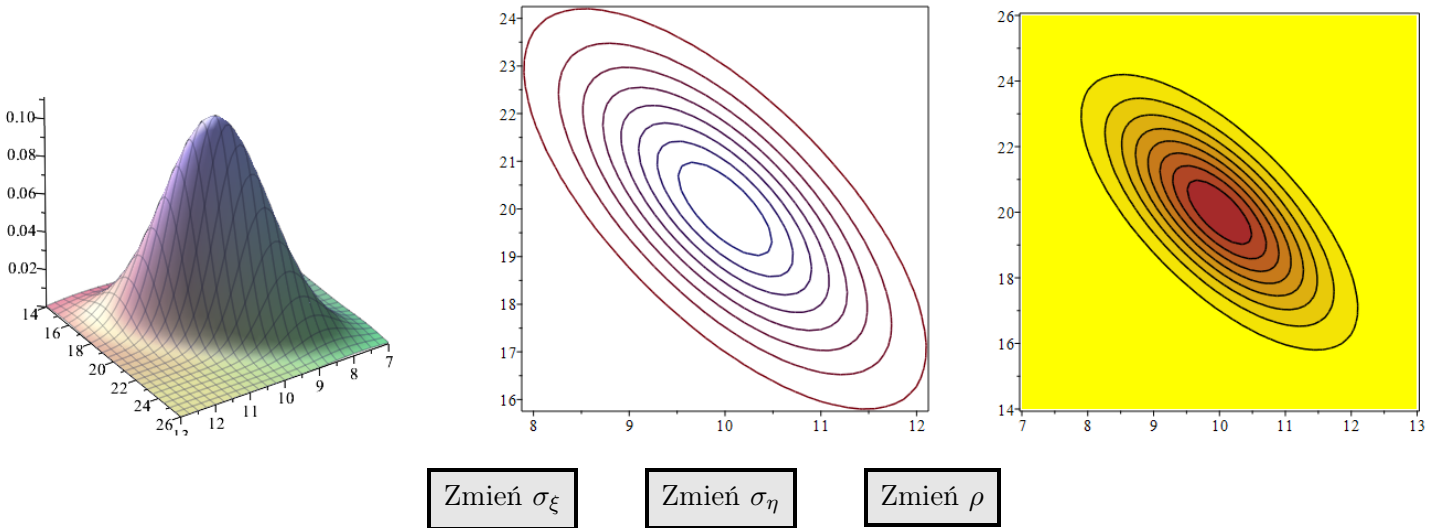
Istnieje taka prosta M , że $P((\xi, \eta) \in M) = 1$, wtedy i tylko wtedy, gdy $|\rho| = 1$.

Wzór na gęstość ma postać (ćwiczenie):

$$f(x, y) = \frac{1}{2\pi \sigma_\xi \sigma_\eta \sqrt{1 - \rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left(\frac{(x-m_\xi)^2}{\sigma_\xi^2} - 2 \frac{\rho(x-m_\xi)(y-m_\eta)}{\sigma_\xi \sigma_\eta} + \frac{(y-m_\eta)^2}{\sigma_\eta^2} \right)}. \quad (19.2)$$

Wszystkie parametry mają ciekawą interpretację geometryczną, którą można łatwo zrozumieć analizując powyższy wzór oraz wykres gęstości.

$$m_\xi = 10, m_\eta = 20, \sigma_\xi = 1, \sigma_\eta = 2, \text{ oraz } \rho := -0.7.$$



19.3 Warunkowanie rozkładów normalnych

Twierdzenie – 19.7. Niech wektor losowy X ma rozkład normalny $N_n(\mu, \Sigma)$, $n = n_1 + n_2$.

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Jeżeli macierz Σ_{22} jest nieosobliwa, to:

1. Dla każdego $x_2 \in \mathbb{R}^{n_2}$ rozkład warunkowy $P_{X_1|X_2=x_2}$ jest rozkładem normalnym $N_{n_1}(\mu^*, \Sigma^*)$, gdzie

$$\mu^* = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \quad \Sigma^* = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

2. $E(X_1|X_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$.

Dowód. Ad 1. Zdefiniujmy wektor losowy:

$$Y = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2.$$

Zauważmy, że (ćwiczenie):

$$\begin{bmatrix} Y \\ X_2 \end{bmatrix} = \begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix} \cdot \begin{bmatrix} X_1 \\ X_2 \end{bmatrix},$$

więc na podstawie twierdzenia 18.5 wektor $\begin{bmatrix} Y \\ X_2 \end{bmatrix}$ ma rozkład (ćwiczenie):

$$N_n \left(\begin{bmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \right).$$

To oznacza, że wektory Y oraz X_2 są niezależne, a także Y ma rozkład $N_{n_1}(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

Z niezależności Y oraz X_2 wynika, że $P_{Y|X_2=x_2} = P_Y$ dla dowolnego x_2 .

Ponieważ $X_1 = Y + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}x_2$, więc gdy $X_2 = x_2$, $P_{X_1|X_2=x_2}$ jest rozkładem wektora $X_1 = Y + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}x_2$, co na mocy twierdzenia 18.5 oznacza tezę. \square

Ad 2. $E(X_1|X_2 = x_2)$ jest nadzieją rozkładu $P_{X_1|X_2=x_2}$, więc z punktu 1 $E(X_1|X_2 = x_2) = \mu^* = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$.

A więc $E(X_1|X_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$. \square

Rozkład normalny na płaszczyźnie c. d.

Przypominamy, że rozważamy wektor $X = (\xi, \eta)$ o rozkładzie $N_2(\mu, \Sigma)$, przy czym.

$$\mu = \begin{bmatrix} m_\xi \\ m_\eta \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_\xi^2 & \rho \sigma_\xi \sigma_\eta \\ \rho \sigma_\xi \sigma_\eta & \sigma_\eta^2 \end{bmatrix}, \quad (19.3)$$

Założmy, że $|\rho| < 1$. Wtedy X ma gęstość:

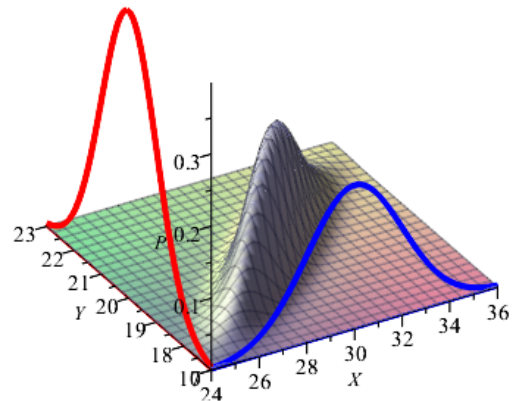
$$f(x, y) = \frac{1}{2\pi \sigma_\xi \sigma_\eta \sqrt{1 - \rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left(\frac{(x-m_\xi)^2}{\sigma_\xi^2} - 2 \frac{\rho(x-m_\xi)(y-m_\eta)}{\sigma_\xi \sigma_\eta} + \frac{(y-m_\eta)^2}{\sigma_\eta^2} \right)}. \quad (19.4)$$

Bezpośrednie znalezienie (przez całkowanie) gęstości rozkładów brzegowych i warunkowych nie jest oczywiste. Wiemy jednak z poprzednich twierdzeń, że:

Rozkłady brzegowe to:

$$P_\xi = N(m_\xi, \sigma_\xi),$$

$$P_\eta = N(m_\eta, \sigma_\eta).$$



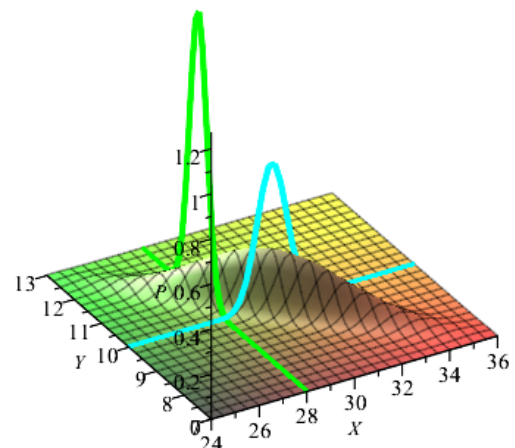
Rozkłady warunkowe to:

$$P_{\eta|\xi=x} = N\left(m_\eta + \rho \frac{\sigma_\eta}{\sigma_\xi} (x - m_\xi), \sigma_\xi^2 (1 - \rho^2)\right),$$

o ile $\sigma_\xi > 0$,

$$P_{\xi|\eta=y} = N\left(m_\xi + \rho \frac{\sigma_\xi}{\sigma_\eta} (y - m_\eta), \sigma_\eta^2 (1 - \rho^2)\right),$$

o ile $\sigma_\eta > 0$.



Znamy też nadzieje warunkowe:

$$E(\eta|\xi = x) = m_\eta + \rho \frac{\sigma_\eta}{\sigma_\xi} (x - m_\xi), \quad E(\xi|\eta = y) = m_\xi + \rho \frac{\sigma_\xi}{\sigma_\eta} (y - m_\eta).$$

Więc:

$$E(\eta|\xi) = m_\eta + \rho \frac{\sigma_\eta}{\sigma_\xi} (\xi - m_\xi), \quad \text{o ile } \sigma_\xi > 0, \quad E(\xi|\eta) = m_\xi + \rho \frac{\sigma_\xi}{\sigma_\eta} (\eta - m_\eta), \quad \text{o ile } \sigma_\eta > 0.$$

Zauważmy, że znamy rozkłady tych zmiennych losowych. Mianowicie:

$$E(\eta|\xi) \sim N(m_\eta, |\rho| \sigma_\eta), \quad E(\xi|\eta) \sim N(m_\xi, |\rho| \sigma_\xi).$$

Inaczej. Przy powyższych założeniach:

$$E(\eta|\xi) = h(\xi), \quad \text{gdzie } h(x) = m_\eta + \rho \frac{\sigma_\eta}{\sigma_\xi} (x - m_\xi),$$

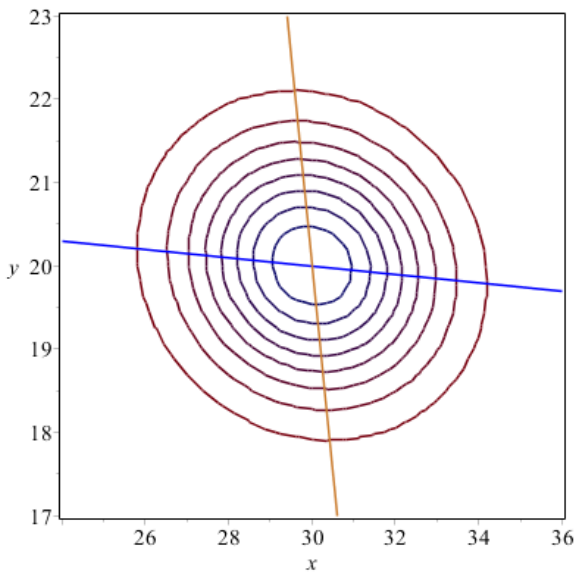
$$E(\xi|\eta) = g(\eta), \quad \text{gdzie } g(y) = m_\xi + \rho \frac{\sigma_\xi}{\sigma_\eta} (y - m_\eta).$$

h, g nazywane są *funkcjami regresji*. Ich wykresami są proste. W następnym punkcie omówimy problem regresji wyjaśniając podstawowe znaczenie funkcji regresji.

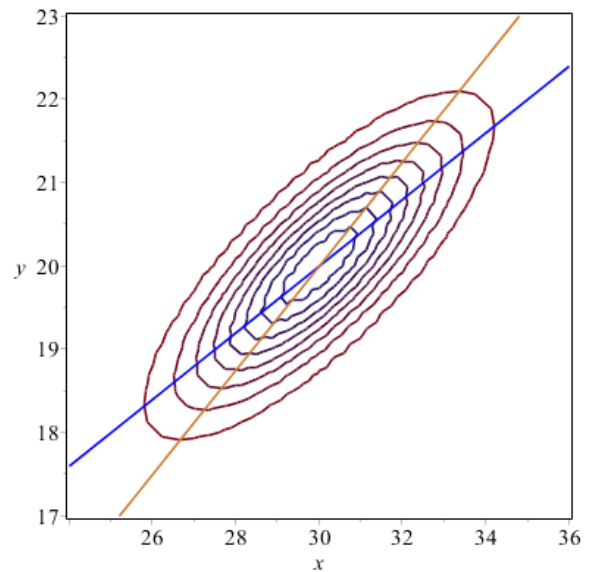
Przykład – 19.8. Zinterpretujemy geometrycznie funkcje regresji

$$h(x) = m_\eta + \rho \frac{\sigma_\eta}{\sigma_\xi} (x - m_\xi), \quad g(y) = m_\xi + \rho \frac{\sigma_\xi}{\sigma_\eta} (y - m_\eta)$$

dla rozkładu normalnego o parametrach $m_\xi = 30$, $m_\eta = 20$, $\sigma_\xi = 2$, $\sigma_\eta = 1$,



$$\rho := -0.1$$



$$\rho := 0.8$$

Przykład – 19.9. Znaleźć $E(2X + Y | X + Y - 2Z)$, gdzie X, Y, Z są niezależnymi zmiennymi losowymi o rozkładzie $N(0, 1)$ każda.

Wektor $\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$ ma rozkład $N_3 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$.

Znajdujemy najpierw rozkład wektora losowego $\begin{bmatrix} 2X + Y \\ X + Y - 2Z \end{bmatrix}$. Znając jego parametry μ oraz Σ stosujemy twierdzenie 19.7. Mamy:

$$\begin{bmatrix} 2X + Y \\ X + Y - 2Z \end{bmatrix} = A \begin{bmatrix} X \\ Y \\ Z \end{bmatrix},$$

gdzie $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -2 \end{bmatrix}$.

$$\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \Sigma = AI_3A^T = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 6 \end{bmatrix}.$$

$$E(2X + Y | X + Y - 2Z) = 0 + 3\frac{1}{6}(X + Y - 2Z - 0) = \frac{1}{2}(X + Y - 2Z).$$

19.4 Wnioskowanie bayesowskie.

Podamy przykład typowej sytuacji, w której mamy do czynienia z warunkowaniem.

Przykład – 19.10. Analityk pracujący dla partii ABC ma określić przedział, w którym zawiera się średnie poparcie dla tej partii. W momencie przystąpienia do pracy dostał wiadomość, że różne sondaże i inne badania wskazują, że zawiera się ono mniej więcej w przedziale od 20% do 25%. Otrzymał także nowe wyniki: na 1070 ankietowanych osób 199 opowiedziało się za partią ABC. Jak analityk powinien wykorzystać obydwie otrzymane informacje, aby rozwiązać postawione przed nim zadanie?

Na podstawie otrzymanych za pomocą sondażu wielkości łatwo, korzystając z CTG, estymować wartość oczekiwaną oraz jej przedział ufności. Mamy: $\hat{m} = x = 199/1070 = 0.18598$, $s = \sqrt{x(1-x)/1070} = 0.011895$ estymuje odchylenie standardowe σ , więc przedział ufności na poziomie ufności $0.95 = [\hat{m} - 1.96s, \hat{m} + 1.96s] = [0.162674, 0.20930]$. Aby jednak wykorzystać całą dostępną informację analityk stosuje tak zwane:

Podejście bayesowskie. Poparcie (frakcja osób popierających) dla partii ABC jest zmienną losową X o rozkładzie $N(M, \sigma)$, przy czym analityk zakłada, że parametr M też jest zmienną losową o rozkładzie $N(m_0, \sigma_0)$, gdzie parametry m_0, σ_0 są ustalone na podstawie wstępnych informacji (a priori). Szukamy rozkładu łącznego wektora losowego (M, X) , a następnie rozkład warunkowy $P_{M|X=x}$, co następnie pozwoli mu wykonać zadanie.

Na podstawie wstępnej informacji ustalmy parametry rozkładu $N(m_0, \sigma_0)$. Przyjmujemy $m_0 = 0.225$, $\sigma_0 = 0.0225/1.96 = 0.01276$. Rozkład ten nazywamy rozkładem **a priori** parametru M , a przedział $[0.2, 0.25] = [m_0 - 1.96\sigma_0, m_0 + 1.96\sigma_0]$ **przedziałem ufności a priori**.

Wyznaczamy kolejno:

$$f_{(M,X)}(m, x) = f_{X|M=m}(x) \cdot f_M(m) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2}\left(\frac{m-m_0}{\sigma_0}\right)^2},$$

$$f_X(x) = \int_{\mathbb{R}} f_{(M,X)}(m, x) dm,$$

$$f_{M|X=x}(m) = \frac{f_{(M,X)}(m, x)}{f_X(x)} = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{m-m_1}{\sigma_1}\right)^2},$$

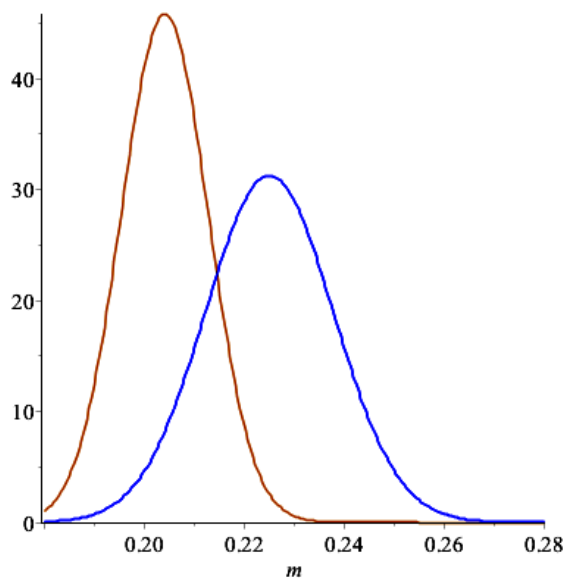
gdzie (ćwiczenie)

$$m_1 = \frac{m_0\sigma^2 + \sigma_0^2x}{\sigma^2 + \sigma_0^2}, \quad \sigma_1 = \sqrt{\frac{\sigma\sigma_0}{\sigma^2 + \sigma_0^2}}.$$

Otrzymujemy wartości: $m_1 = 0.20413$, $\sigma_1 = 0.00870$.

Znamy więc dokładnie rozkład $P_{M|X=x}$. Nazywamy go rozkładem **a posteriori** parametru M . Ponieważ wiemy, że $X = x$, to M – średnie poparcie dla partii ABC – zawiera się z prawdopodobieństwem 0.95 w przedziale $[m_1 - 1.96\sigma_1, m_1 + 1.96\sigma_1] = [0.18708, 0.22118]$. Jest on nazywany **a posteriori** przedziałem ufności lub **bayesowskim** przedziałem ufności.

Zauważmy, że przedział a posteriori jest znacznie węższy niż przedział a priori oraz węższy niż tradycyjny przedział ufności. Jego środek jest średnią ważoną średniej m_1 oraz x .



Gęstość rozkładu M : a priori oraz a posteriori

19.5 Liczby pseudolosowe z rozkładu normalnego

W wielu sytuacjach używamy liczb pseudolosowych z rozkładu normalnego jedno i wielowymiarowego. Jak wiemy już, przykład 12.4, w przypadku rozkładu ciągłego, a takim rozkładem jest $N(m, \sigma)$ można je otrzymywać generując liczby pseudolosowe u_1, u_2, u_3, \dots z rozkładu $U(0, 1)$ i biorąc $\Phi_{m,\sigma}^{-1}(u_1), \Phi_{m,\sigma}^{-1}(u_2), \Phi_{m,\sigma}^{-1}(u_3), \dots$. Można też brać liczby $z_i = \Phi^{-1}(u_i)$ z rozkładu $N(0, 1)$ oraz $x_i = m + \sigma z_i$, $i = 1, 2, 3, \dots$. Obydwa sposoby wymagają jednak wyznaczania wartości Φ^{-1} , co jest numerycznie dość uciążliwym zadaniem. Istnieją jednak szybsze metody. Opiszemy jedną z nich.

Metoda Boxa–Mullera Losujemy niezależnie od siebie dwie liczby u_1, u_2 z rozkładu $U(0, 1)$ i obliczamy

$$z_1 = \sqrt{-2 \ln u_1} \cos(2\pi u_2), \quad z_2 = \sqrt{-2 \ln u_1} \sin(2\pi u_2).$$

Liczby z_1, z_2 reprezentują dwie niezależne zmienne losowe o rozkładach $N(0, 1)$.

Niech $Z = h(g(U))$, przy czym $P_U = U([0, 1]^2)$ (więc $f_U = I_{[0, 1]^2}$) oraz określamy $g(u_1, u_2) = (\sqrt{-2 \ln u_1}, 2\pi u_2)$, $h(r, \theta) = (r \cos \theta, r \sin \theta)$.

Mamy kolejno:

$$f_{g(U)}(w) = |\text{Jac}_w g^{-1}| f_U(g^{-1}(w)) = \frac{1}{2\pi} e^{-\frac{1}{2} w_1^2} w_1, \quad w_1 > 0, \quad 0 < w_2 < 2\pi.$$

$$f_Z(z) = f_{h(g(U))}(z) = |\text{Jac}_z h^{-1}| f_{g(U)}(h^{-1}(z)) = \frac{1}{\sqrt{z_1^2 + z_2^2}} \frac{1}{2\pi} e^{-\frac{1}{2}(z_1^2 + z_2^2)} \sqrt{z_1^2 + z_2^2} = \frac{1}{2\pi} e^{-\frac{1}{2}(z_1^2 + z_2^2)}.$$

X ma więc rozkład $N_2(0, I_2)$, gdzie I_2 jest macierzą identycznościową. Więc Z_1 oraz Z_2 są niezależne i mają rozkłady normalne $N(0, 1)$. \square

Mając do dyspozycji ciąg o długości d liczb pseudolosowych z_1, \dots, z_d z rozkładu $N(0, 1)$ wiemy, że stanowi on wektor pseudolosowy z z rozkładu $N_d(0, I_d)$. Wtedy też $x = \mu + \Sigma^{\frac{1}{2}} z$ jest wektorem pseudolosowym z rozkładu $N_d(\mu, \Sigma)$.

19.6 Pytania

Pytanie 19.1. Czy można tak dobrać stałą c , że funkcja $f: f(x, y) = ce^{-2x^2 - 3y^2 + 2x}$ jest gęstością rozkładu normalnego? Uzasadnić. ROZWIĄZANIE

Pytanie 19.2. Wektor losowy o współrzędnych S, T ma rozkład jednostajny na kwadracie $[0, 1]^2$.

(1) Wskazać dwoma sposobami nadzieję matematyczną oraz macierz kowariancji wektora o współrzędnych $S + T, S - T$.

(2) Czy zmienne losowe $S + T, S - T$ są niezależne? ROZWIĄZANIE

Pytanie 19.3. Wskazać rozkład wektora (X, Y) oraz rozkład warunkowy $P_{X|Y=y}$, gdy wiemy, że $P_X = N(m, \sigma)$ oraz $P_{Y|X=x} = N(x, \sigma_x)$, gdzie $m \in \mathbb{R}, \sigma > 0, \sigma_x > 0$ są znane. ROZWIĄZANIE

Pytanie 19.4. Wektor losowy Y ma rozkład normalny $N_n(0, \sigma I_n)$, gdzie $\sigma > 0$ jest dane, a I_n jest macierzą identycznościową. Wykazać, że wektor $W = AY$ ma też rozkład $N_n(0, \sigma I_n)$, gdy A jest macierzą izometrii. ROZWIĄZANIE

Pytanie 19.5. Niech wektor X o współrzędnych ξ, η ma rozkład ciągły normalny. Wykazać, że proste regresji są prostopadłe, wtedy i tylko wtedy, gdy ξ, η są niezależne. ROZWIĄZANIE

Pytanie 19.6. Znaleźć $E(X + Y + Z|X - 2Y)$, gdzie X, Y, Z są niezależnymi zmiennymi losowymi o rozkładzie $N(m, 1)$ każda, $m \in \mathbb{R}$. ROZWIĄZANIE

Rozdział 20

Regresja

Problem regresji pojawia się w różnych kontekstach i ma różne interpretacje i sposoby rozwiązywania. Ogólnie rzecz biorąc można niezbyt formalnie powiedzieć, że mając dwie wielkości, powiedzmy x oraz y zależne od wspólnego parametru chcemy ustalić czy i jak bardzo y zależy od x . Inaczej, czy istnieje taka funkcja h , że $y \cong h(x)$, przy czym przybliżenie y za pomocą $h(x)$ powinno być możliwie jak najlepsze. Czasami też żądamy, aby sama funkcja h była możliwie prosta, na przykład, aby była liniowa, wykładnicza, lub innej wygodnej dla nas postaci. Poniżej sformułujemy i rozwiążemy problem regresji w przypadku, gdy wielkości x oraz y mają charakter losowy.

20.1 Ogólny problem regresji

Niech X będzie k -wymiarowym wektorem losowym, Y zmienną losową określoną na tej samej przestrzeni probabilistycznej (Ω, Σ, P) . Chcemy wskazać funkcję $h : \mathbb{R}^k \rightarrow \mathbb{R}$, taką, że:

$$Y = h(X) + \varepsilon, \quad (20.1)$$

gdzie ε jest „możliwie małe”.

Taka funkcja h nazywana jest **funkcją regresji Y względem X** .

Frazę „możliwie małe” można rozumieć na różne sposoby. Bardzo często żąda się, żeby $E(\varepsilon) = 0$ oraz żeby wielkość $E(\varepsilon^2) = D^2(\varepsilon)$ była najmniejsza. Problem polega więc na wskazaniu takiej funkcji $h : \mathbb{R}^k \rightarrow \mathbb{R}$, że $E((Y - h(X))^2)$ osiąga możliwie najmniejszą wartość. Zadanie to ma rozwiązanie przy całkiem naturalnych założeniach.

Poszukiwanie funkcji regresji w oparciu o powyższe kryterium jest dość ogólną wersją tak zwanej metody najmniejszych kwadratów. Mniej ogólną wersję poznamy później przy omawianiu problemu regresji liniowej.

Twierdzenie – 20.1. *Przy powyższych oznaczeniach założmy, że $E(Y^2) < \infty$. Niech*

$$h(x) = E(Y|X = x), \text{ dla } x \in \mathbb{R}^k.$$

Wtedy dla każdej funkcji borelowskiej $g : \mathbb{R}^k \rightarrow \mathbb{R}$, takiej, że $\int_{\mathbb{R}^k} g^2 dP_X < \infty$:

$$E((Y - h(X))^2) \leq E((Y - g(X))^2).$$

Twierdzenie to jest prostym wnioskiem z twierdzenia 20.7, które będzie wykazane w następnym punkcie.

W kilku przykładach wyznaczaliśmy już nadzieje warunkowe. Przykłady te znajdują teraz dodatkową interpretację, gdyż bez trudu możemy wypisać w każdym z nich funkcję regresji. Przypomnijmy tylko, strona 107, że nadzieja warunkowa $E(Y|X = x)$, a więc i funkcja regresji, jest jednoznacznie określona jedynie w punktach x , które mogą być przyjęte przez zmienną X .

W przykładzie 13.1 funkcja regresji Y względem X przyjmuje istotne wartości w trzech punktach: $h(0) = \frac{7}{2}$, $h(1) = \frac{14}{3}$, $h(2) = \frac{14}{3}$. Natomiast funkcja regresji X względem Y przyjmuje istotne wartości w sześciu punktach, np. $h(1) = 0$, $h(4) = \frac{10}{7}$.

W przykładzie 13.2 funkcja regresji Y względem X jest określona jako $h(x) = \frac{x}{2}$ dla $0 < x < 1$, natomiast funkcja regresji X względem Y jest dana jako $h(y) = \frac{1-y}{-\ln y}$ dla $0 < y < 1$.

W przykładzie 14.2 funkcja regresji Y względem X jest dana jako $h(0) = 0$, $h(1) = \frac{7}{2}$, $h(2) = 7$.

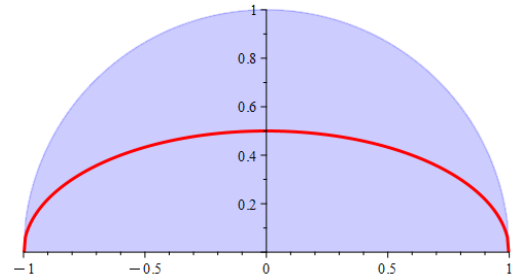
W przykładzie 14.9 funkcja regresji Y względem wektora $(X, Z)^T$ określona jest wzorem:

$$h(x, z) = \begin{cases} \frac{x}{2}, & \text{dla } 0 < x < 1, z = 0, \\ \frac{x+1}{2}, & \text{dla } 0 < x < 1, z = 1. \end{cases}$$

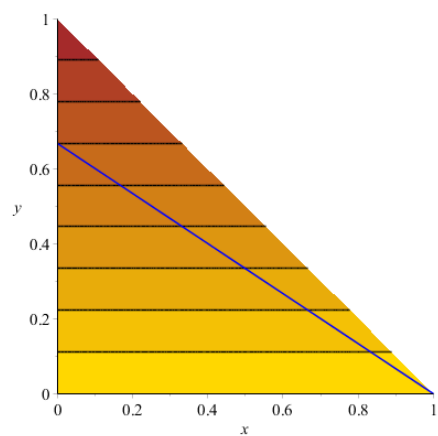
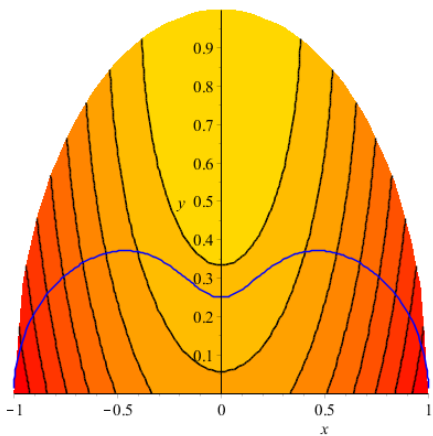
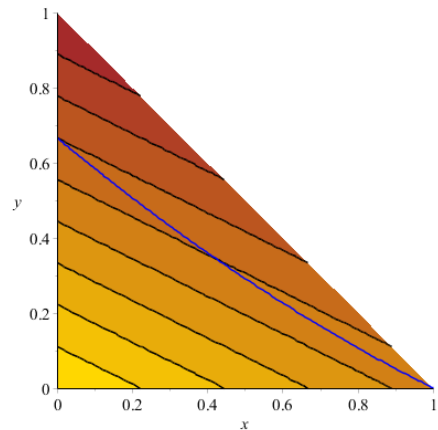
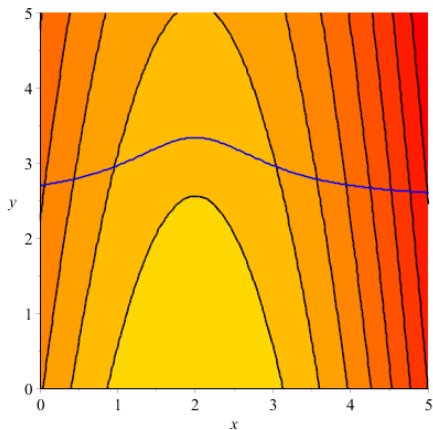
Podamy kilka dalszych przykładów funkcji regresji.

Przykład – 20.2. Ten dość trywialny przykład uwypukla znaczenie warunkowania dla problemu regresji. Niech X będzie wektorem losowym, Y zmienną losową, $E(Y) \in \mathbb{R}$. Załóżmy, że istnieje taka funkcja borekowska h , że $Y = h(X)$. Wtedy oczywiście h jest funkcją regresji. Ale z powyższego twierdzenia wynika, że $E(Y|X) = E(h(X)|X) = h(X)$. Ta ostatnia równość jest własnością 3. w twierdzeniu 13.26.

Przykład – 20.3. Niech (X, Y) będzie wektorem losowym o rozkładzie jednostajnym na półkolu $x^2 + y^2 \leq 1$, $y \geq 0$. Wtedy dla każdego $x \in [-1, 1]$ rozkład warunkowy $P_{Y|X=x}$ jest rozkładem jednostajnym na odcinku $(0, \sqrt{1-x^2})$ (ćwiczenie), więc $E(Y|X = x) = \frac{1}{2}\sqrt{1-x^2}$. Funkcja $h(x) = \frac{1}{2}\sqrt{1-x^2}$ dla $|x| < 1$ oraz $h(x) = 0$ dla $|x| \geq 1$.



Następne cztery wykresy mogą pomóc wyobrazić sobie przebieg funkcji regresji dla różnych rozkładów ciągłych. Na każdym wykresie znajduje się poziomicowy wykres gęstości $f_{(X,Y)}$ pewnego rozkładu wraz z wykresem funkcji $h: h(x) = E(Y|X = x)$.



Jak widzimy w niektórych sytuacjach funkcja regresji jest funkcją afiniczną. Tak zawsze jest w przypadku rozkładów normalnych. Możemy mianowicie przeformułować twierdzenie 19.7.2.

Twierdzenie – 20.4. Niech $\begin{bmatrix} X \\ Y \end{bmatrix}$ ma rozkład normalny, $\Sigma_X > 0$. Wtedy funkcja regresji, h , jest funkcją afiniczną.

$$h(x) = E(Y|X = x) = \mu_Y + \Sigma_{YX}\Sigma_X^{-1}(x - \mu_X).$$

Tak więc nazwa „funkcja regresji” użyta w przypadku rozkładu normalnego na płaszczyźnie okazuje się szczególnym przypadkiem ogólnej definicji.

20.2 Nadzieja warunkowa jako rzutowanie

Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną. W zbiorze zmiennych losowych $\Omega \rightarrow \mathbb{R}$ wprowadzamy relację równoważności utożsamiającą X i Y , gdy $P(X = Y) = 1$. Pisząc X będziemy dalej rozumieć klasę równoważności X względem tej relacji. Niech $\mathcal{A} \subset \Sigma$ będzie σ -algebrą. Definiujemy

$$\mathcal{H}_{\mathcal{A}} = L^2(\Omega, \mathcal{A}, P) = \{X : X \text{ jest } \mathcal{A}\text{-mieralne, } E(X^2) < \infty\}.$$

Oznaczamy, $\mathcal{H} = \mathcal{H}_{\Sigma}$.

Każdy zbiór $\mathcal{H}_{\mathcal{A}}$ wraz z dodawaniem i mnożeniem przez liczby jest przestrzenią wektorową. Wynika to natychmiast z elementarnej nierówności $(x + y)^2 \leq 2(x^2 + y^2)$. Faktycznie $\mathcal{H}_{\mathcal{A}}$ jest podprzestrzenią wektorową \mathcal{H} .

Definiujemy iloczyn skalarny na \mathcal{H} wzorem:

$$\varphi(X, Y) = E(X \cdot Y).$$

Iloczyn skalarny dyktuje normę na \mathcal{H} w standardowy sposób:

$$\|X\| = \sqrt{\varphi(X, X)}$$

oraz metrykę d :

$$d(X, Y) = \|X - Y\|.$$

Dowodzi się, że dla każdej σ -algebry \mathcal{A} przestrzeń $\mathcal{H}_{\mathcal{A}}$ wraz z metryką d jest przestrzenią zupełną, to znaczy $\mathcal{H}_{\mathcal{A}}$ jest tak zwaną przestrzenią Hilberta. W szczególności $\mathcal{H}_{\mathcal{A}}$ jest zbiorem domkniętym w \mathcal{H} .

Zdefiniowana powyżej metryka definiuje zbieżność zwaną *zbieżnością średnio kwadratową* i oznaczaną $\xrightarrow{L^2}$. W szczególności: $X_n \xrightarrow{L^2} X \iff E((X_n - X)^2) \rightarrow 0$, dla $n \rightarrow \infty$.

Określa się prostopadłość wektorów: $X \perp Y \iff \varphi(X, Y) = 0$. Dla każdej podprzestrzeni wektorowej domkniętej $\mathcal{G} \subset \mathcal{H}$ określa się zbiór

$$\mathcal{G}^{\perp} = \{X \in \mathcal{H} \mid \forall Y \in \mathcal{G} : X \perp Y\}.$$

Dowodzi się, że $\mathcal{H} = \mathcal{G} \oplus \mathcal{G}^{\perp}$. Przez $P_{\mathcal{G}}$ oznaczamy rzutowanie prostopadłe na \mathcal{G} .

Mamy więc: dla każdego $Y \in \mathcal{H}$:

$$Y = P_{\mathcal{G}}Y + P_{\mathcal{G}^{\perp}}Y, \quad \|Y\|^2 = \|P_{\mathcal{G}}Y\|^2 + \|P_{\mathcal{G}^{\perp}}Y\|^2.$$

Przypominamy, że odległość Y od \mathcal{G} określa się następująco:

$$d(Y, \mathcal{G}) \stackrel{\text{def}}{=} \inf\{\|Y - Z\| : Z \in \mathcal{G}\}.$$

Korzystając w twierdzenia Pitagorasa w formie:

$$\|Y - Z\|^2 = \|Z - P_{\mathcal{G}}Y\|^2 + \|P_{\mathcal{G}^{\perp}}Y\|^2, \quad \text{dla } Y \in \mathcal{H}, Z \in \mathcal{G}.$$

widzimy, że wektor $P_{\mathcal{G}}Y$ realizuje odległość Y od \mathcal{G} , która wynosi $\|P_{\mathcal{G}^{\perp}}Y\|$.

Twierdzenie – 20.5. Niech $\mathcal{A} \subset \Sigma$ będzie σ -algebrą. Niech $\pi = P_{\mathcal{H}_{\mathcal{A}}}$ będzie rzutowaniem prostopadłym na $\mathcal{H}_{\mathcal{A}}$. Wtedy dla każdego $Y \in \mathcal{H}$:

$$E(Y|\mathcal{A}) = \pi Y, \quad D^2(Y) = D^2(E(Y|\mathcal{A})) + E((Y - E(Y|\mathcal{A}))^2).$$

Dowód. Wiemy już, że $\mathcal{H}_{\mathcal{A}}$ jest podprzestrzenią wektorową domkniętą.

Dla $Y \in \mathcal{H}$: πY , z definicji π , jest zmienną losową \mathcal{A} -mierzalną.

Dla dowolnego zbioru $A \in \mathcal{A}$ funkcja charakterystyczna $\mathbb{1}_A \in \mathcal{H}_{\mathcal{A}}$, więc otrzymujemy $\varphi(Y - \pi Y, \mathbb{1}_A) = 0$. Inaczej:

$$0 = E((Y - \pi Y) \cdot \mathbb{1}_A) = \int_{\Omega} (Y - \pi Y) \cdot \mathbb{1}_A dP = \int_A Y - \pi Y dP = \int_A Y dP - \int_A \pi Y dP.$$

Stąd πY spełnia warunek: $\int_{\mathcal{A}} Y dP = \int_{\mathcal{A}} \pi Y dP$. Są więc spełnione warunki (C) i (M) w definicji nadziei warunkowej, zob. definicja 13.12. Z jednoznaczności nadziei warunkowej otrzymujemy pierwszą żadaną równość.

Dla dowodu drugiej równości zauważmy, że:

$$\begin{aligned} D^2(Y) &= E(Y^2) - E(Y)^2 = \|Y\|^2 - E(Y)^2 = \|\pi Y\|^2 + \|Y - \pi Y\|^2 - E(Y)^2 \\ &= E(E(Y|\mathcal{A})^2) - E(E(Y|\mathcal{A}))^2 + E((Y - E(Y|\mathcal{A}))^2) \\ &= D^2(E(Y|\mathcal{A})) + E((Y - E(Y|\mathcal{A}))^2). \end{aligned} \quad \square$$

Uwaga – 20.6. Powyższe twierdzenie wzmacnia twierdzenie 14.11 o obniżaniu wariancji. Dodatkowo wyjaśnia, jak duża jest różnica między wariancjami przed i po warunkowaniu. Różnica ta wynosi $E((Y - E(Y|\mathcal{A}))^2)$, co można także zapisać jako $D^2(Y - E(Y|\mathcal{A}))$.

Przy warunkowaniu różnica wariancji równa jest wariancji różnicy.

Ustalmy teraz wektor losowy $X : \Omega \rightarrow \mathbb{R}^k$. Na podstawie twierdzenia 13.19

$$\mathcal{H}_{\sigma(X)} = \left\{ Z : \Omega \rightarrow \mathbb{R} : Z = \alpha(X), \alpha : \mathbb{R}^k \rightarrow \mathbb{R} \text{ jest borelowskie, } \int_{\mathbb{R}^k} \alpha^2 dP_X < \infty \right\}.$$

Jako wniosek z twierdzenia 20.5 otrzymujemy:

Twierdzenie – 20.7. Dla dowolnego wektora losowego $X : \Omega \rightarrow \mathbb{R}^k$ oraz zmiennej losowej $Y : \Omega \rightarrow \mathbb{R}$ takiej, że $E(Y^2) < \infty$ odzworowanie:

$$L^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_X) \ni \alpha \rightarrow E((Y - \alpha(X))^2) \in \mathbb{R}$$

przyjmuje wartość najmniejszą gdy $\alpha(x) = E(Y|X = x)$.

Jest to równoważne sformułowanie twierdzenia 20.1.

20.3 Regresja liniowa

W pewnych przypadkach zamiast ogólnego problemu regresji rozpatruje się zagadnienie, w którym na funkcję regresji nakłada się dodatkowe ograniczenia. Takie podejście może być podyktowane różnymi względami. Jednym z nich może być trudność w analitycznym (lub nawet numerycznym) wyznaczeniu funkcji regresji. Czasem, funkcja regresji jest jednoznacznie wyznaczona tylko w skończonej liczbie punktów – tak jest w przypadku rozkładów dyskretnych na zbiorze skończonym – co niekiedy utrudnia jej interpretację. Poniżej omawiamy problem regresji liniowej.

Niech X będzie k -wymiarowym wektorem losowym, Y zmienną losową określoną na tej samej przestrzeni probabilistycznej (Ω, Σ, P) , $E(Y^2) < \infty$. Chcemy wskazać funkcję **afiniczną** $h : \mathbb{R}^k \rightarrow \mathbb{R}$, taką, że:

$$Y = h(X) + \varepsilon, \tag{20.2}$$

gdzie ε jest możliwie małe. h – funkcja regresji liniowej Y względem X .

Zauważmy, że w przypadku, gdy wektor $(X, Y)^T$ ma rozkład normalny, powyższy problem jest tożsamy z ogólnym problemem regresji. W niektórych innych przypadkach też tak się może zdarzyć.

Jak jednak wskazują poznane dotychczas przykłady, nie zawsze tak jest.

Aby rozpatrzeć przypadek regresji liniowej dla dowolnych rozkładów skoncentrujemy się na sytuacji dwuwymiarowej. Niech ξ, η będą zmiennymi losowymi. Często wiemy, że zmienne ξ, η są mocno skorelowane to znaczy ich współczynnik korelacji ρ jest na moduł bliski 1, ich wspólny rozkład $P_{(\xi, \eta)}$ może być skupiony na zbiorze leżącym blisko pewnej prostej. Powstaje wtedy problem znalezienia tej prostej. Nawet gdy ρ jest bliski zeru, poszukiwanie takiej prostej ma pewien sens. Jeżeli jest to prosta o równaniu $y = ax + b$, możemy napisać:

$$\eta = a\xi + b + \varepsilon,$$

gdzie ε jest zmienną losową reprezentującą popełniany błąd. Funkcją regresji liniowej jest funkcja h , $h(x) = ax + b$ dla $x \in \mathbb{R}$. Powyższą prostą nazywa się *prostą regresji liniowej*.

Współczynniki a oraz b znajdziemy elementarną metodą najmniejszych kwadratów zastosowaną już w przypadku ogólnym.

Metoda najmniejszych kwadratów.

Szukamy takich a, b , że wielkość

$$E((\eta - (a\xi + b))^2).$$

jest najmniejsza.

Twierdzenie – 20.8. Załóżmy, że wariancje zmiennych ξ oraz η istnieją oraz $\sigma_\xi^2 > 0$. Wtedy istnieje dokładnie jedna para liczb a, b taka, że funkcja $E((\eta - (a\xi + b))^2)$ ma w punkcie a, b wartość najmniejszą. Wielkości te wynoszą:

$$a = \rho \frac{\sigma_\eta}{\sigma_\xi}, \quad b = m_\eta - \frac{\rho \sigma_\eta m_\xi}{\sigma_\xi}. \quad (20.3)$$

Dowód. Oznaczmy:

$$f(a, b) = E((\eta - (a\xi + b))^2).$$

Skorzystamy z warunku koniecznego na ekstremum powyższej funkcji dwóch zmiennych. Policzmy w tym celu pochodne cząstkowe funkcji f i znajdziemy punkt, w którym są one równe 0.

$$\begin{aligned} \frac{\partial f}{\partial a} &= E\left(\frac{\partial}{\partial a}(\eta - a\xi - b)^2\right) = E((2(\eta - a\xi - b)(-\xi)) = -2E(\xi\eta) + 2E(\xi^2)a + 2E(\xi)b. \\ \frac{\partial f}{\partial b} &= E\left(\frac{\partial}{\partial b}(\eta - a\xi - b)^2\right) = E((2(\eta - a\xi - b)(-1)) = -2E(\eta) + 2E(\xi)a + 2b. \end{aligned}$$

Mamy więc układ równań liniowych ze względu na a, b :

$$\begin{cases} E(\xi^2)a + E(\xi)b = E(\xi\eta), \\ E(\xi)a + b = E(\eta). \end{cases}$$

Wyznacznik tego układu wynosi $E(\xi^2) - E(\xi)^2$, która to wielkość jest wariancją σ_ξ^2 . Z założenia jest ona różna od 0, a więc nasz układ ma dokładnie jedno rozwiązanie.

Wyraża się ono właśnie wzorami (20.3) (ćwiczenie).

Dla kompletności dowodu trzeba uzasadnić, że funkcja f jest rzeczywiście różniczkowalna, że można „wchodzić” z pochodną pod znak nadziei matematycznej oraz że w wyliczonym punkcie (a, b) funkcja f osiąga wartość najmniejszą. Dwa pierwsze punkty wynikają z ogólnych twierdzeń o różniczkowaniu pod znakiem całki. Trzeci punkt może być uzasadniony na różne sposoby – na przykład za pomocą standardowego warunku wystarczającego na ekstremum (ćwiczenie). \square

Prosta regresji ma więc równanie:

$$y = ax + b = \rho \frac{\sigma_\eta}{\sigma_\xi} x + m_\eta - \frac{\rho \sigma_\eta m_\xi}{\sigma_\xi} = \rho \frac{\sigma_\eta}{\sigma_\xi} (x - m_\xi) + m_\eta.$$

Jak już wiemy, gdy (ξ, η) ma rozkład normalny funkcja regresji liniowej pokrywa się z funkcją regresji.

Przykład – 20.9. Rozkład wektora losowego (X, Y) skupiony jest na trójkącie D o wierzchołkach $(0, 0)$, $(0, 1)$, $(1, 0)$ i ma gęstość proporcjonalną do funkcji g : $g(x, y) = x + 2y^2$. Znajdziemy funkcję regresji i funkcję regresji liniowej Y względem X .

Mamy kolejno:

$$f_{X,Y}(x, y) = \frac{g(x, y)}{\int_D g(x, y) d(x, y)} = 3x + 6y^2$$

dla $(x, y) \in D$.

$$f_X(x) = \int_0^{1-x} f(x, y) dy,$$

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)},$$

$$E(Y|X=x) = \int_0^{1-x} y f_{Y|X=x}(y) dy$$

$$= \frac{3(x^2 - x + 1)(1-x)}{4x^2 - 2x + 4}.$$

$$E(X) = \int_0^1 x f_X(x) dx = 0.35,$$

$$E(Y) = \int_0^1 y f_Y(y) dy = 0.425,$$

$$D^2(X) = \int_0^1 x^2 f_X(x) dx - E(X)^2 = 0.061,$$

$$D^2(Y) = \int_0^1 y^2 f_Y(y) dy - E(Y)^2 = 0.069,$$

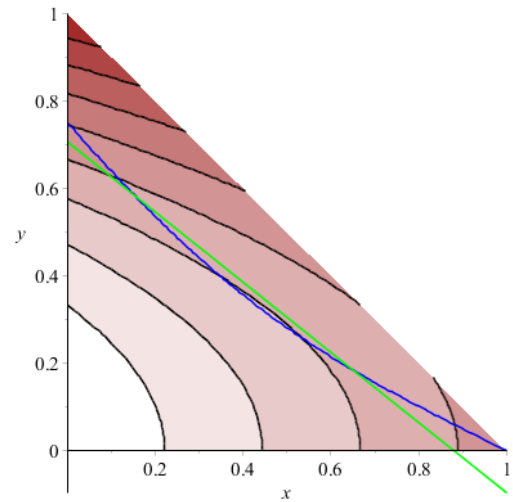
$$\text{cov}(X, Y) =$$

$$\int_D xy f_{(X,Y)}(x, y) d(x, y) - E(X)E(Y) = -0.047,$$

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{D^2(X)D^2(Y)}} = -0.705.$$

$$y = ax + b,$$

gdzie $a = \frac{\text{cov}(X,Y)}{D^2(X)} = -0.801$, $b = E(Y) - aE(X) = 0.705$.



Funkcja regresji i funkcja regresji liniowej

Zauważmy, że gdy wektor (X, Y) ma rozkład dyskretny to funkcja h , regresji Y względem X , jest istotnie określona tylko w punktach x_i takich, że $P(X = x_i) > 0$. Natomiast funkcja regresji liniowej jest określona dla wszystkich x .

20.4 Regresja jako narzędzie statystyczne

Pojęcie regresji odgrywa niezmiernie ważną rolę w badaniach statystycznych. Jakkolwiek występujące tam idee są podobne do tych przedstawionych powyżej, to warto wyjaśnić nieco inny kontekst w którym są stosowane. Przedstawimy jedno ze występujących w statystyce podejść do problemu regresji.

Rozpatruje się k wielkości X_1, \dots, X_k (tak zwanych zmiennych niezależnych – niekoniecznie zmiennych losowych) oraz wielkość Y (zwaną zmienną zależną). Chcemy ustalić związek Y oraz X_1, \dots, X_k postaci $Y = H(X_1, \dots, X_k) + \varepsilon$, gdzie H jest funkcją, a ε zmienną losową „możliwie małą”. Jednak w statystyce nie znamy na ogół rozkładów rozpatrywanych wektorów losowych. Co więcej nie zawsze zakłada się, że są to wektory losowe. Zamiast tego dane są obserwacje $(x_1, y_1), \dots, (x_n, y_n)$, gdzie każde $x_i = (x_{i1}, \dots, x_{ik}) \in \mathbb{R}^k$ reprezentuje wielkości X_1, \dots, X_k , natomiast y_i reprezentuje odpowiadającą im w określony sposób wielkość Y . Można teraz sformalizować – na różne sposoby – problem regresji. Oto jeden z nich:

Znaleźć taką funkcję $H : \mathbb{R}^k \rightarrow \mathbb{R}$, że dla każdego $i = 1, \dots, n$

$$y_i = H(x_i) + \varepsilon_i,$$

przy czym $\varepsilon_1, \dots, \varepsilon_n$ są niezależnymi zmiennymi losowymi o rozkładzie $N(0, \sigma)$ każda, a wielkość σ jest minimalna.

Zazwyczaj zakłada się szczególną postać funkcji H (np. afiniczność) lub własności (np. różniczkowalność). Założenie o normalności $\varepsilon_1, \dots, \varepsilon_n$ nie jest niezbędne, niemniej w wielu przypadkach jest naturalne, a także korzystne ze względów technicznych. Istnieje wiele metod rozwiązywania powyższego zagadnienia regresji. Na przykład, przy założeniu, że funkcja H jest afiniczna, stosuje się często metodę najmniejszych kwadratów, która okazuje się tożsama z omówioną przez nas metodą zastosowaną do przypadku rozkładu dyskretnego (najczęściej jednostajnego) gdyby był on skupiony dokładnie na punktach $(x_1, y_1), \dots, (x_n, y_n)$. W przypadku dwuwymiarowym zachodzi wtedy odpowiednik twierdzenia 20.8, przy czym występujące tam parametry rozkładów wyznaczone są na podstawie obserwacji. Podobnie jest w wyższych wymiarach.

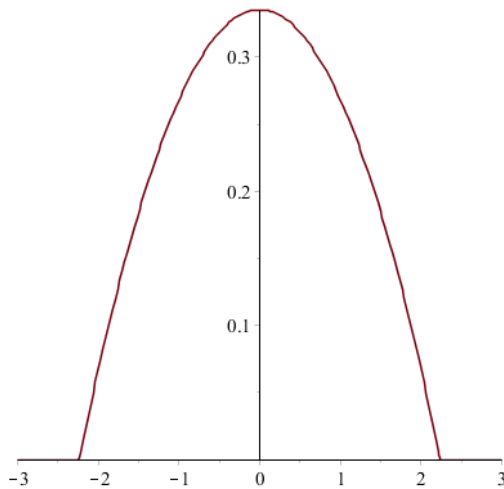
Poniżej podamy przykład tak zwanej metody nieparametrycznej, która pozwala wyznaczać funkcję, która do pewnego stopnia odpowiada ogólnej funkcji regresji.

Niech $(x_1, y_1), \dots, (x_n, y_n)$ będzie danym ciągiem punktów płaszczyzny. Dla danego punktu $x \in \mathbb{R}$ określamy wartość $H(x)$ wzorem:

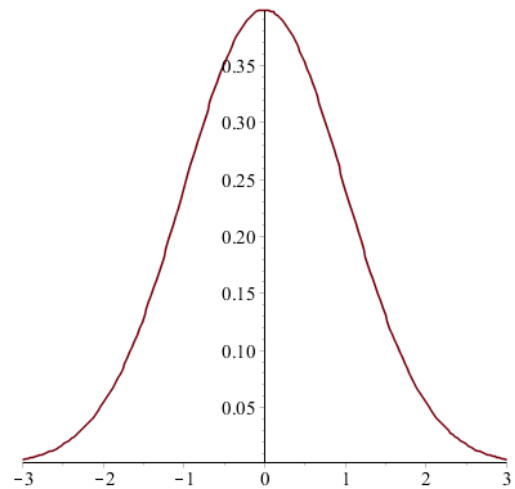
$$H(x) = \frac{\sum_{i=1}^n y_i w_i}{\sum_{i=1}^n w_i}, \quad (20.4)$$

gdzie w_i są nieujemnymi wagami. Chcemy je dobrać tak, aby na wartość $H(x)$ największy wpływ miały te punkty, których współrzędne x -owe są blisko x .

Jednym ze sposobów dobrania wag jest tak zwana metoda jądrowa. Wybieramy funkcję K zwaną jądrem, która spełnia następujące własności: (1) jest gęstością, (2) przyjmuje wartość największą dla $x = 0$, (3) $K(-x) = K(x)$ dla każdego x . Na przykład:



$$K_1(x) = \frac{3}{4\sqrt{5}} \left(1 - \frac{x^2}{5}\right) \mathbb{1}_{(-\sqrt{5}, \sqrt{5})}(x)$$



$$K_2(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Ustala się też parametr (szerokość pasmo) $h > 0$. Jako wagi bierze się:

$$w_i = K\left(\frac{x - x_i}{h}\right).$$

Otrzymujemy więc dla każdego wielkość $x \in \mathbb{R}$:

$$H(x) = \frac{\sum_{i=1}^n y_i K\left(\frac{x - x_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)}, \quad (20.5)$$

zwaną *estymatorem Nadarayi–Watsona*.

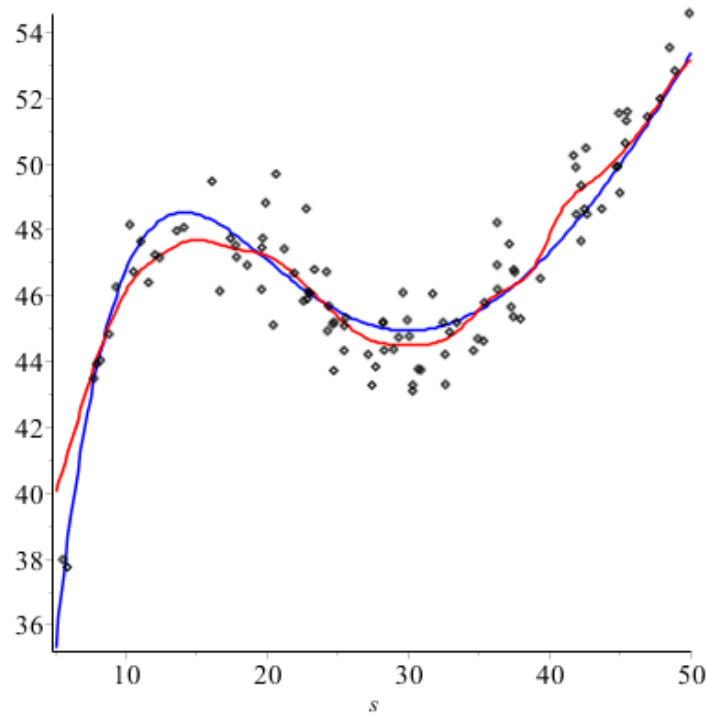
Istnieje bogata literatura, patrz [9], która daje wskazówki jak optymalnie dobrać jądro K oraz – co ważniejsze – szerokość pasma h . Gdy h jest bliskie zeru, wpływ na wartość $H(x)$ mają praktycznie tylko punkty leżące najbliżej. Duża wartość h oznacza większą gładkość wykresu H .

Przykład – 20.10. Aby zademonstrować działanie metody jądrowej Nadarayi–Watsona rozważmy wielkości X oraz Y określone następująco. X jest zmienną losową o rozkładzie jednostajnym na przedziale (a, b) , Y jest zmienną losową określoną jako $Y = h(X) + \varepsilon$, gdzie h jest zadaną przez nas funkcją, a ε ma rozkład normalny $N(0, \sigma)$. Losujemy próbkę prostą x_1, \dots, x_n reprezentującą X , próbkę prostą $\varepsilon_1, \dots, \varepsilon_n$ reprezentującą ε oraz obliczamy $y_i = h(x_i) + \varepsilon_i$ dla $i = 1, \dots, n$.

Zauważmy, że $E(Y|X = x) = E(h(X)|X = x) + E(\varepsilon|X = x) = h(x) + 0 = h(x)$.

Korzystając z wybranych punktów obliczamy wartości estymatora H .

Na wspólnym rysunku porównujemy na tle naszych punktów wykresy funkcji h oraz H . Przyjeliśmy $a = 5$, $b = 50$, $n = 100$, $h(x) = x + 10xe^{-0.1x}$, $\sigma = 1$, jądro K_2 i szerokość pasma $h = 1.8$.



Zmień h Porównaj z Ćwiczeniem 20.3.D.

20.5 Pytania

Pytanie 20.1. Dane są obserwacje $(x_1, y_1), \dots, (x_n, y_n)$, gdzie każde $x_i = (x_{i1}, \dots, x_{ik}) \in \mathbb{R}^k$ reprezentuje określone wielkości X_1, \dots, X_k , natomiast y_i reprezentuje odpowiadające im w pewien określony sposób wielkości Y_i . Na podstawie tych danych należy wyznaczyć takie liczby $\hat{\beta}_1, \dots, \hat{\beta}_k$, że sumy $\sum_{j=1}^k \hat{\beta}_j x_{ij}$ możliwie najlepiej przybliżają wielkości y_i dla $i = 1, \dots, n$ w sensie kryterium najmniejszych kwadratów. Zaproponować metodę rozwiązania tego problemu. (ROZWIĄZANIE)

Pytanie 20.2. Przy powyższych oznaczeniach, założmy, że $k < n$ oraz, że rząd $X = k$. Wyznaczyć rzutowanie prostopadłe na podprzestrzeń wektorową $V = \text{Im } X$. (ROZWIĄZANIE)

Pytanie 20.3. Rozwiązać problem sformułowany w pytaniu 20.1 wykorzystując powyższy wynik. (ROZWIĄZANIE)

Pytanie 20.4. Dla danego ciągu punktów płaszczyzny $(x_1, y_1), \dots, (x_n, y_n)$ wskazać takie liczby a, b , że wartość $\sum_{i=1}^n (y_i - (ax_i + b))^2$ jest najmniejsza. (ROZWIĄZANIE)

Pytanie 20.5. Dla danego ciągu punktów płaszczyzny $(x_1, y_1), \dots, (x_n, y_n)$ wskazać takie liczby a, b obliczone na podstawie twierdzenia 20.8 przy założeniu, że twierdzenie to stosuje się to wektora losowego (ξ, η) mającego rozkład jednostajny na zbiorze $\{(x_1, y_1), \dots, (x_n, y_n)\}$. (ROZWIĄZANIE)

Pytanie 20.6. Podobnie jak w pytaniu 20.1 dane są obserwacje $(x_1, y_1), \dots, (x_n, y_n)$, gdzie każde $x_i = (x_{i1}, \dots, x_{ik}) \in \mathbb{R}^k$ reprezentuje określone wielkości X_1, \dots, X_k , natomiast y_i reprezentuje odpowiadające im w pewien określony sposób wielkości Y_i . Obecnie zakładamy, że Y_i są zmiennymi losowymi o wartościach oczekiwanych μ_i , przy czym zakładamy, że dla każdego i zachodzi związek liniowy:

$$\mu_i = \beta_1 x_{i1} + \dots + \beta_k x_{ik},$$

gdzie β_1, \dots, β_k nie zależą od i . Zbadać własności estymatorów $\hat{\mu}$ parametru μ oraz $\hat{\beta}$ parametru β określonych jako:

$$\hat{\mu} = X(X^T X)^{-1} X^T Y, \quad \hat{\beta} = (X^T X)^{-1} X^T Y,$$

gdzie Y jest wektorem losowym o współrzędnych Y_i . (ROZWIĄZANIE)

Rozdział 21

Odpowiedzi do pytań

Pytanie 2.1. Niech $A, B \in \Sigma$. Wykaż, że jeżeli $P(A) = 1$, to $P(A \cap B) = P(B)$.

Wskazówka. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, $P(A \cup B) = P(A) = 1$. POWRÓT DO ZADANIA

Pytanie 2.2. Wyprowadzić wzór na $P(A \cup B \cup C)$. Uogólnić na dowolną sumę zbiorów.

Wskazówka. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.

POWRÓT DO ZADANIA

Pytanie 2.3. Gracz otrzymuje 13 kart wylosowanych z talii 52 kart i podaje liczbę otrzymanych asów oraz sumaryczną liczbę pozostałych figur. Wskazać zdarzenia, zdarzenia elementarne. Czy za pomocą prawdopodobieństw zdarzeń można obliczyć prawdopodobieństwo tego, że: (a) gracz otrzymał same piki, (b) gracz nie dostał asa?

Wskazówka. $\Omega = \{K \subset \{1, \dots, 52\} : \#K = 13\}$ – zdarzenia elementarne.

$\Sigma = \sigma(\{K_{ij} : i = 0, 1, 2, 3, 4, j = 0, 1, \dots, 12\})$ – zdarzenia, gdzie $K_{ij} \subset \{1, \dots, 52\}$ – zawiera dokładnie i asów oraz dokładnie j pozostałych figur.

Ad (a). Nie. Ad (b). Tak. POWRÓT DO ZADANIA

Pytanie 2.4. Gracz rzuca trzema monetami jednocześnie tak długo, aż na wszystkich trzech pojawi się orzeł. Opisać przestrzeń probabilistyczną odpowiadającą temu doświadczeniu.

Wskazówka. $\Omega = \{1, 2, 3, \dots\}$, $\Sigma = \mathcal{P}(\Omega)$, $p_i = \left(\frac{7}{8}\right)^{i-1} \frac{1}{8}$. POWRÓT DO ZADANIA

Pytanie 2.5. Niech \mathcal{F} oznacza rodzinę wszystkich przedziałów (a, b) , gdzie $a < b$, $a, b \in \mathbb{R}$. Wykazać, że $\sigma(\mathcal{F}) = \mathcal{B}(\mathbb{R})$.

Wskazówka. Każdy przedział (a, b) należy do $\mathcal{B}(\mathbb{R})$, więc $\sigma(\mathcal{F}) \subset \mathcal{B}(\mathbb{R})$. Zawieranie w drugą stronę wynika ze znanego faktu, że każdy otwarty podzbiór \mathbb{R} jest przeliczalną sumą otwartych przedziałów. POWRÓT DO ZADANIA

Pytanie 2.6. Do gry w pliszki potrzebne są dwa patyki: jeden o długości co najmniej 70 cm, drugi o długości 10 do 20 cm. Aby je uzyskać gracze łamią znalezione przez siebie patyki o długości 1 m w losowo wybranym punkcie. Jakie jest prawdopodobieństwo, że mogą grać?

Wskazówka. $\frac{1}{5}$. POWRÓT DO ZADANIA

Pytanie 3.1. Sformułować i wykazać twierdzenie mówiące o tym, że $P(\cdot | W)$ określa przestrzeń probabilistyczną.

Wskazówka. Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną. Niech $W \in \Sigma$, $P(W) > 0$. Niech $\Sigma_W = \{A \subset W : A \in \Sigma\}$, Niech $P_W(A) = P(A | W)$. Wtedy (W, Σ_W, P_W) jest przestrzenią probabilistyczną. POWRÓT DO ZADANIA

Pytanie 3.2. Czy jest prawdą, że $P(A | W_1 \cup W_2) = P(A | W_1) + P(A | W_2)$?

Wskazówka. Nie. POWRÓT DO ZADANIA

Pytanie 3.3. Jak można osłabić założenia w twierdzeniu o prawdopodobieństwie całkowitym?

Wskazówka.

(i') $P(W_i) > 0$ dla każdego $i = 1, \dots, N$, $N \leq \infty$

(ii') $P(W_i \cap W_j) = 0$, dla wszystkich $i \neq j$,

(iii') $P(W_1 \cup \dots \cup W_N) = 1$.

POWRÓT DO ZADANIA

Pytanie 3.4. W przykładzie 3.2 obliczyć $P(K_1 | K_3)$.

Wskazówka. $\frac{23}{288}$. POWRÓT DO ZADANIA

Pytanie 3.5. Rzucamy kostką do gry. Wskazać dwa nietrywialne zdarzenia niezależne w przestrzeni probabilistycznej odpowiadającej temu eksperymentowi.

Wskazówka. Na przykład: $A = \{1, 2\}$, $B = \{2, 3, 4\}$. POWRÓT DO ZADANIA

Pytanie 3.6. Uogólnić sytuację opisaną jako schemat Bernoulliego przy założeniu, że każde doświadczenie może mieć trzy różne wyniki

Wskazówka. Niech $\Omega = \{1, 2, 3\}$, $\Sigma = \mathcal{P}(\Omega)$, a miara P zadana jest przez warunki:

$$P(\{1\}) = p_1, \quad P(\{2\}) = p_2, \quad P(\{3\}) = p_3$$

gdzie $0 < p_i < 1$ spełniają $p_1 + p_2 + p_3 = 1$. Gdy mamy n doświadczeń, rozważamy przestrzeń $(\Omega^n, \Sigma^n, P^n)$.

Ustalmy takie k_i , $0 \leq k_i \leq n$ dla $i = 1, 2, 3$, że $k_1 + k_2 + k_3 = n$. Interesuje nas zbiór A składający się z ciągów, które zawierają dokładnie k_1 jedynek, k_2 dwójek, a więc także k_3 trójek. Chcemy obliczyć $P^n(A)$, gdzie $A = \{\omega = (\omega_1, \dots, \omega_n) \in \Omega^n : \sum_{s:\omega_s=i} 1 = k_i\}$.

Dla każdego $\omega = (\omega_1, \dots, \omega_n) \in A$ mamy $P^n(\{\omega\}) = p_1^{k_1} p_2^{k_2} p_3^{k_3}$.

Ponieważ zdarzenie A składa się z $\frac{n!}{k_1!k_2!k_3!} = \binom{n}{k_1} \binom{n-k_1}{k_2}$ elementów, to

$$P^n(A) = \frac{n!}{k_1!k_2!k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}.$$

POWRÓT DO ZADANIA

Pytanie 4.1. Niech F_1, \dots, F_k będą dystrybuantami, $\lambda_1, \dots, \lambda_k > 0$, $\lambda_1 + \dots + \lambda_k = 1$. Czy $F = \lambda_1 F_1 + \dots + \lambda_k F_k$ jest dystrybuantą?

Wskazówka. Tak. POWRÓT DO ZADANIA

Pytanie 4.2. Q jest rozkładem dyskretnym zadany przez ciągi: $\{x_i\} = (1, 2, 3, \dots)$, $\{p_i\} = (\frac{1}{6}, \frac{5}{6} \cdot \frac{1}{6}, (\frac{5}{6})^2 \cdot \frac{1}{6}, \dots)$, patrz wzór (4.5). Obliczyć $Q(A)$, gdzie A jest zbiorem liczb parzystych.

Wykazać, że zbiory liczb parzystych i liczb pierwszych są zdarzeniami zależnymi.

Wskazówka. $Q(A) = \frac{5}{11}$. Niech B będzie zbiorem liczb pierwszych. Biorąc cztery kolejne liczby pierwsze, możemy stwierdzić, że $Q(B) > 0.3908214735$, więc $Q(A)Q(B) > 0.1776461243$. Tymczasem $Q(A \cap B) = Q(\{2\}) = \frac{5}{36} \approx 0.1388888889$. POWRÓT DO ZADANIA

Pytanie 4.3. Wskazać rozkład ciągły, którego gęstość jest dodatnia we wszystkich punktach $x \in \mathbb{R}$.

Wskazówka. Można wziąć taką funkcję dodatnią, np. ciągłą, g że $\int_{-\infty}^{\infty} g(x) dx < \infty$. Gęstością jest wtedy f :

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

POWRÓT DO ZADANIA

Pytanie 4.4. Wykazać, że dla dowolnej dystrybuanty zbiór punktów nieciągłości jest co najwyżej przeliczalny.

Wskazówka. W istocie jest to uogólnienie uwagi 4.7. POWRÓT DO ZADANIA

Pytanie 4.5. Niech F będzie dystrybuantą, a $h : \mathbb{R} \rightarrow \mathbb{R}$ funkcją. Zaproponować założenia o h gwarantujące, że $F \circ h$ jest dystrybuantą.

Wskazówka. h rosnąca, $\lim_{x \rightarrow -\infty} h(x) = -\infty$, $\lim_{x \rightarrow \infty} h(x) = \infty$. POWRÓT DO ZADANIA

Pytanie 4.6. Wskazać dystrybuantę rozkładu jednostajnego na:

(a) przedziale $(-2, 2)$, (b) sumie przedziałów $(-4, -2)$, $(2, 4)$.

Wskazówka.

POWRÓT DO ZADANIA

 Ad (a).

$$F(x) = \begin{cases} 0, & x < -2 \\ \frac{x+2}{4}, & -2 \leq x < 2 \\ 1, & 2 \leq x. \end{cases}$$

Ad (b).

$$F(x) = \begin{cases} 0, & x < -4 \\ \frac{x+4}{4}, & -4 \leq x < -2 \\ \frac{1}{2}, & -2 \leq x < 2 \\ \frac{x}{4}, & 2 \leq x < 4 \\ 1, & 4 \leq x. \end{cases}$$

Pytanie 5.1. Rzucono trzema kostkami i podano informację, czy, i na których kostkach wypadła „6”. Na odpowiedniej przestrzeni probabilistycznej podać przykłady funkcji, które są oraz które nie są zmiennymi losowymi.

Wskazówka. $\Omega = \{1, \dots, 6\}^3$, $\Sigma = \sigma(A_0, A_1, A_2, A_3)$, gdzie $A_i = \{\omega : \omega_i = 6\}$ dla $i = 1, 2, 3$, $A_0 = \Omega \setminus (A_1 \cup A_2 \cup A_3)$. Zmienną losową na przykład jest: S – suma „6”, które wypadły na kostkach, X określona jako: $X = 1$, gdy na trzeciej kostce wypadła „6”, $X = 0$ w przeciwnym przypadku. Zmienną losową nie jest, na przykład: S – suma uzyskanych oczek, X – liczba uzyskanych jedynek. POWRÓT DO ZADANIA

Pytanie 5.2. Salę oświetla 150 żarówek: po 15 żarówek w 10 rzędach. Wiadomo, że prawdopodobieństwo zepsucia się pojedynczej żarówki w ciągu nachodzącego tygodnia wynosi $p = 0.06$. Wiemy też, że żarówki psują się niezależnie od siebie. Niech X oznacza liczbę rzędów, w których po tygodniu świeci co najmniej 14 żarówek. Wskazać rozkład X .

Wskazówka. $B(10, s)$, gdzie $s = (1-p)^{15} + 15p(1-p)^{14} \approx 0.7510544178$. POWRÓT DO ZADANIA

Pytanie 5.3. W przykładzie 3.2: X , odpowiednio Y , oznaczają liczbę kolejnych dni, w których nieprzerwanie sprząta Kaja, odpowiednio Leon, licząc od dnia zawarcia umowy. Znaleźć rozkład X oraz rozkład Y . Czy zmienne te są niezależne?

Wskazówka. X ma rozkład $B(1, \frac{1}{2})$. Y ma rozkład dany przez ciągi wartości $0, 1, 2, 3, \dots$ oraz prawdopodobieństw $\frac{1}{2}, \frac{1}{2} \cdot \frac{1}{6}, \frac{1}{2} \cdot \frac{5}{6} \cdot \frac{1}{6}, \frac{1}{2} \cdot (\frac{5}{6})^2 \cdot \frac{1}{6}, \dots$, porównaj wzór (4.5). Zmienne są zależne, bo na przykład: $P(X=0, Y=0) = 0$, $P(X=0) = P(Y=0) = \frac{1}{2}$. POWRÓT DO ZADANIA

Pytanie 5.4. Niech $G \subset \mathbb{R}^2$ będzie zbiorem otwartym o mierze skończonej, (X, Y) wektorem losowym o rozkładzie $U(G)$. Podać przykład takiego zbioru G , że X, Y są: (a) niezależne, (b) zależne.

Wskazówka. Ad (a). G jest iloczynem kartezjańskim. Ad (b). G jest trójkątem. POWRÓT DO ZADANIA

Pytanie 5.5. Dane są dwie zmienne losowe niezależne X, Y o rozkładzie $U(0, 1)$ każda. Znajdź dystrybuantę i gęstość zmiennych losowych $\min(X, Y)$, $\max(X, Y)$. Czy te zmienne są niezależne? Wskazać dystrybuantę rozkładu $P_{\min(X, Y)|X=a}$ dla $a \in (0, 1)$.

Wskazówka. $F_{\min(X, Y)}(x) = 1 - (1-x)^2$, $f_{\min(X, Y)}(x) = 2 - 2x$,
 $F_{\max(X, Y)}(x) = x^2$, $f_{\max(X, Y)}(x) = 2x$,

$P_{\min(X, Y)|X=a}$ ma dystrybuantę $F: F(x) = \begin{cases} 0 & \text{dla } x < 0, \\ x & \text{dla } 0 \leq x < a, \\ 1 & \text{dla } a \leq x. \end{cases}$ POWRÓT DO ZADANIA

Pytanie 5.6. W schemacie klasycznym, gdy zbiór $\Omega = \{1, \dots, 6\}$, definiujemy zmienne losowe m, M jako $m(i, j) = \min(i, j)$, $M(i, j) = \max(i, j)$. Znaleźć: rozkład wektora losowego (m, M) oraz rozkład $P_{M|m=1}$.

Wskazówka.

| | | | | | | |
|-----------------|---|---|---|---|---|---|
| $m \setminus M$ | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| 2 | 0 | 1 | 2 | 2 | 2 | 2 |
| 3 | 0 | 0 | 1 | 2 | 2 | 2 |
| 4 | 0 | 0 | 0 | 1 | 2 | 2 |
| 5 | 0 | 0 | 0 | 0 | 1 | 2 |
| 6 | 0 | 0 | 0 | 0 | 0 | 1 |

$$P_{(m,M)}(\{i,j\}) = \begin{cases} \frac{1}{36} & \text{dla } i = j, \\ \frac{1}{18} & \text{dla } i < j, \\ 0 & \text{dla } i > j. \end{cases}$$

$P_{M|m=1}$ jest dany przez ciągi 1, 2, 3, 4, 5, 6 oraz $\frac{1}{11}, \frac{2}{11}, \dots, \frac{2}{11}$. POWRÓT DO ZADANIA

Pytanie 6.1. Podać przykłady zmiennych losowych o rozkładzie dyskretnym oraz o rozkładzie ciągłym, dla których nadzieja matematyczna nie jest skończona.

Wskazówka. $P(X = n) = \frac{\pi^2}{6n^2}, n = 1, 2, 3, \dots$

Gęstość f : $f(x) = 0$ dla $x < 1$, $f(x) = \frac{1}{x^2}$ dla $x \geq 1$. POWRÓT DO ZADANIA

Pytanie 6.2. Podać przykłady zmiennych losowych o rozkładzie dyskretnym oraz o rozkładzie ciągłym, dla których nadzieja matematyczna jest skończona, ale wariancja jest nieskończona.

Wskazówka. $P(X = n) = \frac{1}{\zeta(3)} \frac{1}{n^3}, n = 1, 2, 3, \dots$, gdzie $\zeta(3) = \sum_{n=1}^{\infty} \frac{1}{n^3}$.

Gęstość f : $f(x) = 0$ dla $x < 1$, $f(x) = \frac{2}{x^3}$ dla $x \geq 1$. POWRÓT DO ZADANIA

Pytanie 6.3. Dane są niezależne zmienne losowe o rozkładzie $U(0, 1)$ każda. Obliczyć $E(\min(X, Y))$.

Wskazówka. Niech $Z = \min(X, Y)$.

Sposób 1. $F_Z(z) = 1 - (1 - z)^2, f_Z(z) = 2 - 2z, E(Z) = \int_0^1 z f_Z(z) dz = \frac{1}{3}$.

Sposób 2. $E(Z) = \int_{[0,1]^2} \min(x, y) d(x, y) = \frac{1}{3}$. POWRÓT DO ZADANIA

Pytanie 6.4. Przeprowadzić dowód twierdzenia 6.17 w przypadku, gdy (X, Y) ma rozkład dyskretny na zbiorze skończonym.

Wskazówka. X ma rozkład zadany przez ciągi: $x_1, x_2, \dots, x_n, p_1, p_2, \dots, p_n$ (czyli $P(X = x_i) = p_i$) zaś Y ma rozkład zadany przez ciągi: $y_1, y_2, \dots, y_m, q_1, q_2, \dots, q_m$ (czyli $P(Y = y_j) = q_j$).

Wtedy rozkład łączny wektora losowego (X, Y) ma rozkład skupiony w punktach (x_i, y_j) , a z niezależności wynika, że $P((X, Y) = (x_i, y_j)) = P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) = p_i q_j$.

Sposób 1. XY przyjmuje wartości z_k . przy czym $P(XY = z_k) = \sum_{i,j:z_k=x_i y_j} p_i q_j$. Mamy więc

$$E(XY) = \sum_k z_k P(XY = z_k) = \sum_{i,j} x_i y_j p_i q_j = \sum_i x_i p_i \sum_j y_j q_j = E(X)E(Y).$$

Sposób 2. Biorąc g jako $g(x, y) = xy$ mamy: $E(XY) = \sum_{i,j} x_i y_j p_i q_j = E(X) \cdot E(Y)$. POWRÓT DO ZADANIA

Pytanie 6.5. Dana jest taka funkcja $f: [a, b] \rightarrow \mathbb{R}$, że $J = \int_a^b f(x) dx < \infty$. Wskazać taką zmienną losową X , że $J = E(X)$.

Wskazówka. $J = E(X)$, gdzie $X = (b-a)f(U)$, U – zmienna losowa o rozkładzie $U(a, b)$. POWRÓT DO ZADANIA

Pytanie 6.6. Zmienna losowa X ma rozkład $P(X = i) = 2^{-i}$ dla $i = 1, 2, 3, \dots$, a funkcja g jest dana wzorem; $g(i) = (-1)^{i+1} \frac{2^i}{i}$. Czy istnieje $E(g(X))$? Czy jest zbieżny szereg $\sum_{i=1}^{\infty} g(i)P(X = i)$?

Wskazówka. $\sum_{i=1}^{\infty} g(i)^+ P(X = i) = \infty, \sum_{i=1}^{\infty} g(i)^- P(X = i) = \infty$, więc $E(g(X))$ nie istnieje.

$\sum_{i=1}^{\infty} g(i)P(X = i) = \sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i} = \ln 2$ POWRÓT DO ZADANIA

Pytanie 7.1. Wykazać, że dla zmiennej losowej X mającej skończoną nadzieję matematyczną m i dla każdego dodatniego $\varepsilon > 0$

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

Wskazówka. Pierwsza część nierówności Czebyszewa zastosowana do $|X|$. POWRÓT DO ZADANIA

Pytanie 7.2. Niech $m = E(X)$, $\sigma^2 = D^2(X)$, $a < m < b$. Oszacować z góry $P(X \leq a)$ i z dołu $P(X < b)$.

Wskazówka. $\frac{\sigma^2}{(m-a)^2} \cdot 1 - \frac{\sigma^2}{(b-m)^2}$. POWRÓT DO ZADANIA

Pytanie 7.3. Niech X ma rozkład o gęstości f : $f(x) = 0$ dla $x < 0$ oraz $f(x) = e^{-x}$ dla $x \geq 0$. Dla $\varepsilon > 0$ obliczyć $P(X \geq \varepsilon)$ i porównać otrzymany wynik z oszacowaniem wynikającym z nierówności Czebyszewa.

Wskazówka. $E(X) = 1$, $P(X \geq \varepsilon) = e^{-\varepsilon} < \frac{1}{\varepsilon}$. POWRÓT DO ZADANIA

Pytanie 7.4. Sprawdzić, że w przypadku rozkładu jednostajnego reguła 3σ się trywializuje.

Wskazówka. Gdy $P_X = U(a, b)$, to $3\sigma = \sqrt{3} \cdot \frac{b-a}{2}$ i przedział $(a, b) \subset (m - 3\sigma, m + 3\sigma)$. POWRÓT DO ZADANIA

Pytanie 7.5. Niech $f : [0, 1] \rightarrow \mathbb{R}$ będzie funkcją ciągłą, X_1, X_2, X_3, \dots będzie ciągiem niezależnych zmiennych losowych o rozkładzie $B(1, p)$ każda, $S_n = X_1 + \dots + X_n$. Wykazać, że

$$\lim_{n \rightarrow \infty} E\left(f\left(\frac{S_n}{n}\right) - f(p)\right) = 0,$$

przy czym zbieżność ta jest jednostajna względem p .

Wskazówka. Niech $K = \sup |f|$. Ustalmy $\eta > 0$, $\varepsilon > 0$.

Niech $\Omega_1 = \{|\frac{S_n}{n} - p| < \varepsilon\}$, $\Omega_2 = \{|\frac{S_n}{n} - p| \geq \varepsilon\}$ będzie podziałem Ω .

$$\begin{aligned} \left|E\left(f\left(\frac{S_n}{n}\right) - f(p)\right)\right| &= \left|\int_{\Omega} f\left(\frac{S_n}{n}\right) - f(p) dP\right| \leq \int_{\Omega} \left|f\left(\frac{S_n}{n}\right) - f(p)\right| dP \\ &\leq \int_{\Omega_1} \left|f\left(\frac{S_n}{n}\right) - f(p)\right| dP + \int_{\Omega_2} \left|f\left(\frac{S_n}{n}\right) - f(p)\right| dP \\ &\leq \sup_{|x| \leq \varepsilon} \{|f(p+x) - f(p)|\} + 2KP \left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right). \end{aligned}$$

Ponieważ f jest jednostajnie ciągła, to pierwszy składnik jest mniejszy od $\eta/2$ dla dostatecznie małego ε , natomiast drugi składnik na podstawie nierówności Czebyszewa zmierza do 0. POWRÓT DO ZADANIA

Pytanie 7.6. Wskazać ciąg wielomianów, które jednostajnie aproksymują daną funkcję ciągłą f na przedziale $[0, 1]$.

Wskazówka. $W_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$. POWRÓT DO ZADANIA

Pytanie 8.1. Dwóch graczy wykonuje 10 rzutów kostką. Jakie jest prawdopodobieństwo tego, że obydwaj otrzymają tyle samo „6”?

Wskazówka. $\sum_{k=0}^n \binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k} \approx 0.24209$ dla $n = 10$. POWRÓT DO ZADANIA

Pytanie 8.2. Czy/kiedy suma dwóch niezależnych zmiennych losowych o rozkładach dwumianowych ma rozkład dwumianowy?

Wskazówka. Jeżeli $P_X = B(n, p)$, $P_Y = B(m, q)$, $p = q$, to $P_{X+Y} = B(n+m, p)$. POWRÓT DO ZADANIA

Pytanie 8.3. Ile rodzynek podczas wyrabiania ciasta trzeba średnio przeznaczyć na bułeczkę, aby losowo wybrana bułeczka z prawdopodobieństwem 0,95 lub większym zawierała co najmniej jedną rodzynekę? Jakie wtedy będzie prawdopodobieństwo tego, że losowo wybranej bułeczce będzie co najmniej 5 rodzynek?

Wskazówka. X – liczba rodzynek w bułeczce ma rozkład Poissona P_λ , gdyż jest dużo rodzynek (doświadczeń) i małe prawdopodobieństwo, że jedna z nich trafi do danej bułeczki (sukces). Trzeba tak dobrać λ , żeby $P(X \geq 1) \geq 0.95$.

Mamy kolejno $1 - e^{-\lambda} \geq 0.95$, $\lambda \geq 2.995732274$. $P(X \geq 5) \geq 0.1840201545$. POWRÓT DO ZADANIA

Pytanie 8.4. Przeprowadzić dowód twierdzenia 8.11.

Wskazówka. $F_{X+Y}(z) = P(X + Y \leq z) = \iint_A f_{(X,Y)}(x, y) d(x, y) = \iint_A f_X(x)f_Y(y) d(x, y)$, gdzie $A = \{(x, y) : x + y \leq z\}$. Stosujemy zmianę zmiennych: $s = x + y$, $t = x$. Mamy więc:

$$F_{X+Y}(z) = \iint_{\mathbb{R} \times (-\infty, z)} f_X(t)f_Y(s-t) d(s, t) = \int_{-\infty}^z \left(\int_{\mathbb{R}} f_X(t)f_Y(s-t) dt \right) ds.$$

$$f_{X+Y}(z) = \frac{d}{dz}F_{X+Y}(z) = \int_{\mathbb{R}} f_X(t)f_Y(z-t) dt.$$

POWRÓT DO ZADANIA

Pytanie 8.5. Znaleźć rozkład sumy niezależnych zmiennych losowych o rozkładzie $U(0, a)$ każda.

Wskazówka.

$$f_{X+Y}(z) = I_{(0,2a)} \left(\frac{1}{a} - \left| \frac{1}{a} - \frac{z}{a^2} \right| \right).$$

POWRÓT DO ZADANIA

Pytanie 8.6. Wykazać, że minimum dwóch niezależnych zmiennych losowych o rozkładach wykładniczych ma rozkład wykładniczy.

Wskazówka. $F_{\min(X,Y)}(z) = P(\min(X, Y) \leq z) = 1 - P(X > z, Y > z) = 1 - (1 - F_X(z))(1 - F_Y(z)) = 1 - e^{-\lambda z}e^{-\mu z} = 1 - e^{-(\lambda+\mu)z}$. POWRÓT DO ZADANIA

Pytanie 9.1. Czy suma niezależnych zmiennych losowych o rozkładach normalnych ma rozkład normalny? Czy założenie niezależności jest istotne?

Wskazówka. Tak, można skorzystać z twierdzenia 8.11. Jest istotne: $X + (-X) = 0$. POWRÓT DO ZADANIA

Pytanie 9.2. Sformułować odpowiednik reguły 1.96, gdy $\alpha = 0.01$.

Wskazówka. Reguła 2.58: Jeżeli zmienna losowa X ma rozkład normalny $N(m, \sigma)$, to

$$P(X \in (m - 2.58\sigma, m + 2.58\sigma)) \approx 0.99.$$

POWRÓT DO ZADANIA

Pytanie 9.3. Niech X_1, X_2 będą niezależnymi zmiennymi losowymi o rozkładzie standardowym normalnym każda. Czy zmienne losowe $X_1 - X_2, X_1 + X_2$ są niezależne? Czy założenie normalności jest istotne?

Wskazówka. Są niezależne: można wykorzystać twierdzenie 5.33, aby otrzymać gęstość wektora $(X_1 - X_2, X_1 + X_2)$, a następnie stwierdzić, że jest ona iloczynem gęstości różnicy i sumy.

Założenie normalności jest istotne. Niech X_1, X_2 będą niezależne i mają rozkład $B(1, \frac{1}{2})$ każda. Gdy suma = 2, to różnica = 0. Formalnie: $P(X_1 + X_2 = 2, X_1 - X_2 = 0) = \frac{1}{4}$, $P(X_1 + X_2 = 2)P(X_1 - X_2 = 0) = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$.

POWRÓT DO ZADANIA

Pytanie 9.4. Partia ABC wie, że ma poparcie nie większe niż 10%. Zamawia sondaż, aby stwierdzić czy zdobędzie co najmniej 5% poparcie. Chciałaby mieć 99% pewności, że wynik sondażu oddaje prawdziwe preferencje wyborców z dopuszczalnym błędem nie większym niż 2%. Jak duża powinna być próbka ankietowanych osób?

Wskazówka.

$$n \geq \left(\frac{\Phi^{-1}(1 - \alpha)}{b} \right)^2 \frac{9}{100} \approx 541.189443051267.$$

POWRÓT DO ZADANIA

Pytanie 9.5. Pewna agencja prowadzi rekrutację pracowników w kilku różnych krajach, które stosują różne systemy punktowania, niemniej trudność używanych testów jest porównywalna. Agencja otrzymuje listy punktów uzyskanych przez kandydatów ze wszystkich krajach i na tej podstawie chce wybrać 100 najlepszych kandydatów. W jaki sposób może postąpić agencja, aby wybór był racjonalny?

Wskazówka. Każdy kraj przedstawia listę, powiedzmy x_1, x_2, \dots, x_k , gdzie x_i jest oceną punktową i -tego kandydata. Na podstawie tej listy można obliczyć średnią: $\bar{x} = \frac{1}{k} \sum_{i=1}^k x_i$ oraz odchylenie standardowe z próby¹: $s = \sqrt{\frac{1}{k} \sum_{i=1}^k (x_i - \bar{x})^2}$. Można wyznaczyć znormalizowane punkty: $z_i = \frac{x_i - \bar{x}}{s}$. Teraz można połączyć tak uzyskane listy ze wszystkich krajów, a następnie uporządkować malejąco całą listę i wybrać pierwszych 1000 kandydatów.

POWRÓT DO ZADANIA

Pytanie 10.1. Wykazać uwagę 10.1.

Wskazówka. Niech $B \in \mathcal{B}(\mathbb{R}^n)$. Wtedy $X^{-1}(B) = (X^{-1}(B) \cap \{X = Y\}) \cup (X^{-1}(B) \cap \{X \neq Y\})$. Stąd $P(X^{-1}(B)) = P(X^{-1}(B) \cap \{X = Y\})$. Podobnie: $P(Y^{-1}(B)) = P(Y^{-1}(B) \cap \{X = Y\})$. Wystarczy zauważyć, że $X^{-1}(B) \cap \{X = Y\} = Y^{-1}(B) \cap \{X = Y\}$.

POWRÓT DO ZADANIA

Pytanie 10.2. Wykazać, że granica stochastycznie zbieżnego ciągu zmiennych losowych jest wyznaczona jednoznacznie p.w..

Wskazówka. Załóżmy, że $X_n \xrightarrow{s} X$ oraz $X_n \xrightarrow{s} Y$. Wykażemy, że $P(X = Y) = 1$. Ustalmy $\varepsilon > 0$. Z nierówności trójkąta, $|X - Y| \leq |X - X_n| + |Y - X_n|$, widzimy, że $\{|X - Y| > \varepsilon\} \subset \{|X - X_n| > \frac{\varepsilon}{2}\} \cup \{|Y - X_n| > \frac{\varepsilon}{2}\}$. Stąd $P(|X - Y| > \varepsilon) \leq P(|X - X_n| > \frac{\varepsilon}{2}) + P(|Y - X_n| > \frac{\varepsilon}{2})$. Gdy $n \rightarrow \infty$ prawa strona zmierza do 0, a stąd. $P(|X - Y| > \varepsilon) = 0$. Czyli dla każdego $\varepsilon > 0$ $P(|X - Y| \leq \varepsilon) = 1$. $\{X = Y\} = \bigcap_{\varepsilon > 0} \{|X - Y| \leq \varepsilon\}$.

POWRÓT DO ZADANIA

Pytanie 10.3. Niech X_1, X_2, X_3, \dots będzie ciągiem takich niezależnych zmiennych losowych, że $P_{X_i} = B(1, p_i)$ dla $i = 1, 2, 3, \dots$. Wykazać, że:

$$(1) X_n \xrightarrow{s} 0 \iff \lim_{n \rightarrow \infty} p_n = 0.$$

$$(2) X_n \xrightarrow{1} 0 \iff \sum_{n=1}^{\infty} p_n < \infty.$$

Wskazówka. Ad (1). Niech $\varepsilon < 1$. Wtedy $\{|X_n| \geq \varepsilon\} = \{X_n = 1\}$.

Ad (2). $\{X_n \rightarrow 0\} = \Omega \setminus \bigcap_N \bigcup_{n \geq N} \{X_n = 1\}$. Stosuje się lemat Borela–Cantellego do zdarzeń $A_n = \{X_n = 1\}$, gdyż $X_n \xrightarrow{1} 0 \iff P(A) = 0$.

POWRÓT DO ZADANIA

Pytanie 10.4. Niech $X_n \xrightarrow{s} X$ i niech $f: \mathbb{R} \rightarrow \mathbb{R}$ będzie funkcją jednostajnie ciągłą. Wykazać, że $f(X_n) \xrightarrow{s} f(X)$. Jak można osłabić założenia, gdy wiemy, że $X = c \in \mathbb{R}$?

Wskazówka. Ustalmy $\varepsilon > 0$ i weźmy takie $\delta > 0$, że $|x - y| \leq \delta$ implikuje $|f(x) - f(y)| \leq \varepsilon$. W takim razie

$$P(|X_n - X| \leq \delta) \leq P(|f(X_n) - f(X)| \leq \varepsilon)$$

i wystarczy przejść z n do nieskończoności.

Z dowodu widać, że gdy $X = c$, lub nawet bardziej ogólnie, gdy X przyjmuje wartości w zbiorze zwartym, wystarczy założyć, że f jest ciągła na tym zbiorze.

POWRÓT DO ZADANIA

Pytanie 10.5. Niech $X_n \xrightarrow{1} X$ i niech $f: \mathbb{R} \rightarrow \mathbb{R}$. Zaproponować założenie dotyczące f gwarantujące, że $f(X_n) \xrightarrow{1} f(X)$.

Wskazówka. Załóżmy, że $P_X(B) = 0$, gdzie $B = \{a \in \mathbb{R} : f \text{ nie jest ciągła w } a\}$. Dla dowodu weźmy zbiór $A = \{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}$. Wtedy dla $\omega \in A \cap (\Omega \setminus X^{-1}(B))$ zachodzi $f(X_n(\omega)) \rightarrow f(X(\omega))$, natomiast $P(A \cap (\Omega \setminus X^{-1}(B))) = 1$.

POWRÓT DO ZADANIA

Pytanie 10.6. Niech X_1, X_2, X_3, \dots będzie ciągiem niezależnych zmiennych losowych o takim samym rozkładzie. Wykazać, że jeżeli ciąg $\frac{S_n}{n}$ jest zbieżny z prawdopodobieństwem 1, to zmienne X_i mają skończoną nadzieję matematyczną m i $\frac{S_n}{n} \xrightarrow{1} m$.

Wskazówka. Kolejno mamy: $\frac{X_n}{n} = \frac{S_n}{n} - \frac{S_{n-1}}{n} \xrightarrow{1} 0$.

$$\{\frac{X_n}{n} \rightarrow 0\} \subset \bigcup_N \bigcap_{n \geq N} \{|X_n| < n\} = \Omega \setminus \bigcap_N \bigcup_{n \geq N} \{|X_n| \geq n\}.$$

W związku z tym: $P(\bigcap_N \bigcup_{n \geq N} \{|X_n| \geq n\}) = 0$, więc z lematu Borela–Cantellego $\sum_{n=1}^{\infty} P(|X_n| \geq n) < \infty$. Ponieważ nasze zmienne losowe mają ten sam rozkład, to $P(|X_n| \geq n) = P(|X_1| \geq n)$, więc $\sum_{n=1}^{\infty} P(|X_1| \geq n) < \infty$. Z lematu 10.18 otrzymujemy wiadomość, że $E(|X_1|)$ jest skończona, a więc także $m = E(X_1)$ jest skończona. MPWL dla i.i.d. kończy dowód.

POWRÓT DO ZADANIA

Pytanie 11.1. Czy następujące rodziny rozkładów spełniają warunek Prochorowa?

1. $\{N(m, 1) : m \in \mathbb{R}\}$,
2. $\{N(0, \sigma) : 0 < \sigma < 1\}$,

¹Z pewnych względów, patrz rozdział 12, stosuje się czasem trochę inny wzór: $s = \sqrt{\frac{1}{k-1} \sum_{i=1}^k (x_i - \bar{x})^2}$, co jest istotne, gdy k nie jest dużą liczbą.

3. $\{U(a, b) : a < b < 100\}$.

Odpowiedź uzasadnić.

Wskazówka. Ad 1. Nie. Ustalmy jakikolwiek zbiór zwarty. Jest on zawarty w jakimś przedziale postaci $[-N, N]$. Można obliczyć $P([-N, N])$ i zobaczyć, że dla dużych m jest ono dla każdego ustalonego N małe.

Ad 2. Tak. Rozumowanie podobne jak wyżej.

Ad 3. Nie. Rozumowanie podobne jak wyżej, tylko mniej liczenia. POWRÓT DO ZADANIA

Pytanie 11.2. Nie korzystając z MPWL, wykazać następujące twierdzenie (Chinczyna).

Jeżeli X_1, X_2, X_3, \dots są i.i.d. i mają skończoną nadzieję matematyczną m , to $\frac{S_n}{n} \xrightarrow{s} m$.

Wskazówka. Postępujemy podobnie jak w dowodzie CTG. Niech h będzie funkcją charakterystyczną zmiennej X_i . Wtedy:

$$h(u) = 1 + imu + o(u),$$

$$h_{\frac{S_n}{n}}(u) = h\left(\frac{u}{n}\right)^n = \left(1 + im\frac{u}{n} + o\left(\frac{1}{n}\right)\right)^n \rightarrow e^{ium} = h_m(u).$$

Więc $\frac{S_n}{n} \xrightarrow{d} m$, więc także $\frac{S_n}{n} \xrightarrow{s} m$. POWRÓT DO ZADANIA

Pytanie 11.3. Przeprowadzić dowód CTG dla średnich.

Wskazówka. Tak samo jak dla sum. POWRÓT DO ZADANIA

Pytanie 11.4. Niech S_n oznacza sumę orłów uzyskanych w trakcie n rzutów monetą symetryczną. Niech $\varepsilon > 0$ będzie dowolną liczbą.

Obliczyć:

1. $\lim_{n \rightarrow \infty} P\left(\left|S_n - \frac{n}{2}\right| \geq \varepsilon\right),$
2. $\lim_{n \rightarrow \infty} P\left(\left|S_n - \frac{n}{2}\right| \geq \varepsilon n\right),$
3. $\lim_{n \rightarrow \infty} P\left(\left|S_n - \frac{n}{2}\right| \geq \varepsilon\sqrt{n}\right).$

Wskazówka. Ad 1. $P\left(\left|S_n - \frac{n}{2}\right| \geq \varepsilon\right) = 2 - 2\Phi\left(\frac{\varepsilon}{2\sqrt{n}}\right) \rightarrow 1.$

Ad 2. $P\left(\left|S_n - \frac{n}{2}\right| \geq \varepsilon n\right) = 2 - 2\Phi\left(\frac{\varepsilon\sqrt{n}}{2}\right) \rightarrow 0.$

Ad 3. $P\left(\left|S_n - \frac{n}{2}\right| \geq \varepsilon\sqrt{n}\right) = 2 - 2\Phi\left(\frac{\varepsilon}{2}\right).$ POWRÓT DO ZADANIA

Pytanie 11.5. Niech S_n oznacza sumę orłów uzyskanych w trakcie n rzutów monetą symetryczną. Sformułować jako twierdzenie następujące spostrzeżenia: Gdy wykona się dostatecznie dużo rzutów, to różnica między liczbą uzyskanych orłów i reszek będzie tak wielka jak chcemy, natomiast ich iloraz będzie coraz bliższy 1.

Wskazówka. Twierdzenie. Dla dowolnego $\varepsilon > 0$:

$$1. \lim_{n \rightarrow \infty} P(|S_n - (n - S_n)| \geq \varepsilon) = 1,$$

$$2. \lim_{n \rightarrow \infty} P\left(\left|\frac{n - S_n}{S_n} - 1\right| \geq \varepsilon\right) = 0.$$

Dowód. Punkt 1. jest konsekwencją punktu 1. w poprzednim pytaniu.

$$2. P\left(\left|\frac{n - S_n}{S_n} - 1\right| \geq \varepsilon\right) = 1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{2 - \varepsilon}\right) + \Phi\left(\frac{-\varepsilon\sqrt{n}}{2 + \varepsilon}\right) \rightarrow 0. \quad \square$$

POWRÓT DO ZADANIA

Pytanie 11.6. Wykazać twierdzenie o jednoznaczności rozkładu w przypadku funkcji tworzących.

Wskazówka. Niech $\pi_X = \pi_Y$ będą funkcjami tworzącymi zmiennych losowych X, Y . Wtedy zachodzi też równość $h_X(u) = \pi_X(e^{iu}) = \pi_Y(e^{iu}) = h_Y(u)$ dla $u \in \mathbb{R}$, gdzie h_X, h_Y są odpowiednimi funkcjami charakterystycznymi. Z twierdzenia o jednoznaczności dla funkcji charakterystycznych oznacza to, że $P_X = P_Y$. POWRÓT DO ZADANIA

Pytanie 12.1. Liczby 0.657, 0.773, 0.801, 0.501, 0.123, 0.202 wylosowano z rozkładu $U(0, 1)$. Na tej podstawie wygenerować liczby z rozkładu $B(3, 1/2)$.

Wskazówka. 2, 2, 2, 2, 0, 1. Należy podzielić odcinek $[0, 1]$ punktami $1/8, 4/8, 7/8$. POWRÓT DO ZADANIA

Pytanie 12.2. Niech X_1, X_2, X_3, \dots będą i.i.d. z rozkładu $U(0, a)$, $a > 0$. Zbadać własności estymatora parametru a : $M_n = \max(X_1, \dots, X_n)$.

Wskazówka. Zgodność. Ustalmy $0 < \varepsilon < a$ i zdefiniujmy zdarzenia $A_n = \{M_n \leq a - \varepsilon\}$. Widać, że $P(A_n) = (\frac{a-\varepsilon}{a})^n$. Stąd $\sum_{n=1}^{\infty} P(A_n) < \infty$, co z lematu Borela-Cantellego oznacza, że $P\left(\bigcap_{N=1}^{\infty} \bigcup_{n \geq N} A_n\right) = 0$.

Inaczej:

$$P\left(\bigcap_{\varepsilon > 0} \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{M_n > a - \varepsilon\}\right) = 1,$$

czyli $M_n \xrightarrow{1} a$.

Nieobciążoność.

$$F_{M_n}(x) = \left(\frac{x}{a}\right)^n, \quad f_{M_n}(x) = n \frac{x^{n-1}}{a^n}, \quad E(M_n) = \frac{n}{n+1}a.$$

Estymator jest obciążony.

Nieobciążonym i zgodnym estymatorem parametru a jest więc: $\frac{n+1}{n}M_n$. POWRÓT DO ZADANIA

Pytanie 12.3. Dana jest funkcja $f : [a, b] \rightarrow \mathbb{R}$, o której wiadomo, że spełnia warunek $|f(x)| \leq M$ dla każdego $x \in [a, b]$. Jak można estymować $\int_a^b f(x) dx$, używając metody I z wykładu?

Wskazówka. Rozważyc funkcję $g = f + M$. POWRÓT DO ZADANIA

Pytanie 12.4. Aby obliczyć całkę $J = \int_{-1}^1 f(x) dx$, gdzie $0 < f(x) < 1$ zastosowano metodę I. Wylosowano 1000 punktów z prostokąta $[-1, 1] \times [0, 1]$ i okazało się, że 360 z nich leży pod wykresem funkcji. Na poziomie ufności 0.95 wskazać przedział ufności dla J .

Wskazówka. Przedział ufności jest postaci $(\bar{J}_n - \delta, \bar{J}_n + \delta)$, gdzie $\bar{J}_n = (b-a)c\bar{X}_n$, $\delta = (b-a)c\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$. Mamy $a = -1$, $b = 1$, $c = 1$, $n = 1000$, $\bar{X}_n = \frac{360}{1000} = 0.36$, $\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1.96$. Nie znamy σ , ale na podstawie próby znamy jego przybliżenie, mianowicie $\hat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X}_n)^2$. Ponieważ $X_i = X_i^2$, to $\hat{\sigma}_X^2 = \frac{360}{1000} - \left(\frac{360}{1000}\right)^2 = 0.230400$. W takim razie przedział ufności dla J , to: $[0.660498, 0.779501]$. POWRÓT DO ZADANIA

Pytanie 12.5. W celu estymacji całki $J = \int_a^b f(x) dx$ wylosowano n punktów x_i według rozkładu jednostajnego $U([a, b])$ i wyznaczono sumy $s = f(x_1) + \dots + f(x_n)$ oraz $s_{kw} = f(x_1)^2 + \dots + f(x_n)^2$. Wskazać przedział ufności dla J postaci (c, ∞) na poziomie ufności $1 - \alpha$.

Wskazówka. $c = (b-a) \left(\frac{s}{n} - \frac{\sqrt{n s_{kw} - s^2}}{n} \Phi^{-1}(1 - \alpha) \right)$. POWRÓT DO ZADANIA

Pytanie 12.6. Uzasadnić ostatnie zdanie w dowodzie twierdzenia 12.18.

Wskazówka. Wystarczy wykazać, że

$$\{f(X_t) \rightarrow m, \text{ gdy } t \rightarrow \infty\} \subset \{X_t \rightarrow A^*, \text{ gdy } t \rightarrow \infty\}.$$

Niech $\omega \in \{f(X_t(\omega)) \rightarrow m, \text{ gdy } t \rightarrow \infty\}$ i niech $x_t = X_t(\omega)$. Chcemy pokazać, że $\text{dist}(x_t, A^*) \rightarrow 0$. Gdyby tak nie było, to istniałby taki podciąg t_k oraz $\varepsilon > 0$, że

$$\|x_{t_k} - x\| \geq \varepsilon \text{ dla każdego } x \in A^*. \quad (21.1)$$

Ze zwartości A można wybrać podciąg t_{k_l} oraz punkt $x^* \in A$ takie, że $x_{t_{k_l}} \rightarrow x^*$. Wtedy z ciągłości f wiemy, że $f(x_{t_{k_l}}) \rightarrow f(x^*)$. Ale z naszego założenia: $f(x_{t_{k_l}}) \rightarrow m$. Czyli $m = f(x^*)$, co z określenia A^* oznacza, że $x^* \in A^*$, więc nierówność (21.1) oznacza, że $\|x_{t_{k_l}} - x^*\| \geq \varepsilon$. W granicy $\|x^* - x^*\| \geq \varepsilon$, co daje sprzeczność.

POWRÓT DO ZADANIA

Pytanie 13.1. Znaleźć rozkład warunkowy maksimum liczby oczek w rzucie parą symetrycznych kostek pod warunkiem, że na drugiej kostce wypadła liczba parzysta. Wyznacz jego nadzieję matematyczną.

Wskazówka. Określamy zmienne losowe M – maksimum oczek oraz Y : $Y = 0$, gdy na drugiej kostce jest liczba parzysta, $Y = 1$, gdy na drugiej kostce jest liczba nieparzysta. Interesuje nas rozkład $P_{M|Y=0}$. Wyznaczamy najpierw rozkład wektora losowego (Y, M) .

| | | | | | | |
|-----------------|------|------|------|------|------|------|
| $Y \setminus M$ | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 0 | 2/36 | 1/36 | 5/36 | 2/36 | 8/36 |
| 1 | 1/36 | 1/36 | 4/36 | 2/36 | 7/36 | 3/36 |

Następnie rozkład $P_{M|Y=0}$:

| | | | | | |
|---|------|------|------|------|------|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 2/18 | 1/18 | 5/18 | 2/18 | 8/18 |

Stąd $E(M|Y=0) = \frac{85}{18}$. POWRÓT DO ZADANIA

Pytanie 13.2. Wektor losowy (X, Y) ma rozkład jednostajny na trójkącie o wierzchołkach $(-1, 0)$, $(1, 0)$, $(0, 1)$. Wyznaczyć $E(X|Y=1/2)$, $E(Y|X=1/2)$.

Wskazówka. $E(X|Y=1/2) = 0$, $E(Y|X=1/2) = 1/4$. POWRÓT DO ZADANIA

Pytanie 13.3. Dana jest przestrzeń probabilistyczna (Ω, Σ, P) oraz zmienna losowa $X : \Omega \rightarrow \mathbb{R}$. Czy i jakie zawierania zachodzą pomiędzy $\sigma(X)$, $\sigma(X^2)$ oraz $\sigma(X^3)$: wypowiedzieć i udowodnić twierdzenie, podać odpowiedni przykład jeżeli zawieranie nie zachodzi.

Wskazówka. Wykażemy, że $\sigma(X^2) \subset \sigma(X)$, $\sigma(X) = \sigma(X^3)$.

Dowód. Niech $A \in \sigma(X^2)$. Wtedy istnieje taki zbiór borelowski B , że $A = X^{-2}(B) = \{\omega; X^2(\omega) \in B\} = \{\omega : X(\omega) \in g^{-1}(B)\} = X^{-1}(g^{-1}(B))$, gdzie $g(x) = x^2$. Ponieważ g jest funkcją ciągłą, więc funkcją borelowską, więc $g^{-1}(B)$ jest borelowski, więc $A \in \sigma(X)$, czyli $\sigma(X^2) \subset \sigma(X)$.

Podobnie $\sigma(X^3) \subset \sigma(X)$, tutaj $g(x) = x^3$. Podobnie $\sigma(X) \subset \sigma(X^3)$, tutaj $g(x) = \sqrt[3]{x}$.

Przykład. Niech X będzie zmienną losową o rozkładzie $P(X = -1) = P(X = 1) = 1/2$. Wtedy X^2 jest stałą równą 1. $\sigma(X)$ składa się z czterech elementów, więc nie jest zawarta w $\sigma(X^2)$ składającej się tylko z dwóch elementów. □

POWRÓT DO ZADANIA

Pytanie 13.4. Przeprowadzić dowód własności 7. w twierdzeniu 13.25.

Wskazówka. Podobnie jak dowód własności 8. (po zamianie ról przez strony). POWRÓT DO ZADANIA

Pytanie 13.5. Przeprowadzić dowód własności 2. w twierdzeniu 13.26.

Wskazówka. Krok 1. Niech $Z = \mathbb{1}_A$, gdzie $A \in \mathcal{A}$. Pokażemy, że prawa strona spełnia warunki (M) oraz (C) wymagane od lewej strony. Warunek (M) wynika z faktu, że iloczyn funkcji mierzalnych jest mierzalny. Aby sprawdzić (M) ustalmy dowolne $B \in \mathcal{A}$. Ponieważ $A \cap B \in \mathcal{A}$ mamy $\int_B \mathbb{1}_A E(Y|\mathcal{A}) dP = \int_{A \cap B} E(Y|\mathcal{A}) dP = \int_{A \cap B} Y dP = \int_B \mathbb{1}_A Y dP$, co oznacza warunek (C). POWRÓT DO ZADANIA

Pytanie 13.6. Zakładamy, że $E(Y)$ jest skończona. Czy następujące zdania są prawdziwe? Odpowiedź uzasadnić.

(1) Jeżeli wektor losowy X ma rozkład dyskretny, to zmienna losowa $E(Y|X)$ ma rozkład dyskretny.

(2) Jeżeli wektor losowy X ma rozkład ciągły, to zmienna losowa $E(Y|X)$ ma rozkład ciągły.

Wskazówka. Ad (1). TAK. Wiemy, że $E(Y|X) = \alpha \circ X$, gdzie α jest funkcją borelowską. Z założenia istnieje zbiór K co najwyżej przeliczalny taki, że $P(X \in K) = 1$. Oczywiście $\alpha(K)$ jest co najwyżej przeliczalny, a $P(E(Y|X) \in \alpha(K)) \geq P(X \in K) = 1$, więc $E(Y|X)$ ma rozkład dyskretny.

Ad (2). NIE. Na przykład, gdy X, Y są niezależne, to $E(Y|X)$ jest stałą, więc nie ma rozkładu ciągłego.

POWRÓT DO ZADANIA

Pytanie 14.1. Wektor losowy (X, Y) ma rozkład o gęstości $cx\mathbb{1}_{[0,1]^2}$, gdzie $c \in \mathbb{R}$. Wyznaczyć kolejno: $E(X|Y)$, $E(Y|X)$, $D^2(X)$, $D^2(E(X|Y))$, $D^2(Y)$, $D^2(E(Y|X))$.

Wskazówka. Całka $\int_{[0,1]^2} f_{X,Y} d(x,y)$ musi być równa 1, stąd $c = 2$. Wyznaczamy rozkłady brzegowe i na tej podstawie $D^2(X) = \frac{1}{18}$, $D^2(Y) = \frac{1}{12}$. Wyliczamy $E(Y|X = x) = \frac{1}{2}$, $E(X|Y = y) = \frac{2}{3}$. Tak więc $E(Y|X) = \frac{1}{2}$, $E(X|Y) = \frac{2}{3}$ są stałymi, więc ich wariancje są równe 0. POWRÓT DO ZADANIA

Pytanie 14.2. Zmienna losowa X ma rozkład $U(0, 3)$. Niech $A_i = X^{-1}((i-1, i])$ dla $i = 1, 2, 3$.

Obliczyć $D^2(E(X|\sigma(A_1, A_2, A_3)))$.

Wskazówka. $E(X|\sigma(A_1, A_2, A_3))$ ma rozkład jednostajny dyskretny skupiony na zbiorze $\{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}\}$.

$$D^2(E(X|\sigma(A_1, A_2, A_3))) = \frac{2}{3}. \quad \text{POWRÓT DO ZADANIA}$$

Pytanie 14.3. Uzupełnić dowód twierdzenia 14.10.

(a) Dlaczego $E(Y|\mathcal{A}) \in \Delta$ pw.? (b) Dlaczego zachodzi nierówność (26)?

Wskazówka. Ad (a) Jeżeli $X \leq b$, to $E(X|\mathcal{A}) \leq b$.

Ad (b) Można wziąć $\mathcal{A} = \{\emptyset, \Omega\}$. POWRÓT DO ZADANIA

Pytanie 14.4. Wektor losowy (X, Y) ma rozkład jednostajny na półkolu $x^2 + y^2 \leq 1, y \geq 0$. Znaleźć: $D^2(Y), D^2(E(Y|X)), D^2(X), D^2(E(X|Y))$.

Wskazówka. $D^2(X) = \frac{1}{4}, E(X|Y) = 0, D^2(E(X|Y)) = 0, D^2(Y) = \frac{1}{4} - \frac{16}{9\pi^2}, E(Y|X) = \frac{1}{2}\sqrt{1-X^2}, D^2(E(Y|X)) = \frac{3}{16} - \frac{16}{9\pi^2}$. POWRÓT DO ZADANIA

Pytanie 14.5. Wyznaczyć $D^2(Y)$ oraz $D^2(E(Y|(X, Z)))$ dla zmiennych losowych określonych w przykładzie 14.9.

Wskazówka. Najpierw wyznaczamy: $E(Y^2|X = x, Z = 0) = \frac{x^2}{3}, E(Y^2|X = x, Z = 1) = \frac{x^2+x+1}{3}$. Stosując analogiczne rozumowanie jak w przykładzie 14.9, wyliczamy $E(Y^2)$ jako $E(E(Y^2|(X, Z))) = \frac{1}{9} + \frac{1}{2}p$. Mamy więc:

$$D^2(Y) = E(Y^2) - E(Y)^2 = \frac{1}{9} + \frac{1}{2}p - \left(\frac{1}{4} + \frac{p}{2}\right)^2 = \frac{7}{144} + \frac{1}{4}p - \frac{1}{4}p^2.$$

Wynik uzyskany w przykładzie 14.9 można zapisać w formie:

$$E(Y|(X, Z)) = \frac{X}{2}\mathbb{1}_{\{Z=0\}} + \frac{X+1}{2}\mathbb{1}_{\{Z=1\}} \quad \text{dla } 0 < x < 1.$$

Możemy więc policzyć:

$$\begin{aligned} E(E(Y|(X, Z))^2) &= E\left(\left(\frac{X}{2}\mathbb{1}_{\{Z=0\}} + \frac{X+1}{2}\mathbb{1}_{\{Z=1\}}\right)^2\right) \\ &= E\left(\left(\frac{X}{2}\mathbb{1}_{\{Z=0\}}\right)^2\right) + E\left(\left(\frac{X+1}{2}\mathbb{1}_{\{Z=1\}}\right)^2\right) + 2 \cdot 0 \\ &= \frac{1}{4}E(X^2)(1-p) + \frac{1}{4}E((X+1)^2)p = \frac{1}{4}\left(\frac{1}{3}(1-p) + \frac{7}{3}p\right) = \frac{1}{12} + \frac{1}{2}p. \\ D^2(E(Y|(X, Z))) &= \frac{3}{144} + \frac{1}{4}p - \frac{1}{4}p^2. \end{aligned}$$

POWRÓT DO ZADANIA

Pytanie 14.6. Przy standardowych oznaczeniach i założeniu, że $E(X) \in \mathbb{R}$, porównać

$$E(X|\mathcal{A})^+, E(X^+|\mathcal{A}) \quad \text{oraz podobnie} \quad E(X|\mathcal{A})^-, E(X^-|\mathcal{A}).$$

Wskazówka. Stosujemy nierówność Jensena do funkcji wypukłej $g(x) = \max(0, x)$ oraz $g(x) = -\max(0, -x)$.

$$E(X|\mathcal{A})^+ \leq E(X^+|\mathcal{A}), \quad E(X|\mathcal{A})^- \geq E(X^-|\mathcal{A}).$$

POWRÓT DO ZADANIA

Pytanie 15.1. Niech $a > 0$. Określamy: $X_0 = 1$. Gdy znamy już punkt X_n , określamy X_{n+1} jako liczbę wylosowaną zgodnie z rozkładem $U(0, aX_n)$. Dla jakich a powyższy ciąg jest martyngałem, podmartyngałem, nadmartyngałem.

Wskazówka. Skoro X_{n+1} jest wylosowany według rozkładu jednostajnego na określonym przedziale, to jej nadzieja matematyczna jest środkiem tego przedziału. Inaczej

$$E(X_{n+1}|\sigma(X_1, \dots, X_n)) = \frac{aX_n}{2} = \frac{a}{2}X_n.$$

Dla $a = 2$ mamy martyngał dla $a < 2$ – nadmartyngał, $a > 2$ – podmartyngał. POWRÓT DO ZADANIA

Pytanie 15.2. Niech $X_1, X_2, X_3, \dots, X_n$ będzie ciągiem niezależnych zmiennych losowych o rozkładzie $B(1, p)$ każda. Niech $S_n = X_1 + \dots + X_n$.

- (a) Wskazać takie rozłączne zbiory A_i oraz liczbę N , że $\sigma(X_1, \dots, X_n) = \sigma(A_1, \dots, A_N)$.
- (b) Wykazać, że $\sigma(S_n) \subsetneq \sigma(X_1, \dots, X_n)$ dla $n = 2$.
- (c) Czy $\sigma(X_1, \dots, X_n) = \sigma(S_1, \dots, S_n)$? Odpowiedź uzasadnić.

Wskazówka. Ad (a). Niech $n = 1$. Biorę $A_1 = \{X_1 = 0\}$, $A_2 = \{X_1 = 1\}$. Chociaż nie znamy (Ω, Σ, P) , to widzimy, że A_i są rozłączne, a także, że $P(A_1 \cup A_2) = 1$. Oczywiście $\sigma(A_1, A_2) \subset \sigma(X_1)$. Dla dowolnego zbioru borelowskiego $B \subset \mathbb{R}$ mamy $X_1^{-1}(B) = X_1^{-1}(B \cap \{0, 1\})$. Ten ostatni zbiór jest równy albo zbiorowi pustemu, albo A_1 , albo A_2 , albo Ω . Stąd $\sigma(X_1) \subset \sigma(A_1, A_2)$. Dla $n = 2$ mamy cztery zbiory: $A_1 = \{X_1 = 0, X_2 = 0\}$, $A_2 = \{X_1 = 0, X_2 = 1\}$, $A_3 = \{X_1 = 1, X_2 = 0\}$, $A_4 = \{X_1 = 1, X_2 = 1\}$. Podobnie dla większych n . Formalnie można napisać tak. Niech $N = 2^n$. Dla $1 \leq i \leq N$ niech $\varepsilon_1, \dots, \varepsilon_n$ będzie rozwinięciem liczby $i - 1$ w systemie dwójkowym. Określamy: $A_i = \{X_1 = \varepsilon_1, \dots, X_n = \varepsilon_n\}$.

Ad (b). Ponieważ S_n jest funkcją borelowską zmiennych X_1, \dots, X_n , to zawieranie jest oczywiste. Niech $n = 2$ i niech $A = \{X_1 = 0, X_2 = 1\}$. Widzimy, że $A \in \sigma(X_1, X_2)$. Natomiast $\sigma(S_2) = \sigma(\{S_2 = 0\}, \{S_2 = 1\}, \{S_2 = 2\})$, więc nie zawiera A (nie można go przedstawić jako sumę generatorów). Podobnie jest dla większych n .

Ad (c). Wektor (S_1, \dots, S_n) jest borelowską funkcją wektora (X_1, \dots, X_n) . Mamy też równości: $X_1 = S_1, X_2 = S_2 - S_1, \dots, X_n = S_n - S_{n-1}$, więc wektor (X_1, \dots, X_n) jest borelowską funkcją (S_1, \dots, S_n) . POWRÓT DO ZADANIA

Pytanie 15.3. Wykazać twierdzenie wypowiedziane w przykładzie 15.7.

Wskazówka. Oznaczmy: $X_n = E(X|\mathcal{A}_n)$. $E(X_n) = E(X|\mathcal{A}_n) = E(X) \in \mathbb{R}$. Ponieważ $\mathcal{A}_n \subset \mathcal{A}_{n+1}$, to $E(E(X|\mathcal{A}_{n+1})|\mathcal{A}_n) = E(X|\mathcal{A}_n)$, więc $\{E(X|\mathcal{A}_n)\}$ jest martyngałem względem filtracji $\{\mathcal{A}_n\}$. Dla każdego n zachodzi więc równość:

$$E(X_{n+1}|\mathcal{A}_n) = X_n,$$

a ponieważ dla $i \leq n$ X_i jest \mathcal{A}_i -mierzalna, więc jest też \mathcal{A}_n -mierzalna i stąd wektor (X_1, \dots, X_n) jest \mathcal{A}_n -mierzalny. Tak więc $\sigma(X_1, \dots, X_n) \subset \mathcal{A}_n$ i korzystając z poznanych już własności z powyższej równości otrzymujemy kolejno:

$$E(E(X_{n+1}|\mathcal{A}_n)|\sigma(X_1, \dots, X_n)) = E(X_n|\sigma(X_1, \dots, X_n)),$$

$$E(X_{n+1}|\sigma(X_1, \dots, X_n)) = X_n.$$

POWRÓT DO ZADANIA

Pytanie 15.4. Czy suma, różnica, iloczyn dwóch momentów stopu względem tej samej filtracji jest momentem stopu? Uzasadnić.

Wskazówka. Suma: TAK. $\{\tau_1 + \tau_2 = n\} = \bigcup_{k=1}^{n-1} \{\tau_1 = k, \tau_2 = n - k\} \in \mathcal{A}_n$.

Iloczyn: TAK. Podobnie jak wyżej: $\{\tau_1 \cdot \tau_2 = n\}$ jest sumą zbiorów postaci $\{\tau_1 = k, \tau_2 = l\}$, gdzie $kl = n$.

Różnica: na ogół NIE. Moment stopu ma być większy niż 0. POWRÓT DO ZADANIA

Pytanie 15.5. Podać dowód uwagi 15.16

Wskazówka. $\{\max(\tau_1, \tau_2) \leq n\} = \{\tau_1 \leq n\} \cap \{\tau_2 \leq n\}$,

$\{\min(\tau_1, \tau_2) \leq n\} = \{\tau_1 \leq n\} \cup \{\tau_2 \leq n\}$. POWRÓT DO ZADANIA

Pytanie 15.6. Wykazać uwagę 15.22 i wniosek 15.23.

Wskazówka. Dowód uwagi 15.22. Warunek (15.1) oznacza, że ciąg o wyrazach $E(|X_n|)$ jest ograniczony. Wiemy też, że $E(|X_n|) = E(X_n^+) + E(X_n^-)$. Ciągi o wyrazach nieujemnych $E(X_n^+)$ oraz $E(X_n^-)$ są więc tym bardziej ograniczone. Załóżmy teraz, że ciąg o wyrazach $E(X_n^+)$ jest ograniczony i że mamy do czynienia z podmartyngałem. Zachodzą więc nierówności:

$$E(X_{n+1}|\mathcal{A}_n) \geq X_n \text{ i stąd dla każdego } n: E(X_n) = E(X_n^+) - E(X_n^-) \geq E(X_1),$$

czyli $E(X_n^-) \leq E(X_n^+) - E(X_1)$. Stąd

$$E(|X_n|) = E(X_n^+) + E(X_n^-) \leq 2E(X_n^+) - E(X_1).$$

Zachodzi więc warunek (15.1). W przypadku nadmartyngału jest podobnie.

Dowód wniosku 15.23. Skoro $X_n \geq 0$, to $X_n^- = 0$ i możemy stosować pierwszą część uwagi 15.22, są więc spełnione założenia twierdzenia 15.21 o zbieżności. POWRÓT DO ZADANIA

Pytanie 16.1. Na początku w urnie są trzy białe kule. Co minutę losujemy z urny jedną kulę i jeżeli jest czerwona, to wrzucamy ją z powrotem, a jeżeli jest biała, to z takim samym prawdopodobieństwem wrzucamy tę kulę do urny albo zamiast niej wrzucamy kulę czerwoną. Niech C_n oznacza liczbę czerwonych kul w urnie po upływie n minut. Wskazać rozkład prawdopodobieństwa zmiennej losowej C_3 .

Wskazówka. Można wyliczyć bezpośrednio stosując twierdzenie o prawdopodobieństwie całkowitym.

$$P(C_3 = 0) = \frac{9}{72}, \quad P(C_3 = 1) = \frac{37}{72}, \quad P(C_3 = 2) = \frac{24}{72}, \quad P(C_3 = 3) = \frac{2}{72}.$$

Warto było rozważyć łańcuch Markowa na $M = \{0, 1, 2, 3\}$ z parametrami:

$$\mathbf{p} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 2/3 & 1/3 & 0 \\ 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

i wyznaczyć \mathbf{P}^3 . POWRÓT DO ZADANIA

Pytanie 16.2. Kontynuując przykład 16.2.

(1) Obliczyć prawdopodobieństwo tego, że w trzecim dniu umowy: sprząta Kaja, sprząta Leon.

(2) Używając komputera obliczyć prawdopodobieństwo tego, że Leon sprząta: w dziesiątym dniu umowy, w setnym dniu umowy.

Wskazówka. Ad (1). $\frac{23}{72}, \frac{43}{72}$.

Ad (2). Rozkład łańcucha w chwili t można otrzymać za pomocą t -tej potęgi macierzy przejścia. Odpowiedzi to, odpowiednio: $\frac{12024881}{20155392} \approx 0.5966086395$ oraz ≈ 0.4600000000 . POWRÓT DO ZADANIA

Pytanie 16.3. Uzupełnić lukę w dowodzie twierdzenia 16.6.

Wskazówka.

$$\text{POWRÓT DO ZADANIA} \quad \sum_i \mathbf{p}(i) = \sum_i P(\eta = i) = P(\bigcup_i \{\eta = i\}) = P(\eta \in M) = P(\Omega) = 1.$$

$$\sum_j \mathbf{P}(i, j) = \sum_j P(T(i, \xi_n) = j) = P(\bigcup_j \{T(i, \xi_n) = j\}) = P(T(i, \xi_n) \in M) = P(\Omega) = 1.$$

Pytanie 16.4. Niech ciąg $(X_0, X_1, X_2, X_3, \dots)$ będzie łańcuchem Markowa, Wykazać, że podciąg (X_1, X_3, X_5, \dots) oraz podciąg (X_2, X_4, X_6, \dots) są łańcuchami Markowa o odpowiednich parametrach.

Wskazówka. Niech \mathbf{p} oraz \mathbf{P} będą danymi parametrami rozważanego łańcucha. Wtedy z twierdzenia 16.10 wynika, że:

parametrami łańcucha (X_1, X_3, X_5, \dots) są $\mathbf{P}^T \mathbf{p}$ oraz \mathbf{P}^2 ,
parametrami łańcucha (X_2, X_4, X_6, \dots) są $(\mathbf{P}^2)^T \mathbf{p}$ oraz \mathbf{P}^2 .

POWRÓT DO ZADANIA

Pytanie 16.5. Niech $\{X_t\}$ będzie ciągiem wektorów losowych określonych przez algorytm PRS, patrz punkt 12.4. Zakładając, że A jest zbiorem skończonym, wykazać, że ciąg ten jest łańcuchem Markowa: wskaż macierz przejścia. Zweryfikować uzyskany wynik.

Wskazówka. Oznaczenia ze strony 98. Określamy $X_0 = Y_0$, $X_t = T(X_{t-1}, Y_t)$, dla $t = 1, 2, 3, \dots$, gdzie Y_0, Y_1, Y_2, \dots i.i.d. o rozkładzie $U(A)$,

$$T(x, y) = \begin{cases} x, & \text{gdy } f(x) \leq f(y), \\ y, & \text{gdy } f(x) > f(y), \end{cases}$$

$$\mathbf{P}(x, y) = P(Y_t \in A : T(x, Y_t) = y) = \begin{cases} \frac{\#\{w \in A : f(w) \geq f(x)\}}{\#A}, & \text{gdy } y = x, \\ \frac{1}{\#A}, & \text{gdy } y \neq x, f(x) > f(y), \\ 0, & \text{gdy } y \neq x, f(x) \leq f(y). \end{cases}$$

Dla weryfikacji naszego wyniku warto zauważyć, że dla dowolnego $x \in A$

$$\begin{aligned} \sum_{y \in A} \mathbf{P}(x, y) &= \mathbf{P}(x, x) + \sum_{y \in A, y \neq x} \mathbf{P}(x, y) \\ &= \frac{\#\{w \in A : f(w) \geq f(x)\}}{\#A} + \#\{w \in A : f(w) < f(x)\} \frac{1}{\#A} = 1. \end{aligned}$$

POWRÓT DO ZADANIA

Pytanie 16.6. Wskazać taki ciąg zmiennych losowych X_t , $t = 0, 1, 2, 3, \dots$ o wartościach w zbiorze co najwyżej przeliczanym M , że:

- (1) $\{X_t\}$ jest martyngałem, $\{X_t\}$ jest łańcuchem Markowa.
- (2) $\{X_t\}$ jest martyngałem, $\{X_t\}$ nie jest łańcuchem Markowa.
- (3) $\{X_t\}$ nie jest martyngałem, $\{X_t\}$ łańcuchem Markowa.
- (4) $\{X_t\}$ nie jest martyngałem, $\{X_t\}$ jest nie łańcuchem Markowa.

Wskazówka. Ad (1). Standardowy spacer losowy.

Ad (2). $X_0 = 0$, $X_{t+1} = X_t + Z_{t+1}$ dla $t = 0, 1, 2, \dots$, gdzie Z_1, Z_2, Z_3, \dots – niezależne zmienne losowe przyjmujące wartości całkowite o wspólnej nadziei matematycznej $m = 0$. X_t tworzą martyngał.

$$E(X_{t+1} | \sigma(X_1, \dots, X_t)) = E(X_t | \sigma(X_1, \dots, X_t)) + E(Z_{t+1} | \sigma(X_1, \dots, X_t)) = X_t + E(Z_{t+1}) = X_t.$$

Gdy rozkłady Z_i się zmieniają, to X_t nie tworzą łańcucha Markowa (jednorodnego):

$$P(X_{t+1} = j | X_t = i) = P(Z_{t+1} = j - i) = P_{Z_{t+1}}(j - i)$$

zależy od t .

Ad (3). Spacer losowy po prostej, gdy $p \neq q$.

Ad (4). Podobnie jak (2), tylko z różnymi nadziejami. POWRÓT DO ZADANIA

Pytanie 17.1. Definicje stanu powracającego i stanu okresowego można bez żadnych zmian powtórzyć w przypadku dowolnego łańcucha Markowa. Podać przykład takiego łańcucha, który ma jednocześnie: (1) stan powracający i stan niepowracający, (2) stan okresowy i stan nieokresowy.

Wskazówka. Ad (1). Łańcuch o macierzy przejścia

$$\mathbf{P} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 2/3 & 0 & 1/3 & 0 \\ 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Ad (2). Łańcuch o macierzy przejścia

$$\mathbf{P} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

POWRÓT DO ZADANIA

Pytanie 17.2. Podać przykład wskazujący, że twierdzenie 17.8 jest fałszywe w przypadku, gdy przestrzeń stanów nie jest skończona.

Wskazówka. Spacer losowy po prostej bez barier, gdy $r > 0$ jest nieokresowy, ale w twierdzeniu 17.8 nie zachodzi punkt 2. POWRÓT DO ZADANIA

Pytanie 17.3. Udowodnić, że jeżeli nieredukowalny łańcuch Markowa jest niepowracający, to dla każdego stanu i, j : $\lim_{n \rightarrow \infty} \mathbf{P}^n(i, j) = 0$.

Wskazówka. Ponieważ łańcuch jest niepowracający, to dla każdego j szereg $\sum_{n=1}^{\infty} \mathbf{P}^n(j, j)$ jest zbieżny, więc $\lim_{n \rightarrow \infty} \mathbf{P}^n(j, j) = 0$. Ustalmy teraz dwa stany i, j . Ponieważ łańcuch jest nieredukowalny, to istnieje takie $k \geq 1$, że $\mathbf{P}^k(j, i) > 0$. Dla dowolnego n mamy nierówność:

$$\mathbf{P}^{n+k}(j, j) \geq \mathbf{P}^k(j, i) \mathbf{P}^n(i, j), \text{ więc } \lim_{n \rightarrow \infty} \mathbf{P}^n(i, j) = 0.$$

POWRÓT DO ZADANIA

Pytanie 17.4. R

Wskazówka. o POWRÓT DO ZADANIA zważmy spacer losowy po prostej z barierami w punktach $A = -1$, $B = 1$,

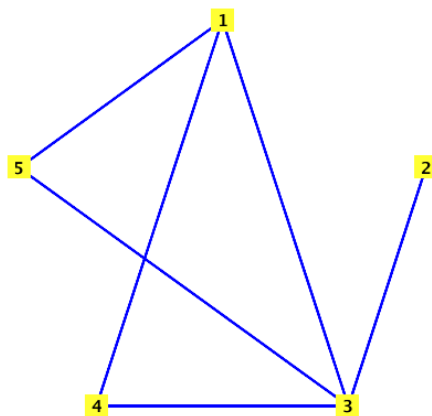
$p = q = 1/4$, $s_A = 9/10$, $s_B = 1/10$. Znaleźć rozkład stacjonarny. Stan stacjonarny $\pi := \begin{bmatrix} 45/68 \\ 18/68 \\ 5/68 \end{bmatrix}$.

Pytanie 17.5. Rozważmy urnowy model Bernoulliego, Przykład 16.7. Niech $k = 3$, $b_0 = 3$. Wyznaczyć oczekiwaną liczbę czerwonych kul w drugiej urnie po pierwszym, drugim, dziesiątym i trzydziestym losowaniu.

Wskazówka. $E(X_1) = 2$, $E(X_2) := \frac{5}{3}$, $E(X_{10}) = \frac{29525}{19683}$, $E(X_{30}) = \frac{102945566047325}{68630377364883} \approx 1.5000000000000072854$.

POWRÓT DO ZADANIA

Pytanie 17.6. Wyznaczyć macierz przejścia spaceru losowego po grafie.



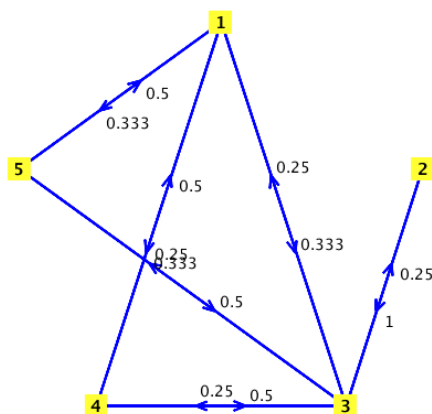
- (1) Czy łańcuch jest okresowy?
- (2) Naszkicować graf skierowany tego łańcucha.
- (3) Wskazać rozkład stacjonarny tego łańcucha.

Wskazówka.

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \end{bmatrix}.$$

Ad (1) Nie jest okresowy.

Ad (2)



Ad (3)

$$\pi = \begin{bmatrix} \frac{3}{12} \\ \frac{1}{12} \\ \frac{4}{12} \\ \frac{2}{12} \\ \frac{2}{12} \end{bmatrix}.$$

Powyższy wynik można było uzyskać dwoma sposobami.

POWRÓT DO ZADANIA

Pytanie 18.1. Znaleźć $\Sigma^{\frac{1}{2}}$, gdy $\Sigma = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$.

Wskazówka. $\Sigma = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & 4/\sqrt{5} \end{bmatrix}$. POWRÓT DO ZADANIA

Pytanie 18.2. Udowodnić twierdzenie 18.5.

Wskazówka. Po rozpisaniu wzoru na $\text{cov}(Y_i, Y_j)$, patrz definicja 18.4, otrzymujemy:

$$\text{cov}(W_i, W_j) = \sum_k a_{ik} \left(\sum_l \text{cov}(Y_k, Y_l) a_{jl} \right).$$

W nawiasie jest (k, j) -ty wyraz macierzy $\text{cov}(Y)A^T$. POWRÓT DO ZADANIA

Pytanie 18.3. Udowodnić twierdzenie 18.10.

Wskazówka.

$$M_W(t) = E\left(e^{t^T(AY+b)}\right) = e^{t^T b} E\left(e^{(A^T t)^T Y}\right) = e^{t^T b} M_W(A^T t).$$

POWRÓT DO ZADANIA

Pytanie 18.4. Wyznaczyć funkcję generującą momenty dla rozkładu Poissona P_λ .

Wskazówka.

$$M_X(t) = E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{\lambda(e^t-1)}.$$

POWRÓT DO ZADANIA

Pytanie 18.5. Znaleźć macierz kowariancji wektora losowego X mającego rozkład o gęstości:

(a) $f(x, y) = ce^{-x^2-4y^2}$, gdzie c jest odpowiednio dobraną stałą.

(b) $f(x, y) = ce^{-x^2-4y^2+2xy-2x}$, gdzie c jest odpowiednio dobraną stałą.

Wskazówka. Ad (a). Mamy $c = \frac{1}{\int_{\mathbb{R}^2} f(x, y) d(x, y)} = \frac{\pi}{2}$, $\Sigma = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/8 \end{bmatrix}$.

Ad (b). $\Sigma = \begin{bmatrix} 2/3 & 1/6 \\ 1/6 & 1/6 \end{bmatrix}$. POWRÓT DO ZADANIA

Pytanie 18.6. Znaleźć macierz kowariancji wektora losowego X mającego rozkład o gęstości $f = c \cdot g \cdot \mathbf{1}_K$, gdzie K jest trójkątem o wierzchołkach $(0, 0)$, $(0, 1)$, $(1, 1)$, $g(x, y) = x^2$, a c jest odpowiednią stałą.

Wskazówka. $c = 4$, $\Sigma = \begin{bmatrix} 2/75 & 1/75 \\ 1/75 & 14/225 \end{bmatrix}$. POWRÓT DO ZADANIA

Pytanie 19.1. Czy można tak dobrać stałą c , że funkcja $f: f(x, y) = ce^{-2x^2-3y^2+2x}$ jest gęstością rozkładu normalnego? Uzasadnić.

Wskazówka. Tak. Dla $c = \frac{1}{\int_{\mathbb{R}^2} f(x, y) d(x, y)}$ (jest dodatnia) f jest gęstością pewnego rozkładu Q . Obliczamy

$M_Q(t) = e^{\frac{1}{2}t_1 + \frac{1}{8}t_1^2 + \frac{1}{12}t_2^2}$. Widać, że jest to funkcja tworząca momenty rozkładu normalnego o parametrach $\mu = \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 1/8 & 0 \\ 0 & 1/12 \end{bmatrix}$. Warto użyć Maple. POWRÓT DO ZADANIA

Pytanie 19.2. Wektor losowy o współrzędnych S, T ma rozkład jednostajny na kwadracie $[0, 1]^2$.

(1) Wskazać dwoma sposobami nadzieję matematyczną oraz macierz kowariancji wektora o współrzędnych $S+T, S-T$.

(2) Czy zmienne losowe $S+T, S-T$ są niezależne?

Wskazówka. Ad (1) Sposób 1. Poprzez twierdzenie 5.33 otrzymujemy gęstość wektora o współrzędnych $S+T, S-T$ (rozkład jednostajny na kwadracie $(0, 0)$, $(1, -1)$, $(2, 0)$, $(1, 1)$) i całkując wyznaczamy wszystkie parametry.

Sposób 2. Twierdzenie 18.5 – dużo szybszy.

$$\mu = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{6} \end{bmatrix}.$$

Sposób 3 (na siłę). Można bezpośrednio wyliczyć sześć powyższych parametrów.

Sposób 4. Wyznaczając funkcje generujące momenty.

Ad. (2) Są zależne. $\{S + T > 1.5\}$ oraz $\{S - T > 0.5\}$ są rozłączne i mają dodatnie prawdopodobieństwa.

POWRÓT DO ZADANIA

Pytanie 19.3. Wskazać rozkład wektora (X, Y) oraz rozkład warunkowy $P_{X|Y=y}$, gdy wiemy, że $P_X = N(m, \sigma)$ oraz $P_{Y|X=x} = N(x, \sigma_x)$, gdzie $m \in \mathbb{R}$, $\sigma > 0$, $\sigma_x > 0$ są znane.

Wskazówka. Jest to w istocie przykład 19.10 z innymi oznaczeniami.

$$P_{\begin{bmatrix} X \\ Y \end{bmatrix}} = M_2 \left(\begin{bmatrix} m \\ m \end{bmatrix}, \begin{bmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 + \sigma_x^2 \end{bmatrix} \right),$$

$$P_{X|Y=y} = N \left(\frac{m\sigma_x^2 + y\sigma^2}{\sigma^2 + \sigma_x^2}, \frac{\sigma_x\sigma}{\sqrt{\sigma^2 + \sigma_x^2}} \right).$$

POWRÓT DO ZADANIA

Pytanie 19.4. Wektor losowy Y ma rozkład normalny $N_n(0, \sigma I_n)$, gdzie $\sigma > 0$ jest dane, a I_n jest macierzą identycznościową. Wykazać, że wektor $W = AY$ ma też rozkład $N_n(0, \sigma I_n)$, gdy A jest macierzą izometrii.

Wskazówka. Jako macierz izometrii, A spełnia warunek $AA^T = I_n$. POWRÓT DO ZADANIA

Pytanie 19.5. Niech wektor X o współrzędnych ξ, η ma rozkład ciągły normalny. Wykazać, że proste regresji są prostopadłe, wtedy i tylko wtedy, gdy ξ, η są niezależne.

Wskazówka. Po rozpisaniu równań prostych regresji $y = h(x)$, $x = g(y)$ otrzymujemy wzór na iloczyn skalarny wektorów do nich prostopadłych: $(\sigma_\xi^2 + \sigma_\eta^2)\varrho$, co kończy dowód, gdyż obydwa warunki są równoważne warunkowi $\varrho = 0$. POWRÓT DO ZADANIA

Pytanie 19.6. Znaleźć $E(X + Y + Z|X - 2Y)$, gdzie X, Y, Z są niezależnymi zmiennymi losowymi o rozkładzie $N(m, 1)$ każda, $m \in \mathbb{R}$.

Wskazówka. $E(X + Y + Z|X - 2Y) = E(X + Y|X - 2Y) + E(Z|X - 2Y)$. Z niezależności drugi składnik $= E(Z) = m$. Wyznaczamy pierwszy składnik.

Wektor $\begin{bmatrix} X \\ Y \end{bmatrix}$ ma rozkład $N_2 \left(\begin{bmatrix} m \\ m \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$.

Wektor losowy $\begin{bmatrix} X + Y \\ X - 2Y \end{bmatrix} = A \begin{bmatrix} X \\ Y \end{bmatrix}$, gdzie $A = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$. Mamy więc:

$$\mu = A \begin{bmatrix} m \\ m \end{bmatrix} = \begin{bmatrix} 2m \\ -m \end{bmatrix}, \quad \Sigma = AI_2A^T = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 5 \end{bmatrix}.$$

Zgodnie z twierdzeniem 19.7 mamy:

$$E(X + Y|X - 2Y) = 2m + \frac{1}{5}(X - 2Y + m).$$

W końcu:

$$E(X + Y + Z|X - 2Y) = \frac{14}{5}m - \frac{1}{5}(X - 2Y).$$

POWRÓT DO ZADANIA

Pytanie 20.1. Dane są obserwacje $(x_1, y_1), \dots, (x_n, y_n)$, gdzie każde $x_i = (x_{i1}, \dots, x_{ik}) \in \mathbb{R}^k$ reprezentuje określone wielkości X_1, \dots, X_k , natomiast y_i reprezentuje odpowiadające im w pewien określony sposób wielkości Y_i . Na podstawie tych danych należy wyznaczyć takie liczby $\hat{\beta}_1, \dots, \hat{\beta}_k$, że sumy $\sum_{j=1}^k \hat{\beta}_j x_{ij}$ możliwie najlepiej przybliżają wielkości y_i dla $i = 1, \dots, n$ w sensie kryterium najmniejszych kwadratów. Zaproponować metodę rozwiązania tego problemu.

Wskazówka. Niech $X = [x_{ij}]$, $i = 1, \dots, n$, $j = 1, \dots, k$ i niech $y \in \mathbb{R}^n$ będzie wektorem kolumnowym o współrzędnych y_i . Wtedy nasz problem można sformułować tak:

Wskazać taki wektor $\hat{\beta} \in \mathbb{R}^k$, że dla każdego $\beta \in \mathbb{R}^k$

$$\|X\hat{\beta} - y\| \leq \|X\beta - y\| \quad \text{czyli} \quad \|X\hat{\beta} - y\|^2 \leq \|X\beta - y\|^2. \quad (21.2)$$

POWRÓT DO ZADANIA

Pytanie 20.2. Przy powyższych oznaczeniach, załóżmy, że $k < n$ oraz, że rząd $X = k$. Wyznaczyć rzutowanie prostopadłe na podprzestrzeń wektorową $V = \text{Im } X$.

Wskazówka. Rozważmy macierz $X^T X$. Jest to macierz symetryczna nieujemnie określona, a ponieważ rząd X jest maksymalny, to $X^T X$ jest dodatnio określona, a więc jest nieosobliwa. Dla $y \in \mathbb{R}^n$ określamy wektor v wzorem:

$$v = X(X^T X)^{-1}y.$$

Zauważmy, że $v \in V$. Natomiast $y - v \in V^\perp$, gdyż dla każdego $w \in V$ kolejno mamy: $w = Xz$ dla pewnego z i wtedy

$$\begin{aligned} \langle w, y - v \rangle &= \langle Xz, y - v \rangle = \langle z, X^T y - X^T v \rangle \\ &= \langle z, X^T y - X^T X(X^T X)^{-1}X^T y \rangle = \langle z, X^T y - X^T y \rangle = 0. \end{aligned}$$

Oznacza to, że rzutowanie prostopadłe na V jest dane wzorem:

$$P_V y = X(X^T X)^{-1}X^T y, \quad y \in \mathbb{R}^n.$$

POWRÓT DO ZADANIA

Pytanie 20.3. Rozwiązać problem sformułowany w pytaniu 20.1 wykorzystując powyższy wynik.

Wskazówka. Niech $\hat{y} = P_V y = X(X^T X)^{-1}X^T y$. Z twierdzenia Pitagorasa wiemy, że dla każdego $\mu \in V$ zachodzi nierówność $\|y - \hat{y}\|^2 \leq \|y - \mu\|^2$. Ponieważ rząd macierzy X jest maksymalny, to dla każdego $\mu \in V$ istnieje dokładnie jeden wektor $\beta \in \mathbb{R}^k$ taki, że $\mu = X\beta$. Niech $\hat{\beta}$ będzie tym elementem, który odpowiada \hat{y} . Mamy więc żadaną nierówność (21.2). Mnożąc lewostronnie przez X^T z równości $\hat{\mu} = X\hat{\beta}$ otrzymujemy $X^T \hat{y} = X^T X \hat{\beta}$, więc w końcu:

$$\hat{\beta} = (X^T X)^{-1}X^T \hat{y} = (X^T X)^{-1}X^T X(X^T X)^{-1}X^T y = (X^T X)^{-1}X^T y.$$

Odpowiedź:

$$\hat{\beta} = (X^T X)^{-1}X^T y. \quad (21.3)$$

POWRÓT DO ZADANIA

Pytanie 20.4. Dla danego ciągu punktów płaszczyzny $(x_1, y_1), \dots, (x_n, y_n)$ wskazać takie liczby a, b , że wartość $\sum_{i=1}^n (y_i - (ax_i + b))^2$ jest najmniejsza.

Wskazówka. Jest to szczególny przypadek poprzedniego problemu. Tutaj macierz X ma postać:

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

Musimy założyć, że rząd $X = 2$, co jest równoważne temu, że nie wszystkie punkty x_i są sobie równe. Wielkości a, b wyliczamy korzystając ze wzoru (21.3) na $\hat{\beta}$. Mianowicie:

$$\begin{bmatrix} b \\ a \end{bmatrix} = \left(\begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

Wprowadźmy standardowo używane oznaczenia: $\bar{x} = \frac{1}{n} \sum_i x_i$, $\overline{xy} = \frac{1}{n} \sum_i x_i y_i$, $\overline{x^2} = \frac{1}{n} \sum_i x_i^2$. Po wykonaniu mnożenia i odwracania macierzy otrzymujemy.

$$\begin{bmatrix} b \\ a \end{bmatrix} = \frac{1}{\overline{x^2} - \bar{x}^2} \begin{bmatrix} \overline{x^2} \cdot \bar{y} - \overline{xy} \cdot \bar{x} \\ \overline{xy} - \bar{x} \cdot \bar{y} \end{bmatrix}.$$

Odpowiedź:

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \quad b = \frac{\overline{x^2} \cdot \bar{y} - \overline{xy} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}. \quad (21.4)$$

POWRÓT DO ZADANIA

Pytanie 20.5. Dla danego ciągu punktów płaszczyzny $(x_1, y_1), \dots, (x_n, y_n)$ wskazać takie liczby a, b obliczone na podstawie twierdzenia 20.8 przy założeniu, że twierdzenie to stosuje się to wektora losowego (ξ, η) mającego rozkład jednostajny na zbiorze $\{(x_1, y_1), \dots, (x_n, y_n)\}$.

Wskazówka. Stosując oznaczenia z poprzedniego pytania łatwo stwierdzić, że:

$$m_\xi = \bar{x}, \quad m_\eta = \bar{y}, \quad \sigma_\xi^2 = \overline{x^2} - \bar{x} \cdot \bar{x}, \quad \sigma_\eta^2 = \overline{y^2} - \bar{y} \cdot \bar{y}, \quad \text{cov}(\xi, \eta) = \overline{xy} - \bar{x} \cdot \bar{y}.$$

Pamiętając, że $\rho = \frac{\text{cov}(\xi, \eta)}{\sigma_\xi \sigma_\eta}$, na podstawie twierdzenia 20.8 mamy:

$$a = \rho \frac{\sigma_\eta}{\sigma_\xi} = \frac{\text{cov}(\xi, \eta)}{\sigma_\xi^2} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2},$$

$$b = m_\eta - \frac{\rho \sigma_\eta m_\xi}{\sigma_\xi} = m_\eta - \frac{\text{cov}(\xi, \eta) m_\xi}{\sigma_\xi^2} = \bar{y} - \frac{\overline{xy} \cdot \bar{x} - \overline{x^2} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\overline{x^2} \cdot \bar{y} - \overline{xy} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}.$$

POWRÓT DO ZADANIA

Pytanie 20.6. Podobnie jak w pytaniu 20.1 dane są obserwacje $(x_1, y_1), \dots, (x_n, y_n)$, gdzie każde $x_i = (x_{i1}, \dots, x_{ik}) \in \mathbb{R}^k$ reprezentuje określone wielkości X_1, \dots, X_k , natomiast y_i reprezentuje odpowiadające im w pewien określony sposób wielkości Y_i . Obecnie zakładamy, że Y_i są zmiennymi losowymi o wartościach oczekiwanych μ_i , przy czym zakładamy, że dla każdego i zachodzi związek liniowy:

$$\mu_i = \beta_1 x_{i1} + \dots + \beta_k x_{ik},$$

gdzie β_1, \dots, β_k nie zależą od i . Zbadać własności estymatorów $\hat{\mu}$ parametru μ oraz $\hat{\beta}$ parametru β określonych jako:

$$\hat{\mu} = X(X^T X)^{-1} X^T Y, \quad \hat{\beta} = (X^T X)^{-1} X^T Y,$$

gdzie Y jest wektorem losowym o współrzędnych Y_i .

Wskazówka. Z założenia $E(Y) = X\beta$, więc

$$E(\hat{\beta}) = E((X^T X)^{-1} X^T Y) = (X^T X)^{-1} X^T E(Y) = (X^T X)^{-1} X^T X \beta = \beta.$$

Podobnie $E(\hat{\mu}) = \mu$, czyli:

$\hat{\mu}$ oraz $\hat{\beta}$ są nieobciążonymi estymatorami parametrów μ oraz β .

Założmy dodatkowo, że zmienne losowe Y_1, \dots, Y_n są niezależne i wszystkie mają wspólną wariancję $\sigma^2 < \infty$. Wtedy $\hat{\mu}$ oraz $\hat{\beta}$ są najlepszymi estymatorami liniowymi dla μ oraz β w następującym sensie. Ustalmy $a \in \mathbb{R}^n$. Odwzorowanie liniowe $S: \mathbb{R}^n \rightarrow \mathbb{R}$ nazywamy najlepszym liniowym estymatorem parametru $a^T \mu$, jeżeli $E(S(Y)) = a^T \mu$ oraz dla każdego odwzorowania liniowego $T: \mathbb{R}^n \rightarrow \mathbb{R}$ takiego, że $E(T(Y)) = a^T \mu$

$$D^2(S(Y)) \leq D^2(T(Y)).$$

Niech $A \in M(m, n)$. Mówimy, że $S: \mathbb{R}^n \rightarrow \mathbb{R}^m$ jest najlepszym liniowym estymatorem parametru $A\mu$, jeżeli dla każdego i i -ta współrzędna S_i jest najlepszym estymatorem $(A\mu)_i$ – i -tej współrzędnej wektora $A\mu$. Udowodnimy:

Twierdzenie Gaussa–Markowa. Dla każdej macierzy $A \in M(m, n)$ estymator $A\hat{\mu}$ jest najlepszym liniowym estymatorem parametru $A\mu$.

Biorąc $A = I_n$ widzimy, że $\hat{\mu}$ jest najlepszym liniowym estymatorem parametru μ . Biorąc $A = (X^T X)^{-1} X^T$ widzimy, że $\hat{\beta}$ jest najlepszym liniowym estymatorem parametru β .

Dowód twierdzenia Gaussa–Markowa. Jak już wiemy $E(\hat{\mu}) = \mu$, więc z liniowości $E(A\hat{\mu}) = AE(\hat{\mu}) = A\mu$. Musimy więc jeszcze porównać estymator $A\hat{\mu}$ z innymi liniowymi nieobciążonymi estymatorami parametru $A\mu$.

Biorąc pod uwagę dwukrokową definicję najlepszego liniowego estymatora wystarczy nasz dowód ograniczyć do przypadku, gdy macierz A ma tylko jeden wiersz, powiedzmy $A = a^T$. Badamy więc estymatory liniowe parametru $a^T \mu$. Niech T będzie takim estymatorem, czyli $T(y) = b^T y$ dla pewnego $b \in \mathbb{R}^n$. Niech, jak w pytaniu 20.1, V oznacza podprzestrzeń wektorową $\text{Im } X$. Pamiętamy, że rzutowanie prostopadłe na V wyraża się wzorem $P_V y = X(X^T X)^{-1} X^T y$ dla $y \in \mathbb{R}^n$.

Zachodzą równoważności: $T(Y)$ jest nieobciążonym estymatorem parametru $a^T \mu \iff b^T \mu = a^T \mu$ dla każdego $\mu \in V \iff \langle b - a, \mu \rangle = 0$ dla każdego $\mu \in V \iff b - a \in V^\perp \iff P_V a = P_V b$. Dla takiego estymatora obliczamy wariancję, pamiętając, że z obecnych założeń wynika równość $\text{cov}(Y) = \sigma^2 I_n$:

$$D^2(T(Y)) = D^2(b^T Y) = b^T \text{cov}(Y) b = \sigma^2 b^T b = \sigma^2 \|b\|^2.$$

Z twierdzenia Pitagorasa

$$D^2(T(Y)) = \sigma^2 (\|b - P_V b\|^2 + \|P_V b\|^2) = \sigma^2 (\|b - P_V a\|^2 + \|P_V a\|^2).$$

Widzimy, że $D^2(T(Y))$ osiąga wartość najmniejszą $\iff b = P_V a \iff b = X(X^T X)^{-1} X^T a$. Wtedy też mamy:

$$T(Y) = b^T Y = (X(X^T X)^{-1} X^T a)^T Y = a^T X(X^T X)^{-1} X^T Y = a^T \hat{\mu}. \quad \square$$

POWRÓT DO ZADANIA

Bibliografia

- [1] Aczel A., *Statystyka w zarządzaniu*, tłum. Z. Czerwiński, W. Latusek, PWN, Warszawa 2000.
- [2] Ash R..B., Doleans-Dade C., *Probability and measure theory 2nd Edition*, Academic Press, New York 2000.
- [3] Asmussen S., Glynn P.W., *Stochastic Simulation: Algorithms and Analysis*, Springer, New York 2007.
- [4] Billingsley P., *Prawdopodobieństwo i miara*, tłum. K. Kizeweter, J. Roguski, PWN, Warszawa 1987.
- [5] Cyganowski S., Kloeden P., Ombach J., *From elementary probability to stochastic differential equations with MAPLE*, Springer, Berlin 2002.
- [6] Feller W., *Wstęp do rachunku prawdopodobieństwa, t. 1*, PWN, tłum. Bartoszyński R., Bielecki B., Warszawa 1977.
- [7] Feller W., *Wstęp do rachunku prawdopodobieństwa, t. 2*, tłum. Bartoszyński R., PWN, Warszawa 1978.
- [8] Fisz M., *Rachunek prawdopodobieństwa i statystyka matematyczna*, PWN, Warszawa 1969.
- [9] Gajek L., Kałuszka M., *Wnioskowanie statystyczne. Modele i metody*, Wydawnictwo Naukowo-Techniczne, Warszawa 1996.
- [10] Ibe O.C., *Fundamentals of Applied Probability and Random Processes*, Elsevier, Lowell MA 2014.
- [11] Iosifescu M., *Skończone procesy Markowa i ich zastosowania*, tłum. Wesolowski J., PWN, Warszawa 1996.
- [12] Jakubowski J., Sztencel R., *Rachunek prawdopodobieństwa dla prawie każdego*, Script, Warszawa 2006.
- [13] Jakubowski J., Sztencel R., *Wstęp do teorii prawdopodobieństwa*, Script, Warszawa 2010.
- [14] Karian Z.A., Tanis E. A., *Probability and statistics, explorations with Maple*, Prentice Hall, New York 1999.
- [15] Krowiak A., *Maple. Podręcznik*, Helion, Gliwice 2012.
- [16] Kulkarni V.G., *Modeling and Analysis of Stochastic Systems, Third Edition*, Chapman and Hall/CRC Press, Boca Raton 2017.
- [17] Madras N., *Lectures on Monte Carlo Methods*, American Mathematical Society, Providence, RI 2002.
- [18] Niemirowicz W., *Symulacje stochastyczne i metody Monte Carlo*, Uniwersytet Warszawski, Warszawa 2013.
- [19] Ombach J., Mazur M., *Rachunek prawdopodobieństwa i statystyka – kurs internetowy*, <http://wazniak.mimuw.edu.pl/index.php>, Warszawa 2006, data dostępu: 07.11.2019.
- [20] Ombach J., *Rachunek prawdopodobieństwa wspomagany komputerowo dla studentów matematyki stosowanej*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2018.

Indeks

A

algorytm PRS, 98

B

bayesowskie przedziały ufności, 147

C

całka względem miary, 38

całkowanie metodami Monte Carlo, 96

całkowanie względem rozkładu, 39

centralne twierdzenie graniczne, 61

CTG, 62

D

δ -Diraca, 18

diagonalizacja macierzy symetrycznej, 139

d -wymiarowa kostka dyskretna, 135

d -wymiarowy spacer losowy, 131

dystrybuanta, 19

rozkładu ciągłego, 23

rozkładu dyskretnego, 21

E

estymator, 93

asymptotycznie nieobciążony, 94

Nadarayi–Watsona, 156

nieobciążony, 94

zgodny

silnie, 93

słabo, 93

F

filtracja, 117

funkcja

borelowska, 27

charakterystyczna, 81

charakterystyczna zbioru, 26

generująca momenty, 140

prosta, 27

regresji, 146, 150

liniowej, 153

schodkowa, 27

tworząca, 87

wypukła, 113

G

gęstość rozkładu, 22

graf, 134

skierowany, 134

z wagami, 134

I

i.i.d., 73

iloczyn kartezjański przestrzeni probabilistycznych, 15

iloczyn skalarny, 44

J

jednorodny łańcuch Markowa, 123

K

kowariancja, 44

L

lemat

Borela–Cantellego, 73

Kroneckera, 72

Toeplitza, 72

liczby

pseudolosowe, 92

z rozkładu normalnego, 148

losowania, 11

Ł

łańcuch

ergodyczny, 132

Markowa, 123

nieredukowalny, 130

nieokresowy, 132

okresowy, 132

M

macierz

dodatnio określona, 139

kowariancji, 140

przejścia łańcucha Markowa, 124

markowskie metody Monte Carlo, 136

martyngał, 117

MCMC, 136

Mersenne Twister Algorithm, 93

metoda

Boxa–Mullera, 148

Monte Carlo

całkowanie, 96

optymalizacja, 98

rozwiązywanie równań, 99

najmniejszych kwadratów, 150, 154
 miara, 6
 Lebesgue'a, 9
 probabilistyczna, 6
 Mocne Prawo Wielkich Liczb, MPWL, 73
 dla i.i.d., 73
 moment, 43
 bezwzględny, 43
 centralny, 43
 stopu, 120

N

nadmartyngał, 117
 nadzieja matematyczna, 37, 39
 wektora, 140
 nadzieja warunkowa, 107
 jako rzutowanie, 152
 względem σ -algebry, 105
 względem wektora losowego, 106
 nieredukowalny łańcuch Markowa, 130
 nierówność
 Cauchy'ego–Schwarza, 44
 Czebyszewa, 46
 Jensena, 113
 Kołmogorowa, 70
 niezależność wektorów losowych, 34

O

obniżanie wariancji, 153
 przez warunkowanie, 113
 odchylenie standardowe, 43
 okres łańcucha Markowa, 132
 optymalizacja metodami Monte Carlo, 98

P

P, 124
 $\Phi_{m,\sigma}$ – dystrybuanta rozkładu normalnego $N(m, \sigma)$,
 60
 podmartyngał, 117
 porównanie rodzajów zbieżności, 69
 potęgi macierzy przejścia, 126
 prawdopodobieństwa
 a posteriori, 14
 a priori, 14
 prawdopodobieństwo
 całkowite, 13
 geometryczne, 10
 powrotu, 130
 warunkowe, 13
 zdarzenia, 6
prawie wszędzie (pw.), 23
 prawo wielkich liczb
 mocne, 73
 mocne dla i.i.d., 73
 słabe, 47
 problem plecakowy, 136

proces
 Poissona, 59
 stochastyczny, 59
 prosta regresji liniowej, 153
 próba
 Bernoulliego, 16
 próbka prosta, 93
 przedział ufności, 94
 przestrzeń
 probabilistyczna, 6
 Pure Random Search, 98

R

regresja, 150
 jako narzędzie statystyczne, 155
 liniowa, 153
 reguła 3 – σ , 46
 rozkład
 $(0, 1, p)$, 21
 Bernoulliego, $(0, 1, p)$, 50
 brzegowy, 30
 ciągły, 22
 dwumianowy
 $B(n, p)$, 21, 50
 ujemny, 55
 dyskretny, 20
 Erlanga, 58
 Gaussa, 60
 geometryczny, 54
 hipergeometryczny, 54
 jednopunktowy δ_c , 49
 jednostajny, 24
 dyskretny na zbiorze skończonym, 49
 $U(a, b)$, 49
 normalny
 $N(m, \sigma)$, 60
 na płaszczyźnie, 144
 standardowy, $N(0, 1)$, 61
 wielowymiarowy, 143
 Pascala, 55
 początkowy łańcucha Markowa, 124
 Poissona, 51
 prawdopodobieństwa, 18
 stacjonarny łańcucha Markowa, 132
 warunkowy, 32
 wektora losowego, 27
 wykładniczy E_λ , 56
 rozwiązywanie równań metodami Monte Carlo, 99

S
 σ -algebra, 6
 $\sigma(\mathcal{F})$, 9
 schemat
 Bernoulliego, 16
 dyskretny
 nieskończony, 9

- skończony, 9
 - klasyczny, 8
 - z niepełną informacją, 8
 - słabe prawo wielkich liczb, 47
 - spacer losowy
 - d -wymiarowy, 125, 131
 - po grafie, 136
 - po płaszczyźnie, 125
 - po prostej, 87, 124
 - stan
 - niewracający, 130
 - okresowy, 131
 - powracający, 130
 - standardowy rozkład normalny, $N(0, 1)$, 61
 - stany komunikujące, 130
 - submartyngał, 117
 - supmartyngał, 117
- Ś**
- średni błąd, 43
- T**
- tożsamość Walda, 120
 - twierdzenie
 - Bayesa, 14
 - Chinczyna, 90, 165
 - de Moivre'a–Laplace'a, 62
 - ergodyczne, 132
 - Halmosa, 118
 - Lebesgue'a, 74
 - Lindeberga–Lévy'ego, 62
 - o ciągłości, 84
 - o ciągu zdarzeń wstępujących, 7
 - o ciągu zdarzeń zstępujących, 8
 - o jednoznaczności, 82
 - o wyborze, 78
 - o zbieżności szeregu, 71
 - Radona–Nikodyma, 104
- U**
- ujemny rozkład dwumianowy, 55
 - urnowy model Bernoulliego, 125
- W**
- wariancja, 43
 - wartość
 - oczekiwana, 39
 - własna, 139
 - warunek Prochorowa, 78
 - warunkowa
 - nadzieja matematyczna, 105
 - wartość oczekiwana, 105, 107
 - wektor losowy, 26
 - wektory losowe
 - niezależne, 34
 - równe prawie wszędzie, 68
 - zależne, 34
 - wnioskowanie bayesowskie, 147
 - współczynnik korelacji, 44
 - wybór
 - losowy, 10
 - strategii, 118
 - wzór Stirlinga, 89
- X**
- X^- , 27
 - X^+ , 27
- Z**
- zbieżność
 - martyngałów, 122
 - rozkładów, 78
 - rozkładów – warunki równoważne, 81
 - rozkładów, \xrightarrow{d} , 68
 - stochastyczna, \xrightarrow{s} , 68
 - średnio kwadratowa, $\xrightarrow{L^2}$, 152
 - według rozkładów, \xrightarrow{d} , 68
 - z prawdopodobieństwem 1, $\xrightarrow{1}$, 68
 - zmiennych losowych, 68
 - zbiory
 - borelowskie, 9
 - zbiór
 - opisywany przez wektor losowy, 26
 - zbiór opisany przez wektor losowy, 26
 - zdarzenia
 - niezależne, 15
 - zdarzenie, 6
 - elementarne, 6
 - niemożliwe, 6
 - pewne, 6
 - przeciwnie, 6
 - ziarno, 92
 - zmienna losowa, 26
 - zmiennie
 - nieskorelowane, 44