

Wojciech Kordecki

Rachunek prawdopodobieństwa i statystyka matematyczna

Wrocław 1998

Spis treści

Wstęp	1
1. Prawdopodobieństwo	2
1.1. Aksjomaty prawdopodobieństwa	2
1.1.1. Przestrzeń zdarzeń	2
1.1.2. Aksjomaty Kołmogorowa	3
1.1.3. Geometryczna i klasyczna definicja prawdopodobieństwa	5
1.1.4. Zadania	6
1.2. Prawa wielkich liczb i symulacja	8
1.2.1. Mocne i słabe prawo wielkich liczb	8
1.2.2. Pierwsze przykłady symulacji	9
1.2.3. Zadania	11
1.3. Prawdopodobieństwo warunkowe	12
1.3.1. Definicja i podstawowe własności	12
1.3.2. Wzór Bayesa	13
1.3.3. Zadania	14
2. Zmienne losowe	15
2.1. Rozkłady zmiennych losowych	15
2.1.1. Definicja zmiennej losowej	15
2.1.2. Dystrybuanta zmiennej losowej	16
2.1.3. Gęstość	18
2.1.4. Zadania	19
2.2. Momenty zmiennych losowych	20
2.2.1. Całka Stieltjesa	20
2.2.2. Wartość oczekiwana	20
2.2.3. Momenty wyższych rzędów	22
2.2.4. Zadania	24
2.3. Rozkłady dyskretne	25
2.3.1. Rozkład dwupunktowy i dwumianowy	25
2.3.2. Rozkład Poissona	26
2.3.3. Zadania	28
2.4. Rozkłady ciągłe	29
2.4.1. Rozkład jednostajny	29
2.4.2. Rozkład wykładniczy	29
2.4.3. Rozkład normalny	31
2.4.4. Zadania	34
3. Twierdzenia graniczne	35
3.1. Nierówność Czebyszewa i prawa wielkich liczb	35
3.1.1. Nierówności Markowa i Czebyszewa	35
3.1.2. Prawa wielkich liczb	36
3.1.3. Zadania	37
3.2. Funkcje charakterystyczne	38
3.2.1. Definicje i własności	38
3.2.2. Rozkład gamma	39

3.2.3.	Zadania	40
3.3.	Centralne twierdzenie graniczne	41
3.3.1.	Twierdzenie Lindeberga-Levy'ego	41
3.3.2.	Rozkłady chi-kwadrat i t-Studenta	42
3.3.3.	Zadania	43
4.	Podstawowe pojęcia statystyki	44
4.1.	Definicje	44
4.1.1.	Słownik pojęć statystycznych	44
4.1.2.	Najważniejsze statystyki	44
4.1.3.	Zadania	46
4.2.	Dystrybuanta empiryczna i histogramy	47
4.2.1.	Dystrybuanta empiryczna	47
4.2.2.	Histogramy	47
4.2.3.	Zadania	47
5.	Estymacja	48
5.1.	Estymacja punktowa	48
5.1.1.	Metoda momentów	48
5.1.2.	Metoda największej wiarygodności	48
5.1.3.	Zadania	50
5.2.	Estymacja przedziałowa	51
5.2.1.	Przedziały ufności	51
5.2.2.	Przedziały ufności dla średniej	51
5.2.3.	Przedziały ufności dla wariancji	53
5.2.4.	Zadania	54
6.	Testowanie hipotez	55
6.1.	Testy parametryczne	55
6.1.1.	Testy dla średniej	56
6.1.2.	Testy dla wariancji	57
6.1.3.	Testy dla dwóch średnich	58
6.1.4.	Zadania	58
6.2.	Testy nieparametryczne	59
6.2.1.	Testy zgodności	59
6.2.2.	Testy niezależności	60
6.2.3.	Zadania	61
Literatura		62

Wstęp

Materiał zawarty w konspekcie wykładu jest podzielony na dwie zasadnicze części – rachunek prawdopodobieństwa i statystykę matematyczną. Przeznaczony on jest na jednosemestralny wykład w wymiarze 4 godzin tygodniowo.

Cały wykład rachunku prawdopodobieństwa jest oparty na aksjomatyce Kołmogorowa, w tym na ścisłym zdefiniowaniu zdarzeń jako podzbiorów przestrzeni zdarzeń elementarnych Ω , tworzących σ -algebrę zdarzeń. Z tego powodu eksponowana jest raczej „geometryczna definicja prawdopodobieństwa” zamiast „kombinatorycznej definicji prawdopodobieństwa”. Z drugiej strony, na samym początku wykładu podaje się prawo wielkich liczb w najprostszej postaci, aby można było wprowadzić intuicje częstościowe, (autor świadomie unika terminu „częstościowa definicja prawdopodobieństwa”, obawiając się jego nieco bałamutnego wydźwięku), co pozwala na ilustrację wykładu symulacjami komputerowymi. Z tego względu, nieco na marginesie wykładu pojawia się też metoda Monte Carlo i generatory liczb losowych.

Druga połowa semestru poświęcona jest na statystykę matematyczną. Nacisk jest położony przede wszystkim na podanie ogólnych metod i ich zrozumienie, a nie na podanie szczegółowych rozwiązań, które można znaleźć w licznych, a częstokroć bardzo dobrych podręcznikach i poradnikach. Ta część wykładu jest uzupełniona o procedury ułatwiające obliczanie wartości niektórych statystyk. Procedury te, napisane głównie w Pascalu, tylko częściowo umieszczone są w tym konspekcie. Wszystkie procedury są dostępne w internecie pod adresem <http://neyman.im.pwr.wroc.pl/~kordecki>.

*Adres
internetowy*

Przy opracowaniu tego konspektu, autor korzystał z wielu podręczników. Przede wszystkim z nieco przestarzałego już, ale dobrego podręcznika M. Fiszera [2] oraz z książek W. Fellera [5] i [6]. Są to jednak książki trudne, przeznaczone raczej dla matematyków i studentów matematyki, niż dla studentów politechnik. Zbiory zadań, to przede wszystkim skrypt T. Inglota, T. Ledwiny i Z. Ławniczak [4] oraz podręcznik J. Grenia [3].

Obecnie na rynku księgarskim i bibliotekach można spotkać wielką liczbę tytułów, poświęconych zagadnieniom rachunku prawdopodobieństwa i statystyki matematycznej. Ich przeglądowi będzie poświęcony ostatni rozdział konspektu.

1. Prawdopodobieństwo

1.1. Aksjomaty prawdopodobieństwa

1.1.1. Przestrzeń zdarzeń

*Zdarzenia
elementarne*

Niech Ω będzie dowolnym zbiorem, zwanym przestrzenią zdarzeń elementarnych. Elementy tej przestrzeni $\omega \in \Omega$ nazywamy zdarzeniami elementarnymi.

Przykład. Rzut monetą: $\Omega = \{\mathcal{O}, \mathcal{R}\}$, gdzie \mathcal{O} jest stroną monety z orłem, a \mathcal{R} jest stroną monety z reszką.

Rzut kostką do gry: $\Omega = \{[1], [2], [3], [4], [5], [6]\}$, gdzie $[i]$ jest tą ścianką kostki, na której jest i oczek.

Strzelanie do tarczy: Ω jest ścianą, na której wyznaczono koło o danej średnicy. Trafienie jest punktem z tej ściany – zakładamy tu, że zawsze w ścianę trafiamy, choć niekoniecznie w tarczę.

Zdarzeniami będziemy nazywać podzbiory przestrzeni zdarzeń elementarnych Ω . Jednakże nie wszystkie takie podzbiory muszą być zdarzeniami. Muszą one jednak spełniać pewne warunki. Aby na przykład określić zdarzenie A jako „wyrzucono co najmniej 5 oczek lub co najwyżej 2 oczka” warto mieć sumę zdarzeń $A = A_{2-} \cup A_{5+}$, gdzie A_{2-} oznacza wyrzucenie co najwyżej 2 oczek, a A_{5+} oznacza wyrzucenie co najmniej 5 oczek. Podobnie, jeśli A będzie zdarzeniem polegającym na trafieniu w tarczę, warto określić zdarzenie przeciwne $A' = \Omega \setminus A$ polegające na nietrafieniu w tarczę. Rozważania te prowadzą nas do następującej definicji.

*σ -algebra
zdarzeń*

Definicja. Niech Ω będzie ustaloną przestrzenią zdarzeń elementarnych. Zdarzeniami nazywamy podzbiory przestrzeni Ω , które tworzą rodzinę (czyli zbiór zbiorów) \mathcal{S} taką, że

- (i) $\emptyset \in \mathcal{S}$,
- (ii) jeżeli $A \in \mathcal{S}$ to $A' \in \mathcal{S}$,
- (iii) jeżeli dla dowolnego ciągu $A_i \in \mathcal{S}$, to

$$A_1 \cup A_2 \cup \dots \cup A_n \dots \in \mathcal{S},$$

gdzie $A = \emptyset$ nazywamy zdarzeniem niemożliwym, a $A' = \Omega \setminus A$ nazywamy zdarzeniem przeciwnym do A . Rodzinę \mathcal{S} nazywamy σ -algebrą zdarzeń.

Występująca w definicji suma może być skończona lub nieskończona. Stosuje się często skrótowy zapis, podobny do zapisu sumy liczb:

$$\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n,$$

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \dots \cup A_n \cup \dots$$

Bezpośrednio z definicji wynikają dalsze własności σ -algebry zdarzeń.

Fakt 1.1.1. Jeżeli Ω jest przestrzenią zdarzeń elementarnych, a \mathcal{S} jest σ -algebrą zdarzeń, to

$$\Omega \in \mathcal{S}. \quad (1.1.1)$$

Dowód. Ponieważ $\emptyset \in \mathcal{S}$ oraz $\Omega = \emptyset'$, to z (i) i (ii) wynika, że $\Omega \in \mathcal{S}$. \square

Zdarzenie $A = \Omega$ nazywamy zdarzeniem pewnym. Podobnie jak dla sumy, stosuje się dla iloczynu zbiorów zapis, podobny do zapisu iloczynu liczb:

$$\bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap \cdots \cap A_n,$$

$$\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \cdots \cap A_n \cap \dots$$

Fakt 1.1.2. Jeżeli dla dowolnego ciągu $A_i \in \mathcal{S}$, to

$$\bigcap_i A_i \in \mathcal{S}.$$

Dowód. Powyższą równość otrzymuje się z prawa de Morgana oraz z własności (ii) i (iii). Ponieważ

$$A = \left(\bigcap_i A_i \right)' = \bigcup_i A_i',$$

oraz $A_i' \in \mathcal{S}$, to również $A = (A')' \in \mathcal{S}$. \square

Występująca w powyższym wzorze suma może być skończona lub nieskończona.

1.1.2. Aksjomaty Kołmogorowa

Podana niżej definicja prawdopodobieństwa zwana aksjomatyczną definicją prawdopodobieństwa, pochodzi od A. N. Kołmogorowa. Była opublikowana w roku 1933 roku¹ i stała się podstawą współczesnej teorii prawdopodobieństwa.

Definicja. Aksjomaty Kołmogorowa. Jeżeli Ω jest przestrzenią zdarzeń elementarnych, a \mathcal{S} jest σ -algebrą zdarzeń, to prawdopodobieństwem nazywamy funkcję $\Pr : \mathcal{S} \rightarrow \mathbb{R}$, czyli funkcję przypisującą liczby zdarzeniom taką, że

$$0 \leq \Pr(A) \leq 1, \quad (1.1.2)$$

$$\Pr(\Omega) = 1, \quad (1.1.3)$$

*Definicja
prawdopodo-
bieństwa*

¹A. N. Kołmogorow (1903 – 1987), miał wtedy zaledwie 30 lat!

jeżeli dla dowolnych $i \neq j$ jest $A_i \cap A_j = \emptyset$, to

$$\Pr\left(\bigcup_i A_i\right) = \sum_i \Pr(A_i). \quad (1.1.4)$$

Jeżeli $A \cap B = \emptyset$ to mówimy, że A i B są wykluczające się. Warunek 1.1.4 oznacza, że dla ciągu parami wykluczających się zdarzeń, prawdopodobieństwo sumy jest równe sumie prawdopodobieństw.

Uwaga. Dla iloczynów zdarzeń nie ma podobnej własności.

Z aksjomatów Kołmogorowa wynikają dalsze własności.

Twierdzenie 1.1.1. *Jeżeli $A \in \mathcal{S}$ to $\Pr(A') = 1 - \Pr(A)$.*

Dowód. Ponieważ $A \cap A' = \emptyset$, to z równości (1.1.3) i (1.1.4) wynika, że $\Pr(A) + \Pr(A') = 1$. \square

Podobnie dowodzi się następującego wyniku.

Twierdzenie 1.1.2. *Jeżeli $B \subset A$, to $\Pr(A \setminus B) = \Pr(A) - \Pr(B)$.*

Kilka następnych użytecznych wzorów, pozostawimy do samodzielnego udowodnienia jako zadania.

*Przestrzeń
probabilistyczna*

Trójkę $(\Omega, \mathcal{S}, \Pr)$ nazywa się przestrzenią probabilistyczną. W następnym punkcie rozpatrzmy dwa specjalne przypadki przestrzeni probabilistycznych. Teraz tylko jeden bardzo prosty (najprostszy nietrywialny) przykład.

Przykład. Rzucamy jeden raz monetą. Wtedy $\Omega = \{\mathcal{O}, \mathcal{R}\}$,

$$\mathcal{S} = \{\emptyset, \{\mathcal{O}\}, \{\mathcal{R}\}, \Omega\}$$

oraz $\Pr(\mathcal{O}) = \Pr(\mathcal{R}) = 1/2$. Wten sposób została określona cała przestrzeń probabilistyczna $(\Omega, \mathcal{S}, \Pr)$.

*Niezależność
zdarzeń*

Niezależność zdarzeń jest własnością nie tylko samych zdarzeń jako zbiorów, ale zależy od określonego na nich prawdopodobieństwa.

Definicja. Ciąg zdarzeń A_1, A_2, \dots (skończony lub nieskończony) jest niezależny, gdy dla dowolnego jego skończonego podciągu $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ zachodzi równość

$$\Pr(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \Pr(A_{i_1}) \Pr(A_{i_2}) \dots \Pr(A_{i_k}). \quad (1.1.5)$$

Jeżeli o pewnych zdarzeniach wiemy, że są niezależne, to znając ich prawdopodobieństwa możemy obliczyć prawdopodobieństwa ich iloczynów.

Przykład. Rzucamy dwoma monetami. Jeżeli wiemy, że rzuty są niezależne i prawdopodobieństwa wyrzucenia zarówno orła jak i reszki są równe, to prawdopodobieństwo wyrzucenia dwóch orłów jest (jak dobrze zresztą wiadomo) równe $\Pr(\mathcal{O}) \cdot \Pr(\mathcal{O}) = (1/2)(1/2) = 1/4$.

1.1.3. Geometryczna i klasyczna definicja prawdopodobieństwa

Rozpatrzmy teraz dwa szczególne przypadki. Jako pierwszy określimy tzw. geometryczną definicję prawdopodobieństwa. Nazwa definicja jest tu myląca, bo tak naprawdę, jest to tylko szczególny przypadek ogólnej, aksjomatycznej definicji prawdopodobieństwa.

*Geometryczna
definicja
prawdopodo-
bieństwa*

Niech Ω będzie pewnym podzbiorem \mathbb{R}^k , gdzie $k = 1, 2, 3$, tzn. Ω jest podzbiorem prostej, (najczęściej odcinkiem), płaszczyzny lub przestrzeni trójwymiarowej. Zdarzeniami z \mathcal{S} będą podzbiory z Ω mające miarę² $m(A)$ – długość, powierzchnię lub objętość. Prawdopodobieństwo zdarzenia A określimy wzorem

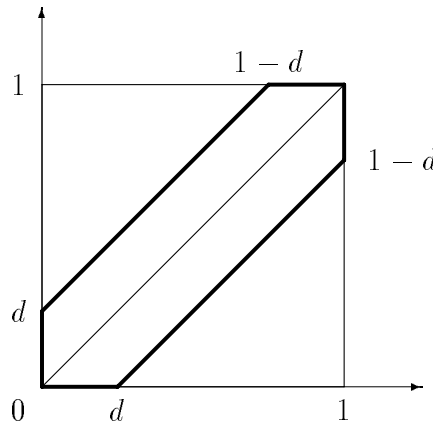
$$\Pr(A) = \frac{m(A)}{m(\Omega)}. \quad (1.1.6)$$

Tak określone prawdopodobieństwo nazywamy geometrycznym.

Przykład. Na odcinku $[0, 1]$ umieszczamy losowo, tzn. zgodnie z prawdopodobieństwem geometrycznym oraz niezależnie, dwa punkty x i y . Przestrzeń Ω możemy wobec tego określić jako kwadrat o wierzchołkach w punktach $(0, 0)$, $(0, 1)$, $(1, 0)$ oraz $(1, 1)$. Teraz zdarzeniami elementarnymi są punkty z tego kwadratu: $\omega = (x, y) \in \Omega$. Dla jednego punktu x zgodnie ze wzorem (1.1.6) mamy $\Pr(\{\omega : x \in [a, b], b > a\}) = b - a$. To samo mamy dla punktu y . Ponieważ punkty są umieszczane niezależnie, to

$$\Pr(\{\omega : \omega \in [a, b] \times [c, d], b > a, d > c\}) = (b - a)(d - c).$$

Iloczyn kartezjański odcinków $[a, b] \times [c, d]$ jest oczywiście prostokątem wewnątrz kwadratu Ω .



Rysunek 1: Prawdopodobieństwo, że $|x - y| < d$.

Niech A_d będzie zdarzeniem polegającym na tym, że $|x - y| < d$, tzn.

$$A_d = \{\omega : |x - y| < d\}.$$

²Nie na wszystkich zbiorach z \mathbb{R}^k da się określić miarę, ale tym zagadnieniem nie będziemy się zajmować

1.1 Aksjomaty prawdopodobieństwa

Zbiór punktów leży wewnątrz obszaru obwiedzionego grubszą linią na rysunku 1. Pole tego obszaru jest równe

$$\Pr(A_d) = 1 - (1 - d)^2.$$

*Klasyczna
definicja
prawdopodo-
bieństwa*

Założmy teraz, że $\Omega = A_1 \cup A_2 \cup \dots \cup A_n$, $A_i \in \mathcal{S}$ oraz wszystkie A_i , $i = 1, 2, \dots, n$ są wykluczające się i mają to samo prawdopodobieństwo, a więc $\Pr(A_i) = 1/n$. Wtedy dla dowolnego zdarzenia A będącego sumą k zdarzeń postaci A_i , czyli dla $A = A_{i_1} \cup \dots \cup A_{i_k}$ ma prawdopodobieństwo $\Pr(A) = k/n$. Zdarzenia $A_i \subset A$ nazywa się wtedy zdarzeniami sprzyjającymi zdarzeniu A , a jego prawdopodobieństwo jest równe stosunkowi liczby zdarzeń sprzyjających do wszystkich n zdarzeń. Tak określone prawdopodobieństwo nosi tradycyjną nazwę „klasycznej definicji prawdopodobieństwa”, choć nie jest definicją, ale pewnym bardzo szczególnym przypadkiem.

Zauważmy, że w tym właśnie przypadku σ -algebrę \mathcal{S} wystarczy ograniczyć do zdarzeń A_i i ich wszystkich sum. Taka σ -algebra jest rodziną 2^n zdarzeń.

Przykład. Rzucamy n razy monetą. Niech A będzie zdarzeniem polegającym na wyrzuceniu m orłów. Przestrzenią zdarzeń elementarnych Ω może tu być zbiór wszystkich ciągów zero-jedynkowych długości n , gdzie jedynka to wyrzucenie orła, a zero to wyrzucenie reszki. Zdarzenia $A_i = \{\omega_i\} = (x_1 x_2 \dots x_n)$, gdzie $i = (x_1 x_2 \dots x_n)_2$ jest liczbą w postaci binarnej odpowiadającą takiemu ciągowi, są wzajemnie wykluczające się i mają te same prawdopodobieństwa. Jest ich 2^n . Zdarzeń sprzyjających zdarzeniu A jest $\binom{n}{m}$. Stąd

$$\Pr(A) = \frac{\binom{n}{m}}{2^n}.$$

1.1.4. Zadania

1. Doświadczenie polega na rzucaniu kostką do gry, aż do wyrzucenia po raz pierwszy sześciu oczek. Określić przestrzeń probabilistyczną.
2. Uogólnić zadanie 1 na przypadek rzucania n kostkami, aż do wyrzucenia po raz pierwszy co najmniej k szóstek w jednym rzucie, $k \leq n$.
3. W zadaniach 1 i 2 obliczyć prawdopodobieństwo, że liczba rzutów będzie równa m , $m = 1, 2, \dots$.
4. Długość l zgięto w dwóch niezależnie wybranych punktach. Obliczyć prawdopodobieństwo, że można w ten sposób utworzyć trójkąt.
5. Z odcinka $[0, 1]$ wybieramy losowo dwie liczby p i q . Jakie jest prawdopodobieństwo tego, że równanie $x^2 + px + q = 0$ będzie miało dwa sprzężone pierwiastki zespolone? Jakie jest prawdopodobieństwo tego, że będzie miało dwa różne pierwiastki rzeczywiste?
6. Dla jakich zdarzeń A i B zachodzi wzór

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A) \cdot \Pr(B) ?$$

7. Udowodnić, że $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

8. Wzorując się na zadaniu 7, sformułować wzór na $\Pr(A \cup B \cup C)$ i spróbować go uogólnić na przypadek sumy n zdarzeń, tzn. znaleźć wzór na obliczenie

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

9. Udowodnić, że

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k},$$

nie wykonując rachunków.

10. W urnie jest $m \geq 3$ białych i $n \geq 3$ czarnych kul. Obliczyć prawdopodobieństwo wylosowania trzech czarnych kul, gdy kule losujemy

- a) bez zwracania,
- b) ze zwracaniem.

1.2. Prawa wielkich liczb i symulacja

1.2.1. Mocne i słabe prawo wielkich liczb

Według potocznych opinii, jeżeli przeprowadzając n obserwacji, zaobserwujemy interesujące nas zjawisko k razy, to prawdopodobieństwo zajścia tego zjawiska powinno wynosić k/n . Iloraz ten jest często przyjmowany za tzw. „statystyczną definicję prawdopodobieństwa”. Określenie to nie jest całkiem poprawne, ale intuicyjnie uzasadnione, a poniższe twierdzenia nadają mu ścisły matematycznie charakter.

Mocne prawo wielkich liczb

Twierdzenie 1.2.1. Niech $(\Omega, \mathcal{S}, \Pr)$ będzie przestrzenią probabilistyczną, a $\{A_i \in \mathcal{S}\}$ będzie nieskończonym ciągiem zdarzeń niezależnych o tym samym prawdopodobieństwie $\Pr(A_i) = p$. Niech $\omega \in \Omega$ będzie dowolnym, ale ustalonym zdarzeniem elementarnym. Oznaczmy przez $N(n, \omega)$ liczbę tych zdarzeń A_i , dla których $\omega \in A_i$, gdy $i = 1, 2, \dots, n$. Wtedy

$$\Pr \left(\left\{ \omega : \lim_{n \rightarrow \infty} \frac{N(n, \omega)}{n} = p \right\} \right) = 1. \quad (1.2.1)$$

Słabe prawo wielkich liczb

Twierdzenie 1.2.2. Przy oznaczeniach z twierdzenia 1.2.1 dla każdego $\varepsilon > 0$ zachodzi równość

$$\lim_{n \rightarrow \infty} \Pr \left(\left\{ \omega : \left| \frac{N(n, \omega)}{n} - p \right| > \varepsilon \right\} \right) = 0. \quad (1.2.2)$$

Twierdzenie 1.2.1 nosi nazwę mocnego prawa wielkich liczb, a twierdzenie 1.2.2 – słabego prawa wielkich liczb. Tak też jest. Z mocnego prawa wielkich liczb łatwo jest wyprowadzić słabe prawo wielkich liczb. Dowodów tych nie będziemy tu przeprowadzać, a słabe prawo wielkich liczb udowodnimy później, w innej, łatwiejszej i ogólniejszej postaci. W tym momencie zajmijmy się interpretacją tych twierdzeń.

W twierdzeniu 1.2.1 zdarzenie elementarne ω jest ustalonym doświadczeniem, polegającym na nieskończonym ciągu obserwacji. W każdej obserwacji badamy, czy zaszło interesujące nas zjawisko, tzn. czy w i -tej obserwacji było $\omega \in A_i$. Tak więc $N(n, \omega)$ jest liczbą zaobserwowanych zjawisk w pierwszych n obserwacjach, w ustalonym doświadczeniu ω . Równość (1.2.1) mówi, że stosunek liczby zjawisk zaobserwowanych w n obserwacjach do liczby obserwacji równej n , dąży do prawdopodobieństwa zaobserwowania zjawiska przy jednej obserwacji dla prawie wszystkich doświadczeń, tzn. mogą się zdarzyć doświadczenia, w których nie zajdzie równość (1.2.1), ale stanie się to z zerowym prawdopodobieństwem.

Podobny charakter ma twierdzenie 1.2.2. Mówi ono, że stosunek liczby zjawisk zaobserwowanych w n obserwacjach do liczby obserwacji różni się niewiele od prawdopodobieństwa zaobserwowania zjawiska przy jednej obserwacji, tzn. prawdopodobieństwo, że różni się więcej niż dowolnie mała liczba ε , dąży do zera, a więc jest dowolnie małe, jeśli tylko weźmiemy dostatecznie duże n .

Przykład. Rzucamy n razy kostką do gry, a n jest „bardzo duże”. Jeżeli liczba wyrzuconych szóstek $N(n)$ daje nam $N(n)/n$ znacznie różniące się od $1/6$, to w myśl twierdzenia 1.2.2 albo zabserwowaliśmy zjawisko rzadkie, które powinno zdarzyć się z małym prawdopodobieństwem (dokładniej, z prawdopodobieństwem dążącym do zera) albo kostka jest fałszywa, tzn. prawdopodobieństwo wyrzucenia szóstki jest różne od $1/6$. Twierdzenie 1.2.1 mówi zaś więcej – zaobserwowaliśmy zjawisko, które zdarzyć się w ogóle nie powinno, bo może zajść z prawdopodobieństwem zero, choć oczywiście tylko w granicy, przy $n \rightarrow \infty$.

1.2.2. Pierwsze przykłady symulacji

Przedstawimy tu dwa przykłady wykorzystania praw wielkich liczb do symulacji, zwanej metodą Monte-Carlo.

Igła Buffona

Przykład. Na płaszczyznę poliniowaną liniami równoległymi, odległymi od siebie o 1, rzucamy losowo igłę o długości 1. Jakie jest prawdopodobieństwo, że igła przetnie linię?

Aby rozwiązać ten przykład, musimy sprecyzować, co rozumiemy przez określenie „na poliniowaną płaszczyznę rzucamy losowo igłę o danej długości”. W tym celu założmy, że linie na płaszczyźnie są ułożone poziomo oraz wprowadźmy oznaczenia: x – odległość dolnego końca igły od najbliższej linii pod igłą, α – kąt między igłą a linią, $0 \leq \alpha < \pi$. Położenie igły jest teraz określone przez parę liczb (x, α) . Losowość rzutu na płaszczyznę rozumiemy w tym przykładzie jako umieszczenie punktu (x, α) w prostokącie $[0, 1] \times [0, \pi]$, zgodnie z geometryczną definicją prawdopodobieństwa. Niech A będzie zdarzeniem polegającym na tym, że igła przetnie linię.

Igła przecina linię wtedy, gdy $\sin \alpha > 1 - x$. Ponieważ pole obszaru spełniającego ten warunek jest równe

$$\int_0^{\pi} \sin \alpha \, d\alpha = 2,$$

a pole prostokąta jest równe π , to $\Pr(A) = 2/\pi$. Jeżeli teraz r_1, r_2, \dots będzie niezależnym ciągiem punktów, każdy z odcinka $[0, 1]$ wybranym zgodnie z geometryczną definicją prawdopodobieństwa, to ciąg $(r_1, \pi r_2), (r_3, \pi r_4), \dots$ będzie niezależnym ciągiem punktów, każdy z prostokąta $[0, 1] \times [0, \pi]$ wybranym zgodnie z geometryczną definicją prawdopodobieństwa. Niech teraz A_i będzie zdarzeniem polegającym na tym, że w i -tym rzucie igła przetnie linię. Zgodnie z twierdzeniem 1.2.1 i przyjętymi w nim oznaczeniami, ze wzoru (1.2.1) wynika, że dokonując coraz to większej liczby rzutów, stosunek ich liczby n do liczby tych rzutów k w których igła przecięła linię, n/k , będzie dążył prawie zawsze do $\pi/2$. Dla n skończonego, n/k będzie tylko przybliżeniem liczby π , o losowym błędzie. Oszacowaniem tego błędu zajmiemy się później.

Symulacja polega na dokonaniu takiego eksperymentu i obliczeniu n/k . Gdy punkty r_i są otrzymywane nie z fizycznego eksperymentu, a generowane są na

drodze programowej, to taki komputerowy eksperyment daje wygodną i taną metodę otrzymywania przybliżeń – w wielu przypadkach wygodniejszą od czysto numerycznych obliczeń.

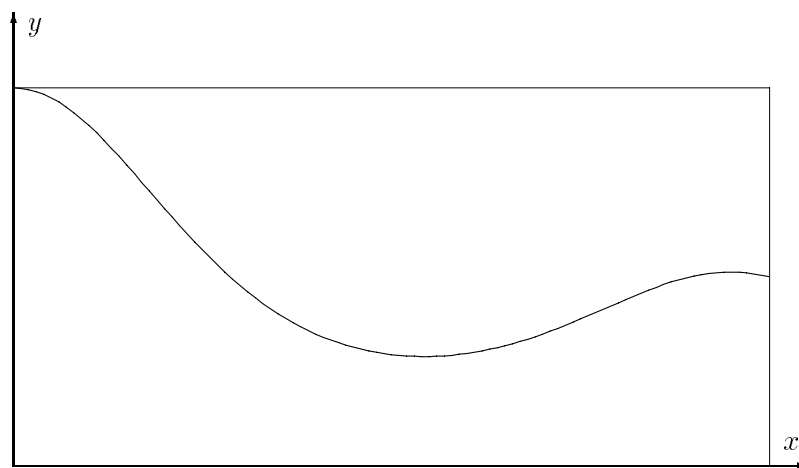
Przykład ten nie ma praktycznego znaczenia. Wartość liczby π jest znana z wystarczającą dokładnością, lepszą od możliwej do otrzymania metodą symulacji. Następny przykład może mieć już znaczenie nieco bardziej praktyczne.

*Całka
obliczona
metodą Monte
Carlo*

Przykład. Obliczyć całkę

$$\int_0^1 \frac{e^{\sin^2 x}}{1+x^2} dx ,$$

(wykres funkcji podcałkowej na rysunku 2).



Rysunek 2: Wykres funkcji z przykładu

Całki tej nie da się obliczyć analitycznie. Metoda Monte Carlo polega na otrzymaniu ciągu n niezależnych punktów (x_i, y_i) z prostokąta $[0, 1] \times [0, 1]$, a następnie obliczeniu liczby k tych punktów, które spełniają nierówność

$$y_i < \frac{e^{\sin^2 x_i}}{1+x_i^2} .$$

Wtedy na podstawie prawa wielkich liczb (twierdzenia 1.2.1 i 1.2.2) można przyjąć, że

$$k/n \approx \int_0^1 \frac{e^{\sin^2 x}}{1+x^2} dx ,$$

gdyż maksimum funkcji podcałkowej wynosi 1.

*Symulacja
w Turbo
Pascalu*

Procedura

```
const n=1000;
var i,k:integer;
    x,y:double;
begin
    randomize;
    k:=0;
    for i:=1 to n do
        begin
            x:=random;
            y:=random;
            if y<exp(-sqr(sin(pi*x)))/(1+sqr(x)) then inc(k);
        end;
        writeln('Całka=',k/n:0:4);
    end;
```

dała dla 5 kolejnych przebiegów³ liczby:

0.5320, 0.4870 0.5340 0.4610 0.5160,

a ta sama procedura, ale dla $n = 10\,000$ dała

0.5045, 0.5011, 0.5046, 0.5006, 0.5045,

podczas gdy prawdziwa wartość całki (obliczona numerycznie przy pomocy programu *Mathematica*) jest równa 0.503498.

1.2.3. Zadania

1*. Niech Ω będzie kwadratem o boku $2R$ oraz A będzie kołem wpisanym w ten kwadrat. Podać metodę obliczania liczby π metodą Monte Carlo i napisać odpowiednią procedurę obliczającą.

2*. Obliczyć całkę

$$\int_0^1 \frac{\arctg(x^2 \sin \pi x)}{\ln^2(1 + (e - 1)x)} dx$$

metodą Monte Carlo.

3*. Hipocykloida jest krzywą określoną przez punkt okręgu koła o promieniu r toczącego się bez poślizgu po wewnętrznej stronie okręgu stałego koła o promieniu R . Jeżeli $m = R/r > 2$ jest liczbą całkowitą, to krzywa ta nie przecina się i ogranicza obszar wewnątrz koła. Dla $m = 4$ krzywa taka jest znana jako *asteroida*.

Napisać procedury obliczające metodą Monte Carlo pole ograniczone takimi krzywymi dla $m > 2$. Porównać te wyniki z wynikami dokładnymi. Dla $m = 3$ pole wynosi $S = 2\pi r^2$, a dla $m = 4$ pole wynosi $S = \frac{3}{8}\pi A^2$.

³Dla następnych 5 przebiegów dostanie się zapewne inne liczby

1.3. Prawdopodobieństwo warunkowe

1.3.1. Definicja i podstawowe własności

Prawdopodobieństwem warunkowym zdarzenia A jest prawdopodobieństwo **tego samego zdarzenia** zwanego skutkiem, ale mające inną wartość, zmodyfikowaną na podstawie informacji o innym zdarzeniu, zwanym przyczyną lub warunkiem. Nie ma natomiast pojęcia zdarzenia warunkowego.

Definicja. Prawdopodobieństwo warunkowe zdarzenia A pod warunkiem zdarzenia B określone jest wzorem

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}, \quad (1.3.1)$$

przy założeniu $\Pr(B) > 0$.

Bezpośrednio z definicji otrzymuje się następujący

Fakt 1.3.1.

$$\Pr(A \cap B) = \Pr(A|B) \Pr(B), \quad (1.3.2)$$

dla $\Pr(B) > 0$.

Własność wyrażona wzorem (1.3.2) jest bardzo użyteczna w sytuacji, gdy prawdopodobieństwa warunkowe są znane w sposób naturalny, a nieznane są prawdopodobieństw iloczynów.

Przykład. W ciemnym pokoju mamy dwie urny: dużą, do której trafiamy z prawdopodobieństwem $3/4$ i małą do której trafiamy z prawdopodobieństwem $1/4$. W małej urnie mamy 2 czarne i 1 białą kulę, a dużej 2 białe i 1 czarną kulę. Jakie jest prawdopodobieństwo, że trafimy na dużą urnę i równocześnie wyciągnięta z urny kula będzie biała.

Rozwiązanie. Niech A będzie zdarzeniem polegającym na trafieniu do dużej urny, B zdarzenie polegające na wylosowaniu kuli białej. Jest jasne, że $\Pr(B|A) = 2/3$, a ponieważ $\Pr(A) = 3/4$, to zgodnie ze wzorem (1.3.2) otrzymujemy

$$\Pr(A \cap B) = \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{2}.$$

Prawdopodobieństwo warunkowe ma wszystkie własności „zwykłego” prawdopodobieństwa, tzn. zachodzi następujące twierdzenie.

Twierdzenie 1.3.1. Dla dowolnej przestrzeni probabilistycznej $(\Omega, \mathcal{S}, \Pr)$ i ustalonego $B \in \mathcal{S}$ takiego, że $\Pr(B) > 0$, prawdopodobieństwo warunkowe $\Pr^B(\cdot) = \Pr(\cdot|B)$ spełnia postulaty Kołmogorowa (1.1.2) – (1.1.4), tzn. $(\Omega, \mathcal{S}, \Pr^B)$ jest przestrzenią probabilistyczną.

Intuicje dotyczące prawdopodobieństwa warunkowego zgodne są z następującym rezultatem.

Twierdzenie 1.3.2. Jeżeli zdarzenia A i B są niezależne i $\Pr(B) > 0$, to

$$\Pr(A|B) = \Pr(A).$$

1.3.2. Wzór Bayesa

Twierdzenia przedstawione w tym punkcie znane są pod tradycyjną nazwą wzorów. Trzeba jednak zwrócić uwagę, że wzory te są prawdziwe tylko przy spełnionych założeniach, wspólnych dla obydwóch twierdzeń.

Wzór na
prawdopodo-
bieństwo
całkowite

Twierdzenie 1.3.3. (prawdopodobieństwo całkowite). Niech A_i będzie ciągiem zdarzeń takim, że $A_i \cap A_j = \emptyset$ dla $i \neq j$ oraz $\sum_i \Pr(A_i) = 1$. Wtedy

$$\Pr(B) = \sum_i \Pr(B|A_i)P(A_i). \quad (1.3.3)$$

Dowód. Twierdzenie udowodnimy w szczególnym przypadku, gdy założenie $\sum_i \Pr(A_i) = 1$ zastąpimy mocniejszym: $\bigcup_i A_i = \Omega$. Ze wzoru (1.3.2) otrzymujemy zależność $\Pr(B|A_i)P(A_i) = \Pr(A_i \cap B)$. Ponieważ A_i są rozłączne, to również rozłączne są $B \cap A_i$ oraz $B = B \cap \bigcup_i A_i$. Stąd

$$\Pr(B) = \Pr\left(B \cap \bigcup_i A_i\right) = \Pr\left(\bigcup_i (B \cap A_i)\right) = \sum_i \Pr(B|A_i)P(A_i).$$

□

Bezpośrednio ze wzorów (1.3.1) i (1.3.2) otrzymuje się następujące

Wzór Bayesa

Twierdzenie 1.3.4. (Bayesa⁴). Przy założeniach z twierdzenia 1.3.3 oraz dla $\Pr(B) > 0$:

$$\Pr(A_i|B) = \frac{\Pr(B|A_i) \Pr(A_i)}{\Pr(B)}, \quad (1.3.4)$$

gdzie $\Pr(B)$ wyznaczone jest ze wzoru (1.3.3).

Przykład. Dwóch strzelców strzela do tarczy. Strzelec 1 trafia z prawdopodobieństwem $2/3$, a strzelec 2 z prawdopodobieństwem $1/2$. Po oddaniu po jednym strzale okazało się, że tarcza została trafiona dokładnie raz. Jakie jest prawdopodobieństwo, że trafił strzelec 1?

Oznaczmy przez A_{11} fakt trafienia w tarczę przez obydwóch strzelców, A_{10} – przez pierwszego, A_{01} – przez drugiego, a przez A_{00} przez żadnego. Ponieważ strzelcy strzelają niezależnie (to trzeba założyć), to $\Pr(A_{11}) = 1/3$, $\Pr(A_{10}) = 1/3$, $\Pr(A_{01}) = 1/6$, $\Pr(A_{00}) = 1/6$. Wszystkie te zdarzenia są rozłączne, więc są spełnione założenia twierdzenia 1.3.4. Przez B oznaczmy fakt trafienia w tarczę dokładnie raz. Szukanym prawdopodobieństwem jest

$$\Pr(A_{10} \cup A_{01}|B) = \Pr(A_{10}|B),$$

gdyż $\Pr(A_{11}|B) = 0$. Wiadomo też, że $\Pr(B|A_{10})\Pr(A_{10}) = 1/3$ oraz $\Pr(B|A_{01})\Pr(A_{10}) = 1/6$, pozostałe prawdopodobieństwa warunkowe są równe

⁴Thomas Bayes †1763

zeru. Tak więc ze wzoru (1.3.3) otrzymujemy $\Pr(B) = 1/2$. Ze wzoru Bayesa (1.3.4) wynika, że

$$\Pr(A_{10}|B) = \frac{1/3}{1/2},$$

co jest szukany prawdopodobieństwem.

Występujące w następnym punkcie zadania 1 i 2 również są typowymi zastosowaniami wzoru na prawdopodobieństwo całkowite i wzoru Bayesa.

1.3.3. Zadania

1. W każdej z trzech urn są po cztery kule – w urnie i -tej jest i kul białych, a reszta czarnych. Z urny pierwszej losujemy jedną kulę i przekładamy do drugiej, z drugiej losujemy jedną kulę i przekładamy do trzeciej, a w końcu z trzeciej urny losujemy jedną kulę. Jakie jest prawdopodobieństwo, że będzie to kula biała?
2. W urnie pierwszej są 2 dwie białe i dwie czarne kule, a w drugiej jedna biała i trzy czarne kule. Z każdej urny losujemy po jednej kuli i następnie losujemy z nich jedną kulę. Okazało się, że wylosowaliśmy kulę białą. Jakie jest prawdopodobieństwo, że kula ta pochodziła z pierwszej urny.
3. Udowodnić, że jeśli $\Pr(A|B) = \Pr(A|B')$, to zdarzenia A i B są niezależne.
4. Przy szeregowej transmisji danych do każdego bajtu dodawany jest jeden bit tak, aby liczba jedynek była parzysta. Prawdopodobieństwo przekłamania dla każdego bitu wynosi p . Obliczyć prawdopodobieństwo niewykrycia przekłamania – dokładne dla dowolnego p oraz przybliżone, gdy p jest bardzo małe.

2. Zmienne losowe

2.1. Rozkłady zmiennych losowych

2.1.1. Definicja zmiennej losowej

Formułowanie problemów wyłącznie w terminach przestrzeni zdarzeń elementarnych, σ -algebry zdarzeń i prawdopodobieństwa określanego bezpośrednio na zdarzeniach, nie jest zbyt wygodne. Często bowiem zamiast określenia całej przestrzeni probabilistycznej $(\Omega, \mathcal{S}, \Pr)$, wystarczy określić charakterystyki liczbowe zdarzeń elementarnych, bez opisywania samych zdarzeń. Na przykład jeśli przy rzutach monetą interesuje nas tylko liczba wyrzuconych orłów, to zamiast opisywać rzut monetą w terminach orłów i reszek, można zastąpić je zerami i jedynkami i dodawać tylko te „losowe” liczby. Co to więc są te „losowe” liczby? Potrzebna jest tutaj następująca definicja.

Zmienna losowa

Definicja. Zmienną losową nazywamy funkcję określoną na przestrzeni zdarzeń elementarnych Ω i przyjmującą wartości rzeczywiste:

$$X : \Omega \rightarrow \mathbb{R},$$

taką, że

$$\{\omega : X(\omega) < x\} \in \mathcal{S}, \quad (2.1.1)$$

dla każdego $x \in \mathbb{R}$.

Zamiast pisać $A = \{\omega : X(\omega) < x\}$ będzie pisać w skrócie $A = \{X < x\}$. Definicja powyższa zapewnia nam, że można powiedzieć: „zdarzenie polegające na tym, że $X < x$ ”.

Bezpośrednio z definicji wynika również, że zdarzeniami są również zbiory:

$$\begin{aligned} &\{\omega : X(\omega) \leq x\}, \\ &\{\omega : X(\omega) > x\}, \\ &\{\omega : X(\omega) \geq x\}, \\ &\{\omega : X(\omega) = x\}, \\ &\{\omega : X(\omega) \in [a, b]\}, \\ &\{\omega : X(\omega) \in [a, b)\}, \\ &\{\omega : X(\omega) \in (a, b]\}, \\ &\{\omega : X(\omega) \in (a, b)\}. \end{aligned}$$

Również w tych przypadkach najczęściej stosować będziemy skrótowy zapis, np. $\{X \in [0, 1]\}$ zamiast pełnego $\{\omega : X(\omega) \in [0, 1]\}$, pamiętając jednak zawsze, że $X = X(\omega)$, czyli że zmienna losowa jest funkcją zdarzeń elementarnych.

Będziemy również pisać $\Pr(X < x)$ zamiast $\{\omega : X(\omega) < x\}$. Podobnie we wszystkich dalszych przypadkach.

Na zmiennych losowych można dokonywać rozmaitych operacji, co precyzuje następujące twierdzenie.

Twierdzenie 2.1.1. *Jeżeli $X(\omega)$ jest zmienną losową, a $h(x)$ jest funkcją przedziałami ciągłą, której dziedzina zawiera zbiór wartości X , to $Y(\omega) = h(X(\omega))$ jest też zmienną losową określoną na tej samej przestrzeni probabilistycznej co X .*

2.1.2. Dystrybuanta zmiennej losowej

Definicja zmiennej losowej zapewnia nam poprawność definicji podanej niżej.

Dystrybuanta

Definicja. Dystrybuantą nazywamy funkcję rzeczywistą zmiennej rzeczywistej, określoną wzorem

$$F(x) = \Pr(\{\omega : X(\omega) < x\}) = \Pr(X < x). \quad (2.1.2)$$

Warunki konieczne i dostateczne

Twierdzenie 2.1.2. *Funkcja F jest dystrybuantą pewnej zmiennej losowej wtedy i tylko wtedy, gdy*

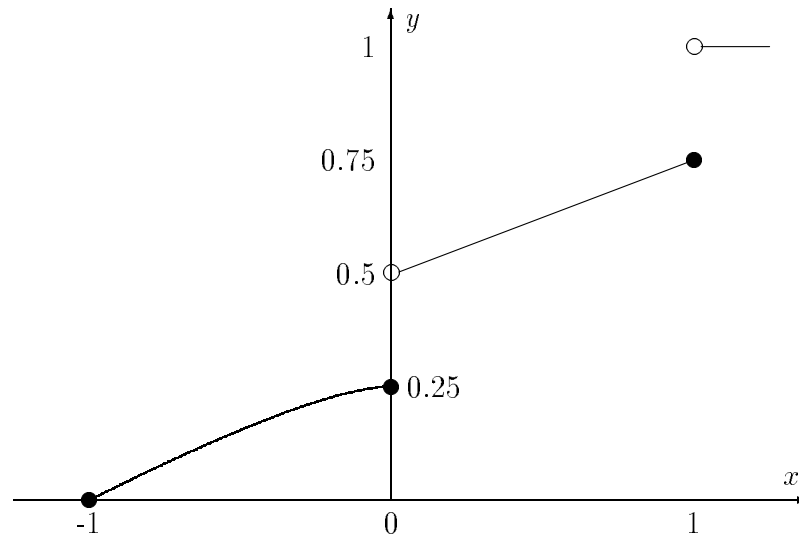
- (a) *jest niemalejąca,*
- (b) $\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1,$
- (c) *jest lewostronnie ciągła.*

Przykład. Należy dobrać stałe A, B, C, D, E i F tak, aby funkcja

$$F(x) = \begin{cases} A & \text{dla } x < -1, \\ Bx^2 + C & \text{dla } -1 \leq x \leq 0, \\ Dx + E & \text{dla } 0 < x \leq 1, \\ F & \text{dla } x > 1 \end{cases}$$

była dystrybuantą pewnej zmiennej losowej. Trzeba więc stałe dobrać tak, aby były spełnione warunki (a) – (c) w twierdzeniu 2.1.2. Najpierw sprawdzimy warunek (b). Wynika z niego, że musi być $A = 0$ oraz $F = 1$. Na to, aby dla $x = -1$ był spełniony warunek (c), $F(x)$ musi być lewostronnie ciągła w tym punkcie, a więc musi być również ciągła, (bo inaczej byłaby tylko prawostronnie ciągła) co oznacza, że $B + C = 0$. Dla spełnienia warunku (a) na odcinku $[0, 1]$ trzeba przyjąć, że $B \leq 0$, a więc również $C \geq 0$. Dalej, w punktach $x = 0$ i $x = 1$ funkcja $F(x)$ jest zawsze lewostronnie ciągła, wystarczy więc sprawdzić warunek (a). Oznacza to, że $C \leq E \leq 1$, $D \geq 0$ oraz $D \leq 1 - E$. Tak więc tylko dwa parametry są wyznaczone jednoznacznie, ($A = 0$ i $F = 1$), a pozostałe są określone przy pomocy układu nierówności i równości. Na rysunku 3 pokazany jest wykres takiej dystrybuanty dla $B = -0.25$, $C = -B = 0.25$, $D = 0.25$, $E = 0.5$.

Przy pomocy dystrybuanty można określić prawdopodobieństwa zdarzeń, o



Rysunek 3: Wykres dystrybuanty z przykładu

których była mowa punkcie 2.1.1.

$$\Pr(X \leq x) = \lim_{t \rightarrow x^+} F(t) = F(x^+),$$

$$\Pr(x_1 \leq X < x_2) = F(x_2) - F(x_1),$$

$$\Pr(X = x) = F(x^+) - F(x),$$

$$\Pr(X > x) = 1 - F(x^+)$$

i tak dalej.

Wśród zmiennych losowych można, ze względu na postać dystrybuanty, wyróżnić dwa typy:

- zmienne losowe typu skokowego (zmienne losowe dyskretne), których dystrybuanta jest przedziałami stała, a ponadto ma tylko skoki w punktach x_i – taka zmienna losowa może przyjmować (z prawdopodobieństwem 1) tylko wartości x_i ,
- zmienne losowe typu ciągłego, których dystrybuantę $F(x)$ można przedstawić w postaci

$$F(x) = \int_{-\infty}^x f(t) dt. \quad (2.1.3)$$

Funkcję $f(x)$ określoną wzorem (2.1.3) nazywa się gęstością.

Uwaga. Nie ma zmiennych losowych ciągłych, bo nie są one funkcjami zmiennej rzeczywistej, a mogą być określone na zbiorach dowolnej natury. Są tylko zmienne losowe **typu ciągłego**, tzn. takie, które mają gęstość.

Własności gęstości omówimy dokładniej w następnym punkcie.

Rozkładem zmiennej losowej określamy jej dystrybuantę lub inne funkcje w pełni ją charakteryzujące, np. gęstość dla zmiennej losowej typu ciągłego lub prawdopodobieństwa $p_i = \Pr(X = x_i)$ dla zmiennej losowej dyskretnej.

W dwóch następnych punktach omówimy najważniejsze rozkłady dyskretne i rozkłady typu ciągłego.

*Niezależność
X i Y*

Niezależność zmiennych losowych definiuje się przy pomocy niezależności zdarzeń lub przy pomocy dystrybuant. Tutaj podamy tylko definicję niezależności dwóch zmiennych losowych, częściowo w terminach dystrybuant. Ogólna definicja zostanie podana później, przy okazji rozpatrywania wektorów losowych i ich dystrybuant.

Definicja. Zmienne losowe X i Y są niezależne, gdy

$$\Pr(X < x, Y < y) = F(x)G(y),$$

gdzie F jest dystrybuantą zmiennej losowej X , a G – zmiennej losowej Y .

2.1.3. Gęstość

Wzór (2.1.3) definiuje gęstość rozkładu. Mówi się też w przypadku, gdy $F(x)$ jest dystrybuantą zmiennej losowej X , że $f(x)$ jest gęstością zmiennej losowej X mimo, że X nie występuje jawnie w definicji gęstości.

*Warunki
konieczne
i dostateczne*

Podobnie jak twierdzenie 2.1.2 podaje warunki konieczne i dostateczne na to, aby dana funkcja była dystrybuantą, poniższe twierdzenie podaje warunki konieczne i dostateczne na to, aby funkcja była gęstością.

Twierdzenie 2.1.3. Funkcja $f(x)$ jest gęstością pewnej zmiennej losowej wtedy i tylko wtedy, gdy

(A) $f(x) \geq 0,$

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

Należy zwrócić uwagę, że gęstość nie musi być określona jednoznacznie, ponieważ można w dowolny sposób ustalić jej wartość w skończonej liczbie punktów. Dzieje się tak, gdyż zarówno we wzorze (2.1.3) definiującym gęstość jak i w warunku (B) w twierdzeniu 2.1.3, wartości całek nie zmieniają się przy zmianie wartości funkcji podcałkowej w skończonej liczbie punktów. Przyjmuje się jednak zawsze, że spełniony jest warunek (A) w twierdzeniu 2.1.3.

*Wzór na
gęstość*

Bezpośrednim wnioskiem ze wzoru (2.1.3) jest

99990TDTiD1000o1006s9999el442

była gęstością pewnej zmiennej losowej. Trzeba więc dobrać stałe tak, aby był spełniony warunek (A) w twierdzeniu 2.1.3, skąd wynika $a \geq 0$ oraz $b \leq \pi/2$. Dla spełnienia warunku (B) w tym twierdzeniu, musi być spełniona równość

$$\int_0^b \cos x \, dx = 1.$$

Stąd po obliczeniu całki otrzymujemy warunek $a \sin b = 1$, przy czym $a > 0$.

Za pomocą gęstości można łatwo wyznaczyć prawdopodobieństwo należenia wartości zmiennej losowej do zbioru.

Twierdzenie 2.1.4. *Jeżeli zmienna losowa X ma gęstość $f(x)$, to*

$$\begin{aligned} \Pr(X \in [a, b]) &= \Pr(X \in [a, b)) \\ &= \Pr(X \in (a, b]) = \Pr(X \in (a, b)) = \int_a^b f(x) \, dx. \end{aligned} \quad (2.1.5)$$

2.1.4. Zadania

1. Dla jakich parametrów a i b funkcja $F(x) = a \arctg x + b$ jest dystrybuantą pewnej zmiennej losowej?

2. Niech

$$F(x) = \begin{cases} 1 - e^{-x^2}, & \text{dla } x \geq 0, \\ 0, & \text{dla } x < 0, \end{cases}$$

będzie dystrybuantą zmiennej losowej X , (sprawdzić, że jest to dystrybuanta). Wyznaczyć $\Pr(X > 1)$, $\Pr(1 < X \leq 2)$, $\Pr(X = 3)$.

3. Dla jakich wartości a i b funkcja $f(x) = \frac{a}{1 + (bx)^2}$ jest gęstością pewnej zmiennej losowej? Wyznaczyć dystrybuantę.

4. Niech dystrybuanta $F(x)$ zmiennej losowej X będzie funkcją ściśle rosnącą i ciągłą, a funkcja $h(x)$ będzie różnowartościowa. Znaleźć dystrybuantę zmiennej losowej $h(X)$.

5. Niech

$$f(x) = \begin{cases} a & \text{dla } x \in [0, \pi], \\ 0 & \text{dla } x \notin [0, \pi] \end{cases}$$

będzie gęstością zmiennej losowej X . Wyznaczyć a , znaleźć dystrybuantę zmiennej losowej X oraz korzystając z zadania 4 znaleźć dystrybuantę i gęstość zmiennej losowej $Y = \sin X$.

2.2. Momenty zmiennych losowych

2.2.1. Całka Stieltjesa

Podana w tym punkcie definicja całki Stieltjesa⁵ jest uogólnieniem zwykłej całki. Jej użycie umożliwia podanie jednolitej definicji momentów zmiennych losowych, zarówno dla typu ciągłego, jak i skokowego.

Definicja. Załóżmy, że dystrybuenta $F(x)$ jest przedziałami ciągła i różniczkowalna. Całką Stieltjesa z funkcji $g(x)$ względem dystrybuenty $F(x)$ nazywamy liczbę, określoną wzorem

*Całka
Stieltjesa*

$$\int_a^b g(x) dF(x) = \int_a^b g(x) \frac{dF(x)}{dx} dx + \sum_i g(x_i) (F(x_i^+) - F(x_i)) , \quad (2.2.1)$$

gdzie x_i są punktami skoków dystrybuenty $F(x)$, pomiędzy którymi jest ona ciągła i różniczkowalna.

*Przypadki
szczególne
całki Stieltjesa*

Jeżeli dystrybuenta $F(x)$ ma gęstość, to we wzorze (2.2.1) znika suma, (bo dystrybuenta jest ciągła, więc skoki są równe zero), a całka Stieltjesa sprowadza się do zwykłej całki Riemanna. Jeśli natomiast zmienna losowa jest typu skokowego, to pochodna pomiędzy skokami jest równa zero, (bo dystrybuenta jest tam stała), a całka Stieltjesa redukuje się we wzorze (2.2.1) do sumy.

Przykład. Określmy dystrybuentę $F(x)$ wzorem

$$F(x) = \begin{cases} 0 & \text{dla } x \leq 0, \\ 1 - e^{-x}/2 & \text{dla } x > 0. \end{cases}$$

Wtedy skok w zerze $F(0^+) - F(0) = 1/2$, całka w przedziale $(-\infty, 0)$ jest równa zero, a więc

$$\int_{-\infty}^{\infty} (x+1) dF(x) = \frac{1}{2} \int_0^{\infty} (x+1) e^{-x} dx + \frac{1}{2} = \frac{3}{2}.$$

2.2.2. Wartość oczekiwana

Wartość oczekiwana zmiennej losowej, zwana też nadzieją matematyczną, wartością średnią lub przeciętną, należy do podstawowych pojęć rachunku prawdopodobieństwa.

Definicja. Wartością oczekiwaną zmiennej losowej $g(X)$ o dystrybuancie $F(x) = \Pr(X < x)$ jest liczba oznaczana jako $Eg(X)$ i określona wzorem

$$Eg(X) = \int_{-\infty}^{\infty} g(x) dF(x) , \quad (2.2.2)$$

⁵Thomas Stieltjes, (1856-1895), matematyk holenderski.

o ile całka we wzorze (2.2.2) jest bezwzględnie zbieżna, to znaczy jeżeli istnieje całka

$$\int_{-\infty}^{\infty} |g(x)| dF(x) . \quad (2.2.3)$$

*Wartość
oczekiwana –
przypadki
szczególne*

Jeżeli zmienna losowa jest typu ciągłego o gęstości $f(x)$, to wartość oczekiwana zamiast wzorem (2.2.2), wyraża się wzorem

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

a jeżeli jest dyskretna, to wzorem

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

gdzie $p_i = \Pr(X = x_i)$. Oczywiście w obydwóch tych przypadkach, suma lub całka musi być bezwzględnie zbieżna.

Jeżeli całka (2.2.2) nie istnieje lub jeśli istnieje, ale nie jest bezwzględnie zbieżna, a więc nie istnieje całka (2.2.3), to mówimy, że wartość oczekiwana nie istnieje. W przypadku, gdy $g(x) \geq 0$ i nie istnieje wartość oczekiwana, to mówi się też, wartość oczekiwana jest nieskończona.

W pewnych przypadkach można łatwo stwierdzić, że wartość oczekiwana istnieje. Przykładem są tu dwa, łatwe do udowodnienia fakty.

Fakt 2.2.1. *Jeśli zmienna losowa X ma gęstość $f(x)$, która jest równa zero poza pewnym ograniczonym zbiorem na prostej rzeczywistej, (tzn. istnieje odcinek (a, b) taki, że jeśli $x \notin (a, b)$, to $f(x) = 0$) oraz jeśli $g(x)$ jest ograniczona na tym zbiorze, to istnieje $E g(X)$.*

Fakt 2.2.2. *Jeśli zmienna losowa dyskretna X przybiera tylko skończoną liczbę wartości x_i , to wartość oczekiwana $E g(X)$ istnieje.*

Poniżej przedstawione zostaną dwa najprostsze, ale bardzo ważne przykłady.

Przykład. Rozkład zero-jedynkowy: $\Pr(X = 1) = p$, $\Pr(X = 0) = q = 1 - p$, a więc zmienna losowa X nie przyjmuje innych wartości niż 0 i 1.

$$E X = 0 \cdot q + 1 \cdot p = p .$$

Przykład. Rozkład jednostajny na odcinku $[0, 1]$:

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

Wartość oczekiwana istnieje (patrz fakt 2.2.2) i wyraża się wzorem

$$E X = \int_0^1 x dx = \frac{1}{2} .$$

Dla wartości oczekiwanych prawdziwa jest własność taka jak dla całek, którą można słownie sformułować następująco:

- wartość oczekiwana sumy jest równa sumie wartości oczekiwanych, (o ile wartości oczekiwane istnieją),
- stałą można wyciągnąć przed znak wartości oczekiwanej.

Operator liniowy

Formalnie sformułujemy to w postaci twierdzenia.

Twierdzenie 2.2.1. *Jeżeli istnieją $E X$ i $E Y$, to istnieje $E(aX + bY)$ oraz*

$$E(aX + bY) = aE X + bE Y, \quad (2.2.4)$$

czyli wartość oczekiwana jest operatorem liniowym.

Wartość oczekiwana iloczynu

Podobnie jak dla całek, nie ma ogólnej reguły, jak obliczać wartość oczekiwaną iloczynu zmiennych losowych. Reguła taka istnieje tylko dla niezależnych zmiennych losowych.

Twierdzenie 2.2.2. *Jeżeli zmienne losowe X i Y są niezależne, to*

$$E(XY) = (E X)(E Y). \quad (2.2.5)$$

Wartość oczekiwana zmiennej losowej X , nazywa się momentem rzędu pierwszego i oznaczana jest bardzo często symbolem $m = m_1 = E X$.

2.2.3. Momenty wyższych rzędów

Moment zwykły

Momentem rzędu k , czyli momentem k -tego rzędu lub krótko, k -tym momentem zmiennej losowej X jest wartość oczekiwana jej k -tej potęgi, czyli $E X^k$. Moment rzędu k jest często oznaczany symbolem m_k . Moment ten nazywany jest momentem zwykłym, dla odróżnienia od innego rodzaju momentów omawianych dalej.

Przykład. Dla zmiennej losowej zero-jedynkowej

$$m_k = E X^k = 0^k \cdot p + 1^k \cdot q = p.$$

Przykład. Dla zmiennej losowej rozkładzie jednostajnym na $[0, 1]$

$$m_k = E X^k = \int_0^1 x^k dx = \frac{1}{k+1}.$$

Wariancja

Dalej definiuje się moment centralny rzędu k , (k -ty moment centralny), wzorem $c_k = E(X - E X)^k$. Przypadek $k = 2$ ma szczególną rolę w rachunku prawdopodobieństwa. Moment centralny drugiego rzędu nazywa się wariancją i jest oznaczany symbolem $D^2 X$,

$$D^2 X = E (X - E X)^2. \quad (2.2.6)$$

Wariancja jest nieujemna!

Ze definicji wariancji wynika w szczególności, że $D^2 X \geq 0$. Na oznaczenie wariancji używany jest też symbol $\text{Var } X$. Z wariancją związana jest ważna charakterystyka, zwana średnim odchyleniem standardowym lub dyspersją, oznaczona symbolem σ i określana jako pierwiastek z wariancji: $\sigma = \sqrt{D^2 X}$. Z istnienia momentów wyższych rzędów wynika istnienie momentów niższych rzędów.

Twierdzenie 2.2.3. *Jeżeli istnieje $E X^k$ oraz $l < k$, to istnieje $E X^l$.*

Dla wariancji nie jest prawdziwy bez dodatkowych założeń odpowiednik twierdzenia 2.2.1. Można jednak sformułować twierdzenie następujące.

Twierdzenie 2.2.4.

$$D^2(aX) = a^2 D^2 X, \quad (2.2.7)$$

a jeżeli zmienne losowe X i Y są niezależne, to

$$D^2(X + Y) = D^2 X + D^2 Y \quad (2.2.8)$$

lub oznaczając $\sigma_X = \sqrt{D^2 X}$ i $\sigma_Y = \sqrt{D^2 Y}$,

$$\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}.$$

Wariancja zmiennej losowej X , a zwłaszcza jej dyspersja, jest miarą odchylenia się zmiennej od jej średniej wartości, czyli miarą rozrzutu. Interpretacja taka wynika bezpośrednio ze wzoru (2.2.6). Wariancję łatwiej niż ze wzoru (2.2.6), można obliczyć korzystając z następującej własności.

Fakt 2.2.3.

$$D^2 X = E X^2 - (E X)^2 \quad (2.2.9)$$

albo równoważnie

$$\sigma^2 = m_2 - m_1^2.$$

Dowód. Ze wzoru (2.2.6) wynika, że

$$D^2 X = E (X^2 - 2XE X + (E X)^2) = E (X^2 - 2m_1 X + m_1^2).$$

Następnie z twierdzenia 2.2.1 wynika, że

$$D^2 X = E X^2 - 2m_1 E X + m_1^2 = m_2 - 2m_1^2 + m_1^2 = m_2 - m_1^2$$

□

Ponadto mamy następujący

Fakt 2.2.4. $D^2X = 0$ wtedy i tylko wtedy, gdy zmienna losowa X przyjmuje wartość stałą z prawdopodobieństwem 1, to znaczy, gdy $\Pr(X = x_0) = 1$ dla pewnej wartości x_0 .

Dowód. Jeżeli $\Pr(X = x_0) = 1$, to $m_1 = x_0$ oraz $m_2 = x_0$, a więc $D^2X = 0$. Z drugiej strony, jeśli $D^2X \neq 0$, to oczywiście $D^2X > 0$. Przypuśćmy, że $D^2X = 0$, a równocześnie $\Pr(X = x_0) < 1$. Ze wzoru (2.2.6) otrzymujemy

$$\sigma^2 = D^2X = \int_{-\infty}^{\infty} (x - m)^2 dF(x) \geq 0,$$

gdyż $(x - m)^2 \geq 0$, przy czym $\sigma^2 = 0$ tylko wtedy, gdy funkcja podcałkowa jest tożsamościowo równa zeru, (co nie jest prawdą) albo gdy w dokładnie jednym punkcie $x_0 = m$ dystrybuanta $F(x)$ ma skok równy 1, czyli $F(x_0^+) - F(x_0) = 1$. \square

2.2.4. Zadania

1. Niech R będzie zmienną losową o rozkładzie jednostajnym na odcinku $[0, 1]$. Obliczyć momenty zwykłe i centralne k -tego rzędu.

2. Zmienna losowa X ma rozkład dany tabelą:

x_k	0	1	2	3	4
p_k	0.2	0.1	0.1	0.3	A

Znaleźć stałą A , a następnie wartość oczekiwaną oraz wariancję zmiennych X i $Y = \sin(\pi X/2)$.

3. Rzucamy trzy razy monetą. Niech X będzie liczbą wyrzuconych orłów. Obliczyć EX i D^2X .

4. Zmienna losowa X przyjmuje tylko wartości 2^k dla $k = 1, 2, \dots$ oraz $\Pr(X = 2^k) = 4 \cdot 5^{-k}$. Znaleźć EX oraz D^2X . Czy istnieje trzeci moment zmiennej X ?

5. Zmienna losowa X przyjmuje tylko wartości 3^{k-1} dla $k = 1, 2, \dots$ oraz $\Pr(X = 3^{k-1}) = 2 \cdot 3^{-k}$. Czy istnieje EX ? Obliczyć $E\sqrt{X}$.

6. Rzucamy dwoma kostkami. Obliczyć wartość oczekiwaną i wariancję sumy oczek.

2.3. Rozkłady dyskretne

2.3.1. Rozkład dwupunktowy i dwumianowy

Zmienna losowa X ma rozkład dwupunktowy, gdy z prawdopodobieństwem 1 przyjmuje tylko dwie wartości, tzn. jeśli $\Pr(X = x_1) = p$ i $\Pr(X = x_2) = q$, to $p + q = 1$. Łatwo policzyć momenty: $E X = x_1 p + x_2 q$, co w przypadku $p = q = 1/2$ daje $m = \frac{x_1 + x_2}{2}$, czyli średnią arytmetyczną. Moment zwykły rzędu k wynosi $m_k = x_1^k p + x_2^k q$.

Szczególnym przypadkiem rozkładu dwupunktowego jest rozkład zero-jedynkowy, rozważany już w punkcie 2.2.2. Zajmiemy się tym przypadkiem dokładnie, bo mimo swej prostoty, (a może właśnie dzięki swej prostocie), jest on podstawą wielu dalszych rozważań.

Rozkład zero-jedynkowy

Niech $X_i, i = 1, 2, \dots, n$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie zero-jedynkowym, $\Pr(X_i = 1) = p$, $\Pr(X_i = 0) = q = 1 - p$. Oznaczmy przez X sumę n takich zmiennych losowych:

$$X = X_1 + X_2 + \dots + X_n. \quad (2.3.1)$$

Z twierdzeń 2.2.1 i 2.2.4 otrzymujemy równości: $E X = np$ i $D^2 X = npq$, skąd też $\sigma = \sigma_X = \sqrt{npq}$.

Zbadajmy rozkład zmiennej losowej X , to znaczy obliczymy prawdopodobieństwa $\Pr(X = k) = p_k$. Przede wszystkim zauważmy, że k może zmieniać się tylko od 0 do n . Ponieważ zdarzenie $\{X = k\}$ zachodzi wtedy i tylko wtedy, gdy na dokładnie k pozycjach w sumie pojawi się jedynka – prawdopodobieństwo tego wynosi $p^k q^{n-k}$, a możliwych takich układów k jedynek w ciągu n -elementowym jest $\binom{n}{k}$, to

Rozkład dwumianowy

$$\Pr(X = k) = \binom{n}{k} p^k q^{n-k}. \quad (2.3.2)$$

Wzór (2.3.2) określa rozkład, znany jako rozkład dwumianowy (głównie w literaturze anglosaskiej) lub rozkład Bernoulliego⁶. Zmienną losową X określoną wzorem (2.3.1) można interpretować jako liczbę sukcesów (jedynek) w ciągu niezależnych doświadczeń, zwanych próbami Bernoulliego, gdzie prawdopodobieństwo sukcesu wynosi p , a porażki (zera) wynosi $q = 1 - p$.

Przykład. Wykonujemy ciąg $n = 10$ doświadczeń Bernoulliego. W każdym z nich odnosimy sukces z prawdopodobieństwem $p = 0.9$. Jakie jest prawdopodobieństwo, że w dokładnie jednej próbie lub w nie więcej niż 10% prób odniesiemy sukces. W tym przypadku rozwiązanie jest proste. Zarówno w jednym jak i w drugim przypadku obliczamy

$$\Pr(X = 1) = \binom{10}{1} 0.9^1 0.1^9 = 9 \cdot 10^{-9}.$$

⁶Jakub Bernoulli (1654 – 1705), matematyk szwajcarski, jeden z licznej rodziny Bernoullich, autor *Ars conjectandi*, pierwszego dzieła poświęconego rachunkowi prawdopodobieństwa.

Co jednak będzie, gdy przy tym samym n będzie $p = 0.09$? Obliczenie liczby $0.09 \cdot 0.91^9$ jest już kłopotliwe, a obliczenie prawdopodobieństwa, że uda się nie więcej niż 10% doświadczeń z 1000 prób jest już praktycznie bardzo trudne. Spróbujmy bowiem bezpośrednio i dokładnie obliczyć

$$\Pr(X \leq 100) = \sum_{k=0}^{100} \binom{1000}{k} p^k q^{1000-k}.$$

Napisanie procedur (np. w Pascalu lub C) wykorzystujących wprost wzory (2.3.2) i obliczający współczynniki Newtona ze wzoru rekurencyjnego

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

nie daje zadowalających efektów dla większych n i k .

Wzór
rekurencyjny

Dla wygody przyjmiemy oznaczenie

$$b(n, k, p) = \binom{n}{k} p^k q^{n-k}.$$

Wtedy łatwo jest udowodnić wzór rekurencyjny

$$b(n, 0, p) = q^n, \quad (2.3.3)$$

$$b(n, k+1, p) = \frac{n-k}{k+1} \frac{p}{q} b(n, k, p). \quad (2.3.4)$$

Wzór ten do bezpośredniego zaprogramowania nadaje się w równie małym stopniu jak i wzór (2.3.2), ale stanowić będzie podstawę dla efektywnych przybliżeń rozpatrywanych w następnym punkcie.

2.3.2. Rozkład Poissona

Zmienna losowa X przyjmująca tylko wartości całkowite nieujemne ma rozkład Poissona⁷ z parametrem λ , gdy

$$\Pr(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}. \quad (2.3.5)$$

Definicja tego rozkładu jest poprawna, to znaczy wszystkie prawdopodobieństwa są nieujemne i sumują się do jedynki, gdyż jak wiadomo z rozwinięcia funkcji e^x w szereg Maclaurina,

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Podobnie jak dla rozkładu dwumianowego, również dla rozkładu Poissona przyjmiemy oznaczenie

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

Przybliżenie
rozkładem
Poissona

Twierdzenie 2.3.1. (Poissona) Jeżeli $np = np_n \rightarrow \lambda$ dla $n \rightarrow \infty$, to dla każdego k ,

$$\lim_{n \rightarrow \infty} b(n, k, p) = p(\lambda, k). \quad (2.3.6)$$

Dowód. Dla uproszczenia przyjmijmy $np = \lambda$, zamiast granicy $np \rightarrow \lambda$. Dowód przeprowadzimy przez indukcję.

1°

$$b\left(n, 0, \frac{\lambda}{n}\right) = \left(\frac{\lambda}{n}\right)^0 \left(1 - \frac{\lambda}{n}\right)^n \rightarrow e^{-\lambda},$$

2°

$$\frac{b\left(n, k+1, \frac{\lambda}{n}\right)}{b\left(n, k, \frac{\lambda}{n}\right)} = \frac{\binom{n}{k+1} p}{\binom{n}{k} q} = \frac{n-k}{k+1} \frac{\frac{\lambda}{n}}{1 - \frac{\lambda}{n}} \rightarrow \frac{\lambda}{k+1} = \frac{p(k+1, \lambda)}{p(k, \lambda)},$$

co daje wzór (2.3.5) □

Przybliżenie
dla średnich λ

Uwaga. Twierdzenie 2.3.1 daje dobre przybliżenie rozkładu dwumianowego rozkładem Poissona, gdy n jest duże, ($n \geq 100$), p jest małe, a λ średnie. Wtedy zamiast (2.3.6) można przyjąć

$$b(n, k, p) \approx p(\lambda, k). \quad (2.3.7)$$

Przykład. W $n = 100$ próbach Bernoulliego prawdopodobieństwo sukcesu wynosi $p = 0.98$. Niech X będzie liczbą porażek. Prawdopodobieństwo porażki wynosi $q = 0.02$. Szukamy prawdopodobieństwa, że poniesiemy co najwyżej jedną porażkę. Ponieważ $\lambda = np = 2$, to można zastosować przybliżenie wzorem (2.3.7):

$$\Pr(X \leq 1) = \Pr(X = 0) + \Pr(X = 1) \approx e^{-2} \left(\frac{2^0}{0!} + \frac{2^1}{1!} \right) = 3e^{-2}.$$

Przykład. Prawdopodobieństwo przesłania błędnego bitu wynosi $2.5 \cdot 10^{-9}$ niezależnie od pozostałych. Przesyłamy 10^8 bitów i na końcu dodajemy bit parzystości. Błędny ciąg odbierzemy jako prawdziwy, gdy przekłamaniu ulegnie parzysta liczba bitów. Prawdopodobieństwo odebrania błędnego ciągu jako prawdziwego, wynosi w przybliżeniu $\lambda^2 e^{-\lambda}/2$, gdzie $\lambda = 0.25$, gdyż prawdopodobieństwo przekłamania czterech bitów wynosi $\lambda^4 e^{-\lambda}/24$ i może być pominięte, tak jak i dalsze parzyste liczby przekłamań. Tak więc prawdopodobieństwo odebrania błędnego ciągu jako prawdziwego wynosi w przybliżeniu $\lambda^2 e^{-\lambda}/2! = 2.5^2 e^{-0.25}/2 = 0.2565$.

Na koniec obliczmy parametry w rozkładzie Poissona – wartość oczekiwaną i wariancję.

Twierdzenie 2.3.2. Jeżeli zmienna losowa X ma rozkład Poissona z parametrem λ , to $E X = \lambda$ i $D^2 X = \lambda$.

⁷Siméon Denis Poisson (1781 – 1840), francuski mechanik, fizyk i matematyk

Dowód. Korzystając z twierdzenia 2.3.1 można zauważyć, że przyjmując $\lambda \rightarrow np$ otrzymujemy

$$E X = \lim_{n \rightarrow \infty} np = \lambda$$

oraz

$$D^2 X = \lim_{n \rightarrow \infty} np(1-p) = \lim_{n \rightarrow \infty} \lambda \left(1 - \frac{\lambda}{n}\right).$$

□

2.3.3. Zadania

1. Prawdopodobieństwo sukcesu w jednej próbie wynosi 0.01. Niech X_n będzie liczbą sukcesów w n niezależnych próbach. Znaleźć $\Pr(X \leq 1)$, gdy a) $n = 3$, b) $n = 100$.
2. Książka o 500 stronicach zawiera 50 błędów drukarskich. Obliczyć prawdopodobieństwo tego, że na przypadkowo wybranej stronicy znajdą się co najmniej 3 błędy.
- 3*. Pokazać, że jeśli X ma rozkład dwumianowy przy ustalonym n i p , to $b(n, k, p)$ ma tylko jedną lub dwie równe wartości największe. Znaleźć takie k przy których te wartości są przyjmowane.
- 4*. Pokazać, że analogiczną własność mają prawdopodobieństwa w rozkładzie Poissona.
5. Bity przekazywane szeregowo są przekłamywane niezależnie od siebie, każdy z prawdopodobieństwem $p = 10^{-6}$. Na końcu każdego bloku bitów dodajemy bit parzystości. Jaka jest największa długość bloku, przy której prawdopodobieństwo niewykrycia błędu jest mniejsze od 0.001?

2.4. Rozkłady ciągłe

2.4.1. Rozkład jednostajny

Rozkład jednostajny na odcinku $[a, b]$ ma gęstość określoną wzorem

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{dla } x \in [a, b], \\ 0 & \text{dla } x \notin [a, b]. \end{cases}$$

Momenty zwykłe

$$m_k = \frac{1}{b-a} \int_a^b x^k dx = \frac{1}{b-a} \frac{x^{k+1}}{k+1} \Big|_a^b = \frac{b^{k+1} - a^{k+1}}{(b-a)(k+1)}$$

mają dla $a = 0$ i $b = 1$ szczególnie prostą postać (patrz zadanie 1 do rozdziału 2.2): $m_k = 1/(k+1)$. Stąd obliczamy wariancję

$$\sigma^2 = m_2 - m_1^2 = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{(b-a)^2}{12}.$$

Dla rozkładu jednostajnego momenty centralne nieparzystego rzędu są równe zeru. Łatwo jest też obliczyć momenty rzędu parzystego, co jednak pozostawimy jako zadania 6 i 7 do tego rozdziału. Zauważmy tu jednak, (co będzie pomocne przy rozwiązywaniu tych zadań), że jeśli zmienna losowa X ma rozkład jednostajny na odcinku $[0, 1]$, to zmienna losowa $Y = (b-a)X + a$ ma rozkład jednostajny na odcinku $[a, b]$.

2.4.2. Rozkład wykładniczy

Rozkład wykładniczy z parametrem $\lambda > 0$ ma gęstość określoną wzorem

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{dla } x \geq 0, \\ 0 & \text{dla } x < 0, \end{cases} \quad (2.4.1)$$

której wykres jest przedstawiony na rysunku 4. Ze wzoru (2.4.1) otrzymujemy wzór na dystrybuantę

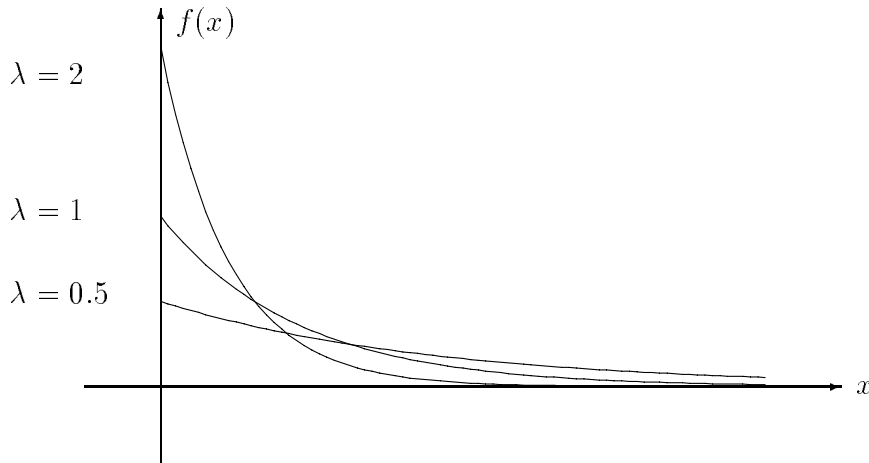
$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{dla } x \geq 0, \\ 0 & \text{dla } x < 0. \end{cases} \quad (2.4.2)$$

Obliczmy dwa pierwsze momenty i wariancję.

$$m_1 = \lambda \int_0^{\infty} x e^{-\lambda x} dx,$$

Wartość
oczekiwana

skąd obliczając całkę przez części, $(\lambda e^{\lambda x} = -(e^{-\lambda x})')$, otrzymujemy



Rysunek 4: Wykres gęstości rozkładu wykładniczego

$$m_1 = \lambda \left(-\frac{x}{\lambda} e^{-\lambda x} \Big|_0^\infty + \frac{1}{\lambda} \int_0^\infty dx \right) = -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^\infty = \frac{1}{\lambda}.$$

Podobnie, ale dwukrotnie całkując przez części, obliczamy

$$m_2 = \lambda \int_0^\infty x^2 e^{-\lambda x} dx = \frac{2}{\lambda^2},$$

Wariancja skąd

$$\sigma^2 = \frac{1}{\lambda^2}.$$

Rozkład wykładniczy jest często używany w teorii niezawodności, gdzie dobrze opisuje czas pracy elementów niestarczających się. Jeżeli zmienna losowa T jest czasem pracy elementu, to $\tau = 1/\lambda$ jest średnim czasem pracy, a parametr $\lambda = 1/\tau$ jest nazywany intensywnością uszkodzeń. Niestarzenie się elementu oznacza, że prawdopodobieństwo uszkodzenia się elementu w czasie t od chwili obecnej, nie zależy od dotychczas przepracowanego czasu t_0 . Własność ta nazywana jest brakiem pamięci rozkładu wykładniczego. Sformułujemy ją w postaci twierdzenia.

Twierdzenie 2.4.1. Jeżeli T jest zmienną losową o rozkładzie wykładniczym, to

$$\Pr(T > t + t_0 | T > t_0) = \Pr(T > t). \quad (2.4.3)$$

Dowód. Ponieważ $\{T > t + t_0\} \subset \{T > t\}$ oraz $\Pr(T > t) = 1 - F(t) = e^{-t/\tau}$, to

$$\begin{aligned} \Pr(T > t + t_0 | T > t_0) &= \frac{\Pr(T > t + t_0)}{\Pr(T > t_0)} \\ &= e^{(t+t_0)/\tau} e^{-t_0/\tau} = e^{-t/\tau} = \Pr(T > t). \end{aligned}$$

□

*Rozkład
Weibulla*

Można też pokazać, czego jednak nie będziemy robić, że tylko rozkład wykładniczy ma własność braku pamięci, to znaczy spełnia równanie (2.4.3).

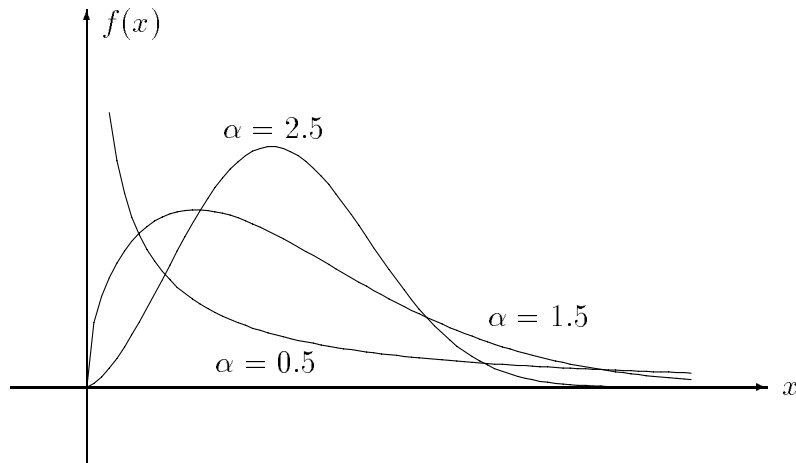
Uogólnieniem rozkładu wykładniczego jest rozkład Weibulla, którego dystrybuanta jest modyfikacją dystrybuanty danej równaniem (2.4.2)

$$F(x) = \begin{cases} 1 - e^{-\lambda x^\alpha} & \text{dla } x \geq 0, \\ 0 & \text{dla } x < 0. \end{cases} \quad (2.4.4)$$

Gęstość tego rozkładu wyraża się wzorem

$$f(x) = \begin{cases} \lambda x^{\alpha-1} e^{-\lambda x^\alpha} & \text{dla } x \geq 0, \\ 0 & \text{dla } x < 0, \end{cases} \quad (2.4.5)$$

której wykres dla $\lambda = 1$ i $\alpha = 0.5, \alpha = 1.5, \alpha = 2.5$ jest przedstawiony na rysunku 5. Przypadek $\alpha = 1$ daje rozkład wykładniczy. Rozkład Weibulla jest



Rysunek 5: Wykres gęstości rozkładu Weibulla

stosowany w teorii niezawodności.

2.4.3. Rozkład normalny

Rozkład normalny $N(0, 1)$ ma gęstość daną wzorem

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad (2.4.6)$$

której wykres jest przedstawiony na rysunku 6. Jeśli zmienna losowa Y ma rozkład normalny $N(0, 1)$, to zmienna losowa

*Rozkład
normalny
 $N(m, \sigma)$*

$$X = \sigma Y + m \quad (2.4.7)$$

ma rozkład normalny $N(m, \sigma)$. W ten sposób definiuje się rozkład normalny o parametrach m i σ . Z drugiej strony, jeśli X ma rozkład normalny $N(m, \sigma)$, co oznaczamy symbolicznie jako $X \sim N(m, \sigma)$, to jak łatwo sprawdzić

$$\tilde{X} = \frac{X - m}{\sigma} \sim N(0, 1). \quad (2.4.8)$$

Operację daną równaniem (2.4.8), (odwrotną do (2.4.7)), nazywa się standaryzacją zmiennej losowej X .

Ponieważ gęstość zmiennej losowej $Y = aX + b$, $a > 0$, gdy X ma gęstość $f(x)$, wyraża się wzorem (patrz zadanie 9),

$$f(x)_Y = \frac{1}{a} f_X \left(\frac{x - b}{a} \right),$$

Gęstość
rozkładu
 $N(m, \sigma)$

to gęstość rozkładu normalnego $N(m, \sigma)$ wyraża się wzorem

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}. \quad (2.4.9)$$

Wzór (2.4.9) jest bezpośrednim wnioskiem ze wzorów (2.4.7) i (2.4.6). Ponieważ gęstość $f(x)$ rozkładu $N(0, 1)$ jest funkcją parzystą, to $xf(x)$ jest funkcją nieparzystą, a więc dla $X \sim N(m, \sigma)$ otrzymujemy $E\tilde{X} = 0$, skąd ze wzorów (2.4.7) i (2.4.8): $EX = m$. Nieco więcej zachodu wymaga obliczenie D^2X . Wystarczy jednak obliczyć, że $D^2\tilde{X} = 1$, (zadanie 10), skąd od razu, ze wzoru (2.4.7) otrzymujemy, że $D^2X = \sigma^2$.

Dystrybuantę zmiennej losowej $X \sim N(0, 1)$ oznacza się tradycyjnie symbolem $\Phi(x)$. Dystrybuanta rozkładu normalnego nie jest funkcją elementarną, dlatego jej wartości trzeba odczytywać z tablic rozkładu normalnego lub korzystać z programów komputerowych. Bez żadnych obliczeń mamy jednak $\Phi(0) = 0.5$.

Ze względu na (2.4.7), wartości dystrybuanty rozkładu $N(m, \sigma)$ można otrzymać z wartości dystrybuanty rozkładu $N(0, 1)$. Czasem w tablicach zamiast wartości dystrybuanty $\Phi(x)$ podaje się wartości funkcji

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt = \Phi(x) - 0.5.$$

Dla $x > 0$ mamy $\Phi(x) = 0.5 + \Phi_0(x)$.

*Reguła
3-sigmowa*

Dla rozkładu normalnego można z tablic odczytać następującą własność, znaną jako reguła trzy-sigmowa.

Fakt 2.4.1. *Jeżeli zmienna losowa X ma rozkład normalny $N(m, \sigma)$, to*

$$\Pr(|X - m| > 3\sigma) < 0.01. \quad (2.4.10)$$

Dowód. Ponieważ $\Pr(\tilde{X} > x) = 1 - \Phi(x) = \Phi(-x)$, to

$$\begin{aligned} \Pr(|X - m| > 3\sigma) &= \Pr\left(\left|\frac{X - m}{\sigma}\right| > 3\right) = 2(1 - \Phi(3)) \\ &= 2 \cdot 0.0044 = 0.0088 < 0.01. \end{aligned}$$

□

Dla rozkładów innych niż normalny, nierówność (2.4.10) nie musi być prawdziwa.

Przykład. Dla zmiennej losowej X o gęstości

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|},$$

jest $EX = 0$, gdyż $f(x)$ jest funkcją parzystą, a całka

$$\int_{-\infty}^{\infty} xf(x) dx = 2 \int_0^{\infty} xf(x) dx$$

jest bezwzględnie zbieżna (i równa $1/\lambda$, będąc wartością oczekiwaną w rozkładzie wykładniczym). Wariancja

$$D^2 X = \frac{\lambda}{2} \int_{-\infty}^{\infty} x^2 e^{-\lambda|x|} dx = \lambda \int_0^{\infty} x^2 e^{-\lambda x} dx = \frac{2}{\lambda}$$

jest drugim momentem zwykłym rozkładu wykładniczego, skąd $\sigma = \sqrt{2}/\lambda$. Obliczamy

$$\Pr(|X - m| > 3\sigma) = 2 \Pr\left(X > \frac{3\sqrt{2}}{\lambda}\right) = e^{-3/\sqrt{2}} \approx 0.1199,$$

a więc nierówność (2.4.10) jest dla tego rozkładu nieprawdziwa.

2.4.4. Zadania

1. Niech R będzie zmienną losową o rozkładzie jednostajnym na odcinku $[0, 1]$, tzn. $f(x) = 1$ dla $x \in [0, 1]$ oraz $f(x) = 0$ dla $x \notin [0, 1]$. Niech $F(x)$ będzie ciągłą i różnowartościową dystrybuantą pewnej zmiennej losowej. Pokazać, że $X = F^{-1}(R)$ ma dystrybuantę $F(x)$.
2. W oparciu o zadanie 1 podać algorytm generowania zmiennej losowej o rozkładzie wykładniczym z parametrem λ , mając do dyspozycji generator liczb losowych o rozkładzie jednostajnym na odcinku $[0, 1]$.
3. Korzystając z tablic rozkładu normalnego obliczyć $\Pr(X < 0.5)$, $\Pr(X > -1)$, $\Pr(1.5 < X \leq 3)$ dla zmiennych losowych $X \sim N(1, 1)$, $X \sim N(-1, 0.5)$, $X \sim N(1.5, 1)$ i $X \sim N(1.5, 0.1)$.
4. Korzystając z tablic rozkładu normalnego obliczyć parametr x_0 spełniający równanie $\Pr(X > x_0) = \alpha$ dla wartości $\alpha = 0.1$ oraz wartości $\alpha = 0.05$ dla zmiennej losowej $X \sim N(0, 1)$ oraz $X \sim N(-0.5, 2)$.
5. Przez analogię do reguły 3-sigmowej sformułować „regułę n -sigmową” jeśli termin „bardzo mały” chcemy rozumieć jako „mniejszy od 0.0001”. Jak należałoby rozumieć termin „bardzo mały” chcąc mieć „regułę 2-sigmową”?
6. Pokazać, że w rozkładzie jednostajnym momenty centralne rzędu nieparzystego są równe zeru.
7. Obliczyć momenty centralne parzystego rzędu dla zmiennych losowych o rozkładzie jednostajnym na odcinku $[a, b]$.
- 8**. Niech X i S będą niezależnymi zmiennymi losowymi takim, że X będzie miała rozkład wykładniczy oraz $\Pr(S = 1) = \Pr(S = -1) = 1/2$. Znaleźć rozkład zmiennej losowej $Z = SX$.
9. Niech $Y = aX + b$, $a > 0$ oraz niech X ma dystrybuantę $F(x)$ i gęstość $f(x)$. Znaleźć dystrybuantę i gęstość zmiennej losowej X .
10. Niech X ma rozkład normalny $N(0, 1)$. Obliczyć $D^2 X$.
- 11*. Korzystając z rozwinięcia gęstości rozkładu normalnego w szereg Maclaurina, znaleźć rozwinięcie dystrybuanty. Korzystając z niego napisać procedurę obliczania wartości dystrybuanty rozkładu $N(m, \sigma)$.

3. Twierdzenia graniczne

3.1. Nierówność Czebyszewa i praw

Przykład. Niech X ma dowolny rozkład spełniający założenia twierdzenia 3.1.2. Wtedy z (3.1.2) mamy

$$\Pr(|X - m| \geq 3\sigma) = \Pr\left(\left|\frac{X - m}{\sigma}\right| \geq 3\right) \leq \frac{1}{9}.$$

Jeżeli zaś $X \sim N(m, \sigma)$, (X spełnia również założenia twierdzenia 3.1.2), to z (2.4.10) wynika oszacowania lepsze o rząd, to znaczy ponad dziesięciokrotnie lepsze.

Przykład. Niech X będzie liczbą sukcesów w 20 próbach, gdzie prawdopodobieństwo sukcesu w jednej próbie wynosi $p = 0.4$. Oszacujmy z dołu prawdopodobieństwo, że liczba sukcesów będzie zawarta między 4 i 12.

Ponieważ $m = E X = 20 \cdot 0.4 = 8$ oraz $\sigma^2 = D^2 X = 20 \cdot 0.4 \cdot 0.6 = 4.8$ to

$$\begin{aligned} \Pr(4 \leq X \leq 12) &= \Pr(|X - m| \leq 4) = 1 - \Pr(|X - m| > 4) \\ &\geq 1 - \Pr(|X - m| \geq 2) \geq 1 - \frac{\sqrt{4.8}}{16} = 0.7, \end{aligned}$$

przyjmując $t = 4/\sqrt{4.8}$ we wzorze (3.1.2).

3.1.2. Prawa wielkich liczb

Z nierówności Czebyszewa można wyprowadzić słabe prawo wielkich liczb dla zmiennych losowych. Takie sformułowanie prawa wielkich liczb jest ogólniejsze, niż sformułowanie tylko dla zdarzeń, jak to zrobiono w punkcie 1.2.1.

Słabe prawo wielkich liczb

Twierdzenie 3.1.3. Niech X_1, X_2, \dots będą ciągiem niezależnych zmiennych losowych o tym samym rozkładzie, o wartości oczekiwanej m i wariancji $\sigma^2 < \infty$. Wtedy dla każdego $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \Pr\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - m\right| < \varepsilon\right) = 1. \quad (3.1.3)$$

Dowód. Dla skrócenia zapisu oznaczmy

$$S_n = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Z niezależności zmiennych losowych X_i wynika, że $D^2 S_n = \sigma^2/n$, a także $E S_n = m$. Z nierówności Czebyszewa (3.1.2) można przyjmując $\varepsilon = t\sigma$ napisać

$$\Pr(|S_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2},$$

co daje tezę twierdzenia dla $n \rightarrow \infty$. □

1. The first step in the process of creating a new product is to identify a market need.

2. The second step is to develop a prototype of the product.

3. The third step is to conduct market research to determine if there is a demand for the product.

4. The fourth step is to create a business plan for the product.

5. The fifth step is to secure funding for the product.

6. The sixth step is to manufacture the product.

7. The seventh step is to distribute the product.

3.2. Funkcje charakterystyczne

3.2.1. Definicje i własności

*Zmienna
losowa
zespolona*

Zmienną losową zespoloną nazywamy funkcję $Z : \Omega \rightarrow \mathbb{C}$ taką, że

$$Z(\omega) = X(\omega) + iY(\omega),$$

gdzie $X(\omega)$ i $Y(\omega)$ są rzeczywistymi zmiennymi losowymi zdefiniowanym w rozdziale 2.1.1.

*Funkcja cha-
rakterystyczna*

Funkcją charakterystyczną nazywamy funkcję zmiennej rzeczywistej $\varphi(t)$ określoną wzorem

$$\varphi(t) = E e^{itX} = \int_{-\infty}^{\infty} e^{itx} dF(x). \quad (3.2.1)$$

Funkcja charakterystyczna ma następujące własności, które w wielu przypadkach pomagają obliczać i badać różne charakterystyki zmiennych losowych i ich rozkładów.

Twierdzenie 3.2.1. *Funkcja charakterystyczna $\varphi(t)$ określona wzorem (3.2.1)*

- (i) *jest jednostajnie ciągła,*
- (ii) $\varphi(0) = 1,$
- (iii) $|\varphi(t)| \leq 1,$
- (iv) $\varphi(-t) = \overline{\varphi(t)}.$

□

*Niezależne
zmiennne
losowe*

Poniższe twierdzenie pomaga obliczać rozkłady i funkcje charakterystyczne sum zmiennych losowych.

Twierdzenie 3.2.2. *Jeżeli zmienne losowe X i Y są niezależne, to*

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t),$$

gdzie $\varphi_X(t)$ i $\varphi_Y(t)$ są funkcjami charakterystycznymi zmiennych losowych X i Y . □

Momenty

Również obliczanie momentów zmiennych losowych może być znacznie prostsze, gdy znamy ich funkcje charakterystyczne.

Twierdzenie 3.2.3. *Istnienie k -tego momentu zmiennej losowej jest równoważne istnieniu k -tej pochodnej jej funkcji charakterystycznej, przy czym*

$$E X^k = \frac{\varphi^{(k)}(0)}{i^k}. \quad \square$$

Bezpośrednio ze wzoru (3.2.1) wynika wzór, analogiczny do znanych wzorów w przekształceniu Laplace'a:

$$\varphi_{aX}(t) = \varphi_X(at).$$

Z kolei z twierdzenia 3.2.2 wynika, że

$$\varphi_{X+a}(t) = e^{ita} \varphi_X(t),$$

gdyż stałą a można traktować jako zmienną losową przyjmującą tylko jedną wartość a , a taka zmienna losowa jest niezależna od dowolnych zmiennych losowych i ma funkcję charakterystyczną (co łatwo sprawdzić) równą e^{ita} .

Omówimy teraz funkcje charakterystyczne rozkładów typu ciągłego, które były przedstawione w rozdziale 2.4. Obliczenie funkcji charakterystycznej rozkładu jednostajnego pozostawimy jako zadania 1 i 3, policzymy natomiast funkcje charakterystyczne rozkładu wykładniczego (strona 29) i normalnego (strona 31).

*Rozkład
wykładniczy*

Przykład. Dla rozkładu wykładniczego z parametrem λ funkcję charakterystyczną obliczamy ze wzoru (3.2.1)

$$\varphi(t) = \int_0^{\infty} e^{itx} \lambda e^{-\lambda x} dx = \frac{1}{1 - \frac{it}{\lambda}}. \quad (3.2.2)$$

*Rozkład
normalny*

Przykład. Dla rozkładu normalnego:

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-it)^2/2} e^{-t^2/2} dx = e^{-t^2/2}.$$

3.2.2. Rozkład gamma

*Rozkład
gamma*

Rozkład gamma definiujemy jako rozkład, którego funkcja charakterystyczna ma postać

$$\varphi(t) = \frac{1}{\left(1 - \frac{it}{b}\right)^p}, \quad (3.2.3)$$

gdzie $b > 0$ i $p > 0$.

*Rozkład
Erlanga*

Z porównania wzorów (3.2.2) i (3.2.3) wynika, że rozkład wykładniczy jest szczególnym przykładem rozkładu gamma. Ponadto z Twierdzenia 3.2.2 otrzymujemy, że jeśli $p = n$ jest liczbą naturalną, to zmienna losowa o takim rozkładzie gamma jest sumą n zmiennych losowych niezależnych, o tym samym rozkładzie wykładniczym. Taki szczególny przypadek rozkładu gamma nazywa się rozkładem Erlanga o n stopniach swobody.

Dla rozkładu gamma można wyznaczyć gęstość. Wyraża się ona wzorem

$$f(x) = \begin{cases} \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} & \text{dla } x \geq 0, \\ 0 & \text{dla } x < 0, \end{cases}$$

gdzie

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx$$

jest funkcją specjalną zwaną funkcją Gamma. Funkcja ta ma ciekawą własność, $\Gamma(p+1) = p\Gamma(p)$, która pozwala ograniczyć się jedynie do znajomości jej wartości dla $0 < p \leq 1$. Proste obliczenia pokazują też, że $\Gamma(1) = 1$, skąd łatwo już otrzymać równość: $\Gamma(n) = (n-1)!$.

3.2.3. Zadania

1. Obliczyć funkcję charakterystyczną rozkładu jednostajnego na odcinku $[0, 1]$.
2. Obliczyć funkcję charakterystyczną rozkładu o gęstości $f(x) = \lambda e^{-\lambda|x|}/2$.
3. Korzystając z wyniku zadania 1 i własności funkcji charakterystycznych, znaleźć funkcję charakterystyczną rozkładu jednostajnego na odcinku $[a, b]$.
4. Korzystając z funkcji charakterystycznej rozkładu $N(0, 1)$, obliczyć momenty k -tego rzędu tego rozkładu.
- 5*. Pokazać, że

$$\int_0^{\infty} \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} dx = 1$$

dla $b > 0$ i $p > 0$.

6. Korzystając z funkcji charakterystycznej rozkładu gamma obliczyć wartość oczekiwaną i wariancję w tym rozkładzie.
7. Niech X i Y będą niezależne o rozkładzie jednostajnym na odcinku $[-0.5, 0.5]$. Udowodnić, że ich suma ma rozkład o gęstości

$$f(x) = \begin{cases} x+1 & \text{dla } x \in [-1, 0), \\ -x+1 & \text{dla } x \in [0, 1], \\ 0 & \text{dla } x \notin [-1, 1]. \end{cases}$$

3.3. Centralne twierdzenie graniczne

3.3.1. Twierdzenie Lindeberga-Levy'ego

Dystrybuantę rozkładu normalnego $N(0, 1)$ oznaczamy jak zwykle jako $\Phi(x)$.

$D^2X < \infty \equiv$
istnieje
wariancja

Twierdzenie 3.3.1. (Lindeberga-Levy'ego) Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi o tym samym rozkładzie, wartości oczekiwanej $m = E X$ i wariancji $0 < \sigma^2 = D^2X < \infty$. Wtedy

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{X_1 + X_2 + \dots + X_n - nm}{\sigma \sqrt{n}} < x \right) = \Phi(x). \quad (3.3.1)$$

Dowód. Funkcja charakterystyczna rozkładu normalnego $N(0, 1)$ jest równa $\varphi(t) = e^{-t^2/2}$. Oznaczmy przez $\varphi_X(t)$ funkcję charakterystyczną zmiennej losowej $X'_i = X_i - m$, a przez $\varphi_{Z_n}(t)$ funkcję charakterystyczną zmiennej losowej

$$Z_n = \frac{X'_1 + X'_2 + \dots + X'_n}{\sigma \sqrt{n}}$$

występującej po lewej stronie wzoru (3.3.1). Uwzględniając znane już własności (paragraf 3.2.1) funkcji charakterystycznej otrzymujemy, że

$$\varphi_{Z_n}(t) = \left(\varphi_X \left(\frac{t}{\sigma \sqrt{n}} \right) \right)^n. \quad (3.3.2)$$

Rozwińmy teraz funkcję $\varphi_X(t)$ w szereg Maclaurina. Ponieważ $\varphi_X(0) = 1$, $\varphi'_X(0) = m$ oraz $\varphi''_X(0) = -\sigma^2$, to

$$\varphi_X(t) = 1 - \frac{1}{2}\sigma^2 t^2 + R_3(t^2),$$

gdzie $R_3(t^2)/t^2 \rightarrow 0$ dla $t \rightarrow 0$. Ze wzoru (3.3.2) otrzymujemy z kolei

$$\varphi_{Z_n}(t) = \left(1 - \frac{t^2}{2n} + R_3(t^2/n) \right)^n$$

oraz

$$\lim_{n \rightarrow \infty} n R_3(t^2/n) = 0.$$

Podstawiając $u = t^2/2n + R_3(t^2/n)$ otrzymujemy

$$\ln \varphi_{Z_n}(t) = n \ln(1 - u) = n \left(-t^2/2n + R_3(t^2/n) \right) = -t^2/2 + n R_3(t^2/n),$$

skąd

$$\lim_{n \rightarrow \infty} \ln \varphi_{Z_n}(t) = -t^2/2,$$

czyli

$$\lim_{n \rightarrow \infty} \varphi_{Z_n}(t) = e^{-t^2/2}.$$

□

Bezpośrednio z twierdzenia 3.3.1 wynika

Twierdzenie 3.3.2. (*Moivre’a-Laplace’a*) Niech $\Pr(Y_n = k) = b(n, k, p)$, (patrz strona 26), czyli Y_n ma rozkład dwumianowy z parametrami n i p . Wtedy

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{Y_n - np}{\sqrt{npq}} < x \right) = \Phi(x).$$

Asymptotyczna normalność

Własność wyrażoną w twierdzeniu 3.3.2 można interpretować tak, że \widetilde{Y}_n ma dla dużych n rozkład w przybliżeniu normalny $N(0, 1)$ albo że Y_n ma w przybliżeniu rozkład normalny $N(np, \sqrt{npq})$. Mówimy też, że Y_n ma rozkład asymptotycznie normalny z parametrami np i \sqrt{npq} .

3.3.2. Rozkłady chi-kwadrat i t-Studenta

Rozkład chi-kwadrat

Niech $X_i, i = 1, 2, \dots, n$ będą niezależnymi zmiennymi losowymi o rozkładzie normalnym $N(0, 1)$. Zmienną losową χ^2 określa się wzorem:

$$\chi^2 = \chi_n^2 = X_1^2 + X_2^2 + \dots + X_n^2. \quad (3.3.3)$$

Mówimy, że zmienna losowa χ^2 określona wzorem (3.3.3) ma rozkład chi-kwadrat Pearsona¹⁰ o n stopniach swobody. Ponieważ $E \chi^2 = E X_i^2 = \sigma^2$ oraz $ED^2 \chi^2 = D^2 X_i^2 = 2$ to z twierdzenia 3.3.1 wynika asymptotyczna normalność rozkładu zmiennej losowej χ^2 o n stopniach swobody, przy czym ten rozkład normalny ma parametry n i $\sqrt{2n}$, czyli

Asymptotyczna normalność χ^2

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{\chi_n^2 - n}{\sqrt{2n}} < x \right) = \Phi(x).$$

Asymptotyczną normalność rozkładu chi kwadrat można określić też przez zdefiniowanie zmiennej $\sqrt{2}\chi^2$, która też ma rozkład asymptotycznie normalny, ale $N(\sqrt{2n-1}, 1)$.

Zbieżność rozkładu chi-kwadrat do rozkładu normalnego jest tak szybka, że dla $n > 30$ dystrybuenta $\Phi(x)$ przybliża dystrybuentę rozkładu chi-kwadrat wystarczająco dobrze. Z tego powodu dla $n >$

Studenta określony jako rozkład zmiennej losowej t danej wzorem

$$t = t_n = \frac{X}{\sqrt{\chi_n^2/n}}, \quad (3.3.4)$$

gdzie $X \sim N(0, 1)$, a χ_n^2 ma rozkład chi-kwadrat o n stopniach swobody oraz X i χ^2 są niezależne. Rozkład t -Studenta jest również asymptotycznie normalny $N(0, 1)$, przy czym podobnie jak dla rozkładu chi-kwadrat, przybliżenie jest dobre już dla $n > 30$.

3.3.3. Zadania

1. Korzystając z twierdzenia Lindeberga-Levy'ego pokazać, że rozkład chi-kwadrat o n stopniach swobody jest asymptotycznie normalny $N(n, \sqrt{2n})$.
2. Niezależne zmienne losowe X_1, X_2, \dots, X_{60} mają rozkład jednostajny na odcinku $[1, 3]$. Niech

$$X = \sum_{k=1}^{60} X_k$$

Obliczyć przybliżoną wartość wyrażeń $\Pr(X > 130)$, $\Pr(X < 115)$ oraz $\Pr(118 < X < 123)$. Które z tych przybliżeń można otrzymać również z nierówności Markowa lub Czebyszewa? Porównać dokładność tych oszacowań.

3. Niezależne zmienne losowe X_1, X_2, \dots, X_{100} mają rozkład Poissona o parametrze $\lambda = 2.5$. Obliczyć przybliżoną wartość wyrażenia

$$\Pr\left(\sum_{k=1}^{100} X_k > 220\right).$$

4. Korzystając z tablic znaleźć χ_α^2 takie, że $\Pr(\chi^2 > \chi_\alpha^2) = \alpha$ dla $\alpha = 0.1$, $\alpha = 0.05$ i $\alpha = 0.01$, gdy zmienna losowa χ^2 ma rozkład chi-kwadrat Pearsona o $n = 2$, $n = 5$, $n = 15$ i $n > 30$ stopni swobody.
5. To samo co w zadaniu 4, ale dla zmiennej losowej $\tau = |t|$, gdy t ma rozkład t -Studenta.
6. Korzystając z tablic znaleźć $\Pr(\chi^2 < x)$ dla zmiennej losowej o rozkładzie chi-kwadrat Pearsona o $n = 2$, $n = 5$, $n = 15$ i $n > 30$ stopniach swobody oraz $x = 0.5$ i $x = 5$.
7. To samo co w zadaniu 6, ale dla zmiennej losowej t , o rozkładzie t -Studenta.

4. Podstawowe pojęcia statystyki

4.1. Definicje

W rozdziale tym podane są najważniejsze pojęcia statystyczne oraz najczęściej spotykane funkcje, którymi posługuje się statystyka. W tym i w następnych rozdziałach skupiono się na matematycznych treściach statystyki, pozostawiając zagadnienia zastosowań do samodzielnej lektury. Popularne podejście do zagadnień statystyki zawiera przystępnie napisana książka F. Clegga [1]. Podręcznik J. Grenia [3] podaje najpotrzebniejsze definicje i wiele gotowych wzorów potrzebnych do rozwiązywania problemów statystycznych, także wiele zadań oraz niezbędne tablice statystyczne.

4.1.1. Słownik pojęć statystycznych

Statystyka posługuje się specyficznym językiem, dlatego najpierw podamy jej podstawowe pojęcia w postaci słownika.

Populacja generalna – zbiorowość statystyczna, tzn. zbiór dowolnych elementów, nieidentycznych z punktu widzenia badanej cechy X .

Próba – część populacji, dostępnej bezpośredniej obserwacji ze względu na ustaloną cechę, i -ty element populacji ma cechę X_i , czyli próbę można traktować jako ciąg zmiennych losowych (X_1, X_2, \dots, X_n) .

Próba prosta – próba, w której cechy elementów X_i są niezależne i o tym samym rozkładzie co cecha X w populacji generalnej.

Statystyka – zmienna losowa będąca dowolną funkcją wyników próby losowej, tzn. dowolną funkcją $Z = f(X_1, X_2, \dots, X_n)$.

Estymator – dowolna statystyka Z służąca do oszacowania nieznanej wartości parametru θ populacji generalnej.

Hipoteza statystyczna – dowolne przypuszczenie dotyczące rozkładu populacji generalnej.

Test statystyczny – reguła postępowania, która na podstawie wyników próby ma doprowadzić do decyzji przyjęcia lub odrzucenia postawionej hipotezy statystycznej.

Podane powyżej hasła będą dalej uzupełniane o nowe pojęcia.

4.1.2. Najważniejsze statystyki

Momenty zmiennych losowych, zwykle $m_k = E X^k$ i centralne $c_k = E (X - m_1)^k$ zwane są w statystyce momentami teoretycznymi. Momenty empiryczne będziemy dla odróżnienia oznaczać dużymi literami M_k i C_k . Niech (X_1, X_2, \dots, X_n) będzie n -elementową próbą prostą. Określimy k -ty moment empiryczny (zwykły) wzorem

$$M_k = \frac{1}{n} \sum_{i=1}^n X_i^k,$$

a k -ty empiryczny moment centralny wzorem

$$C_k = \frac{1}{n} \sum_{i=1}^n (X_i - M_1)^k .$$

Moment k -ty nazywany jest też momentem rzędu k .

*Średnia
empiryczna \bar{X}*

Podobnie jak dla momentów teoretycznych, dla momentu pierwszego i dla drugiego momentu centralnego istnieją specjalne oznaczenia i nazwy. Są to średnia empiryczna

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i , \quad (4.1.1)$$

*Wariancja
empiryczna S^2*

zwana potocznie statystyką „ X z kreską” oraz wariancja empiryczna

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - M_1)^2 , \quad (4.1.2)$$

zwana statystyką „ S kwadrat”. Ponadto wprowadza się jeszcze wariancję empiryczną poprawioną

$$\hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_1)^2 , \quad (4.1.3)$$

zwaną statystyką „ S kwadrat daszkiem”.

Jest łatwe do sprawdzenia, że $E \bar{X} = E X$, gdy X jest cechą w populacji generalnej, (X_1, X_2, \dots, X_n) jest próbą prostą z tej populacji oraz wartość oczekiwana istnieje. Trudniej już sprawdzić, że $E \hat{S}^2 = D^2 X$, co oznacza też, że

$$E S^2 = \frac{n-1}{n} D^2 X .$$

*Własności
estymatorów*

Niech θ będzie pewnym parametrem rozkładu cechy X w populacji generalnej. Przez $Z_n = \bar{\theta}$ oznaczmy estymator tego parametru, czyli jego statystyczne oszacowanie. Będziemy mówić, że statystyka Z_n jest estymatorem nieobciążonym parametru θ , gdy $E Z_n = \theta$. Estymator jest zgodny, gdy

$$\lim_{n \rightarrow \infty} \Pr (|Z_n - \theta| < \varepsilon) = 1$$

dla każdego $\varepsilon > 0$. Estymator jest asymptotycznie nieobciążony, gdy

$$\lim_{n \rightarrow \infty} E Z_n = \theta .$$

Z powyższych określeń wynika, że \bar{X} jest estymatorem nieobciążonym wartości oczekiwanej, a ze słabego prawa wielkich liczb wynika, że jest również estymatorem zgodnym.

Przykład. Proste obliczenia pokazują, że jeżeli X jest cechą w populacji o rozkładzie normalnym $N(m, \sigma)$, to $\bar{X} \sim N(m, \sigma/\sqrt{n})$.

Trudniej jest dowieść następujące ważne wyniki.

Twierdzenie 4.1.1. *Jeżeli X jest cechą w populacji o rozkładzie normalnym $N(m, \sigma)$, to statystyka $n\hat{S}^2/\sigma^2$ ma rozkład chi-kwadrat o $n - 1$ stopniach swobody.*

Zaskakujący jest natomiast wynik następujący.

Twierdzenie 4.1.2. *Jeżeli cecha X w populacji generalnej ma rozkład normalny, to statystyki \bar{X} i S^2 są niezależne.*

Należy tu zwrócić uwagę, że \bar{X} i S^2 o których mowa w twierdzeniu 4.1.2 pochodzą z **tej samej** próby.

4.1.3. Zadania

1. Dokonano 20 niezależnych prób, otrzymując następujące wyniki: 0.50 0.93 0.75 0.89 0.15 0.94 0.16 0.00 0.63 0.57 0.33 0.10 0.14 0.21 0.05 0.15 0.37 0.51 0.09 0.25. Obliczyć \bar{X} , S^2 i \hat{S}^2 .

2. Dokonano 30 niezależnych prób, otrzymując następujące wyniki: 1.05 1.13 0.41 0.12 0.12 0.19 3.02 0.08 3.87 0.54 2.63 0.40 1.15 0.24 0.46 1.07 0.58 0.29 0.56 2.11 0.40 0.04 0.74 1.41 0.18 3.14 0.40 0.64 0.29 2.47. Obliczyć \bar{X} , S^2 i \hat{S}^2 .

3. Przeprowadzić następujący eksperyment: rzucić 50 razy monetą, obliczyć \bar{X} , S^2 i \hat{S}^2 oraz porównać z $E X$, $D^2 X$.

4**. Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi o jednakowych rozkładach normalnych $N(m, \sigma)$. Dla jakiej wartości k estymator

$$W = k \sum_{j=1}^n |X_j - \bar{X}|$$

jest nieobciążonym estymatorem parametru σ ?

5**. Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi o jednakowych rozkładach normalnych $N(m, \sigma)$. Dobrać stałą k tak, aby funkcja

$$W^2 = k \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2$$

była nieobciążonym estymatorem wariancji.

6. Zmienna losowa X ma rozkład jednostajny na $[a, a + 1]$. Otrzymano n niezależnych obserwacji tej zmiennej losowej. Sprawdzić, że

$$T_n = \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{2}$$

jest estymatorem zgodnym i nieobciążonym parametru a .

4.2. Dystrybuanta empiryczna i histogramy

4.2.1. Dystrybuanta empiryczna

Podobnie jak dla momentów, istnieje empiryczny odpowiednik dystrybuanty, zwanej już dalej teoretyczną. Niech (X_1, X_2, \dots, X_n) będzie próbą prostą. Dystrybuanta empiryczna jest funkcją określoną wzorem:

$$F_n(x) = \frac{1}{n} |\{i : X_i < x\}|.$$

Ponieważ wszystkie $X_i = X_i(\omega)$ są funkcjami określonymi na przestrzeni zdarzeń elementarnych Ω , to dystrybuanta empiryczna $F_n(x)$ jest zmienną losową dla każdego ustalonego x . Jeżeli natomiast ustalimy zdarzenie elementarne ω , to $F_n(x)$ jest funkcją zmiennej rzeczywistej, która spełnia wszystkie w

5. Estymacja

5.1. Estymacja punktowa

5.1.1. Metoda momentów

Metoda momentów jest jedną z wielu metod konstruowania estymatorów parametrów rozkładu cechy w populacji generalnej. Przyjmijmy, że parametr θ jest jednoznacznie określony przez wartości pierwszych k momentów teoretycznych cechy. Oznacza to, że

$$\theta = f(m_1, \dots, m_k).$$

Estymator $\hat{\theta}$ parametru θ określa się wzorem

$$\hat{\theta} = f(\hat{m}_1, \dots, \hat{m}_k)$$

gdzie $\hat{m}_i = M_i$ są momentami empirycznymi. W szczególności parametr θ może być funkcją tylko pierwszego momentu teoretycznego m , wtedy estymator jest funkcją tylko statystyki \bar{X} .

Przykład. Niech X ma rozkład jednostajny na odcinku $[a, b]$. Przyjmijmy, że $a = 0$ oraz $\alpha = b$ jest nieznanym parametrem tego rozkładu. Ponieważ

$$E X = \frac{a+b}{2} \quad D^2 X = \frac{(b-a)^2}{12}, \quad (5.1.1)$$

to $\alpha = 2m$, a więc $\hat{\alpha} = 2\bar{x}$. Nie jest to estymator najlepszy, gdyż można spotkać takie dane, że niektóre z nich wyjdą poza prawy koniec tak oszacowanego przedziału. Jeżeli chcemy oszacować oba końce a i b , które nie są znane, to najpierw trzeba rozwiązać układ dwóch równań (5.1.1) obliczając a i b w funkcji m i σ .

Przykład. Dla rozkładu wykładniczego, którego gęstość jest określona wzorem (2.4.1), $E X = 1/\lambda$. Stąd $\hat{\lambda} = 1/\hat{x}$.

5.1.2. Metoda największej wiarygodności

Idea metody największej wiarygodności polega oszacowaniu nieznanymi parametrów tak, aby empiryczne dane były przy tym oszacowaniu najbardziej prawdopodobne. Dla znalezienia takiego estymatora konstruuje się funkcję wiarygodności L . Jest ona różna dla przypadku ciągłego i dyskretnego.

Najpierw omówimy przypadek dyskretny. Niech $\Pr(X = x^{(k)}) = p(x^{(k)}, \theta)$, gdzie $\theta = (\theta_1, \dots, \theta_m)$ jest wektorem parametrów lub w szczególnym przypadku – parametrem. Funkcję wiarygodności określa się wzorem

$$L(\theta) = L(x_1, x_2, \dots, x_n, \theta) = p(x_1, \theta)p(x_2, \theta) \dots p(x_n, \theta). \quad (5.1.2)$$

*Funkcja
wiarygodności*

Zwróćmy tu uwagę, że $x^{(k)}$ jest ciągiem wartości, które dyskretna zmienna losowa może przyjąć, a x_i są danymi, czyli wartościami, otrzymanymi w rzeczywistości, na przykład z eksperymentu.

Dla cechy typu ciągłego o gęstości $f(x, \theta)$, funkcja wiarygodności określona jest wzorem

$$L(\theta) = L(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta)f(x_2, \theta) \dots f(x_n, \theta). \quad (5.1.3)$$

MNW

Estymatorem parametru θ jest ta jego wartość, przy której funkcja wiarygodności osiąga wartość największą. Jest to estymator otrzymany metodą największej wiarygodności – MNW. Jeżeli funkcja L określona wzorami (5.1.2) lub (5.1.3) jest różniczkowalna, jej maksimum można znaleźć szukając miejsca zerowania się pochodnej. Ze względu na to, że L jest iloczynem funkcji, to wygodnie jest badać pochodną nie L , a pochodną $\ln L$.

Przykład. Niech X ma rozkład Poissona z parametrem λ określonym wzorem (2.3.5). Wtedy dla danych (k_1, k_2, \dots, k_n)

$$L(\lambda) = e^{-n\lambda} \frac{\lambda^{k_1+k_2+\dots+k_n}}{k_1!k_2!\dots k_n!},$$

skąd

$$\ln L(\lambda) = -n\lambda + (k_1 + k_2 + \dots + k_n) \ln \lambda - \ln(k_1!k_2!\dots k_n!).$$

Po obliczeniu pochodnej otrzymujemy dla znalezienia maksimum równanie

$$-n + \frac{k_1 + k_2 + \dots + k_n}{\lambda} = 0,$$

a więc $\hat{\lambda} = \bar{x}$.

Przykład. Dla rozkładu wykładniczego z parametrem λ określonego wzorem (2.4.1),

$$L(\lambda) = \lambda^n e^{-\lambda(x_1+x_2+\dots+x_n)}.$$

Po zlogarytmowaniu i zróżniczkowaniu otrzymujemy równanie

$$\frac{n}{\lambda} - x_1 - x_2 - \dots - x_n = 0,$$

a więc $\hat{\lambda} = 1/\bar{x}$.

Jeżeli L nie jest funkcją różniczkowalną, to jej maksimum nie może być znalezione w ten sposób.

Przykład. Niech X ma rozkład jednostajny na odcinku $[0, \alpha]$. Wtedy

$$L(\alpha) = \begin{cases} 1/\alpha^n & \text{dla } x_i < \alpha, \\ 0 & \text{dla } \alpha < \max\{x_i\}. \end{cases}$$

Taka funkcja L osiąga swoje maksimum w punkcie $\max\{x_i\}$, w którym nie ma pochodnej. Stąd $\hat{\alpha} = \max\{x_i\}$.

5.1.3. Zadania

1. Skonstruować metodą momentów estymatory parametrów p i n w rozkładzie dwumianowym.

2. Gęstość wyraża się wzorem

$$f(x) = \begin{cases} 0, & \text{dla } x < \alpha, \\ \beta/x^2, & \text{dla } x \geq \alpha, \end{cases}$$

gdzie $\alpha > 0$. Obliczyć β i następnie wyznaczyć metodą największej wiarygodności estymator parametru α .

3. Niech gęstość wyraża się wzorem

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

Wyznaczyć metodą największej wiarygodności estymator parametru λ .

4. Niech gęstość wyraża się wzorem

$$f(x) = \begin{cases} \alpha x + \beta, & \text{dla } 0 \leq x \leq 1, \\ 0, & \text{poza tym.} \end{cases}$$

Wyznaczyć metodą momentów estymatory parametrów α i β .

5.2. Estymacja przedziałowa

5.2.1. Przedziały ufności

Poziom
ufności

Idea estymacji przedziałowej polega na tym, aby zamiast szacowania parametru θ za pomocą jednej liczby, znaleźć przedział, w którym nieznany nam parametr znajdzie się z zadowalającym nas prawdopodobieństwem. Końce tego przedziału muszą być wobec tego zmiennymi losowymi – statystykami będącymi funkcjami próby $Z_1 = u_1(X_1, X_2, \dots, X_n)$ i $Z_2 = u_2(X_1, X_2, \dots, X_n)$ takimi, aby $\Pr(\theta \in (Z_1, Z_2))$ było bliskie 1. Bliskość jedynki określa się liczbą $1 - \alpha$ i nazywa poziomem ufności. Łatwo jest zauważyć, że im mniejsze α , tym dłuższy jest przedział ufności. Zazwyczaj α przybiera jedną z wartości 0.1, 0.05, 0.01, przy czym wartość $\alpha = 0.05$ jest najczęściej używana – mówimy wtedy o 95 procentowym przedziale ufności.

Sposób określenia przedziału ufności zależy od rozkładu w którym występuje nieznany parametr, od tego czy znamy pozostałe parametry w tym rozkładzie i od liczebności próby. W następnych punktach omówimy szerzej dwa typowe przypadki: przedziały ufności dla parametru $m = E X$ i $\sigma = \sqrt{D^2 X}$. Wiele innych użytecznych wzorów można znaleźć w książce [3]. Wszystkie te wzory opierają się na tej samej zasadzie, omówionej bardziej szczegółowo w następnym punkcie.

5.2.2. Przedziały ufności dla średniej

Rozpatrywane są trzy przypadki, w zależności od przyjętych założeń. We wszystkich jednak przypadkach, przedział ufności jest symetryczny względem średniej empirycznej \bar{X} określonej wzorem (4.1.1).

Rozkład
normalny,
 σ znane

Model I. Populacja generalna ma rozkład $N(m, \sigma)$, odchylenie standardowe jest znane. Nieznany jest parametr m , dla którego szukamy przedziału ufności. Dla próby o liczebności n , końce przedziału ufności wyrażają się wzorami

$$Z_1 = \bar{X} - u_\alpha \frac{\sigma}{\sqrt{n}},$$

$$Z_2 = \bar{X} + u_\alpha \frac{\sigma}{\sqrt{n}},$$

gdzie u_α jest takie, że $\Pr(|U| > u_\alpha) = \alpha$ oraz $U \sim N(0, 1)$. Wtedy

$$\Pr\left(\bar{X} - u_\alpha \frac{\sigma}{\sqrt{n}} < m < \bar{X} + u_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \quad (5.2.1)$$

Aby dla otrzymanych już danych, a więc ustalonego zdarzenia elementarnego ω wyznaczyć przedział ufności, należy w miejsce \bar{X} we wzorze (5.2.1) podstawić \bar{x} .

*Rozkład
normalny,
 σ nieznane*

Model II. Populacja generalna ma rozkład $N(m, \sigma)$, odchylenie standardowe jest nieznane. Nieznany jest też parametr m , dla którego szukamy przedziału ufności. Dla próby o liczebności n końce przedziału ufności wyrażają się wzorami

$$Z_1 = \bar{X} - t_\alpha \frac{S}{\sqrt{n-1}},$$

$$Z_2 = \bar{X} + t_\alpha \frac{S}{\sqrt{n-1}},$$

gdzie t_α jest takie, że $\Pr(|t| > t_\alpha) = \alpha$ oraz t ma rozkład t -Studenta o $n-1$ stopniach swobody. Statystyka $S = \sqrt{S^2}$ określona jest wzorem (4.1.2). Wtedy

$$\Pr\left(\bar{X} - t_\alpha \frac{S}{\sqrt{n-1}} < m < \bar{X} + t_\alpha \frac{S}{\sqrt{n-1}}\right) = 1 - \alpha \quad (5.2.2)$$

lub równoważnie przy pomocy statystyki $\hat{S} = \sqrt{\hat{S}^2}$

$$\Pr\left(\bar{X} - u_\alpha \frac{\hat{S}}{\sqrt{n}} < m < \bar{X} + u_\alpha \frac{\hat{S}}{\sqrt{n}}\right) = 1 - \alpha. \quad (5.2.3)$$

Ponieważ we wzorach (5.2.1), (5.2.2) i (5.2.3) znamy dokładne rozkłady statystyk, to można je stosować nawet przy małych próbach.

*Rozkład
dowolny, duża
próba*

Model III. Rozkład dowolny, ale n musi być duże, (co najmniej kilka dziesiątków) oraz istnieje wariancja, $\sigma^2 = D^2 X < \infty$. Wtedy przedziały ufności wyznaczane są ze wzoru (5.2.1), przy czym zamiast σ podstawiamy S lub \hat{S} , (dla dużego n różnica między S i \hat{S} jest nieznaczna).

Przykład. Przyjmijmy poziom ufności $1 - \alpha = 0.95$. Ponieważ

$$\Pr(|U| > u_\alpha) = 2\Pr(U > u_\alpha) = 2(1 - \Phi(u_\alpha)) = \alpha = 0.05$$

czyli $\Phi(u_\alpha) = 1 - \alpha/2 = 0.975$. Z tablic rozkładu normalnego odczytujemy, że $u_\alpha = 1.96$. Z $n = 100$ danych o rozkładzie normalnym obliczono $\bar{x} = 2.0031$ i $\hat{s} = 0.1967$.

Ze wzoru (5.2.2) po podstawieniu $\sigma = \hat{s}$ otrzymujemy przedział ufności dla średniej

$$\left(2.0031 - 1.96 \frac{0.1967}{10}, 2.0031 + 1.96 \frac{0.1967}{10}\right),$$

a więc przedział $(1.965, 2.042)$.

Gdyby do obliczeń dostępne były tylko początkowe cztery dane: 2.17, 1.85, 1.87, 1.97, to $\bar{x} = 1.9650$ oraz $\hat{s} = 0.1464$. Wtedy korzystając z rozkładu t -Studenta o trzech stopniach swobody, (model II), otrzymujemy $t_\alpha = 3.182$. Ze wzoru (5.2.2) otrzymujemy przedział ufności $(1.918, 2.011)$.

Jeżeli wiadomo byłoby, że $\sigma = 0.2$, to z modelu I, wzór (5.2.1) mamy jak na początku $u_\alpha = 1.96$, skąd przedział ufności (1.926, 2.004). Dane do tego przykładu zostały wzięte z populacji o rozkładzie $N(2, 0.2)$:

2.17 1.85 1.87 1.97 2.27 2.15 1.74 2.06 1.98 1.94 2.01 1.94 2.27 1.71 2.01 1.59 1.73
2.32 1.83 2.07 1.76 1.97 1.85 1.90 2.30 2.03 1.80 2.12 2.09 1.88 2.03 1.95 1.76 1.88
2.04 1.94 2.24 2.11 1.80 2.03 2.30 2.06 1.89 1.87 2.10 2.06 2.09 2.09 2.13 1.79 1.67
1.82 1.82 1.88 2.17 2.60 2.12 1.65 2.18 2.02 2.07 2.12 1.86 2.40 2.14 1.86 2.07 2.05
1.99 2.03 2.02 2.01 2.19 1.96 2.18 1.99 2.27 2.06 1.59 2.27 1.91 2.36 1.72 1.77 1.80
2.42 2.36 1.74 2.05 1.85 2.10 1.92 2.04 2.13 2.01 1.88 1.82 1.94 2.37 1.72.

Na zakończenie zwróćmy uwagę na istotną rolę założenia o rozkładzie normalnym populacji. Tylko przy tym założeniu próba może być mała. Jeżeli to założenie nie jest spełnione, to musi być duża próba. To ostatnie założenie oznacza asymptotyczną normalność na mocy twierdzenia Lindeberga-Levy'ego, (Twierdzenie 3.3.1).

5.2.3. Przedziały ufności dla wariancji

Przedział ufności dla wariancji nie zależy od wartości oczekiwanej $m = E X$. Stąd tylko dwa rozważane przypadki.

*Rozkład
normalny,
mała próba*

Model I. Populacja generalna ma rozkład normalny. Nieznany jest parametr σ , dla którego szukamy przedziału ufności. Próba jest mała ($n < 30$). Przedział ufności określony jest wzorem

$$\Pr\left(\frac{nS^2}{c_2} < \sigma^2 < \frac{nS^2}{c_1}\right) = 1 - \alpha \quad (5.2.4)$$

lub równoważnie wzorem

$$\Pr\left(\frac{(n-1)\hat{S}^2}{c_2} < \sigma^2 < \frac{(n-1)\hat{S}^2}{c_1}\right) = 1 - \alpha, \quad (5.2.5)$$

gdzie c_1 i c_2 spełniają równania

$$\Pr(\chi^2 < c_1) = \Pr(\chi^2 > c_2) = \alpha/2$$

dla zmiennej losowej χ^2 o rozkładzie chi-kwadrat o $n - 1$ stopniach swobody. Zwróćmy uwagę, że tak otrzymany przedział ufności nie jest symetryczny względem s^2 .

Założenie, że próba jest mała ma charakter czysto rachunkowy – dla $n > 30$ rozkład chi-kwadrat jest na tyle zbliżony do normalnego, że tablice zawierają na ogół wartości tylko do $n = 30$.

*Rozkład
normalny,
duża próba*

Model II. Populacja generalna ma rozkład normalny i $n \geq 30$ lub zbliżony do normalnego i próba jest duża. Wtedy przybliżony przedział ufności wyraża się wzorem

$$\Pr\left(\frac{s}{1 + \frac{u_\alpha}{\sqrt{2n}}} < \sigma < \frac{s}{1 - \frac{u_\alpha}{\sqrt{2n}}}\right) \approx 1 - \alpha, \quad (5.2.6)$$

gdzie u_α jest takie, że $\Pr(|U| > u_\alpha) = \alpha$ oraz $U \sim N(m, \sigma)$.

5.2.4. Zadania

1. Dla danych -0.23, 0.61, -0.85, -0.72, -0.39, 0.73, oszacować na poziomie ufności 0.9 wartość oczekiwaną i wariancję, przyjmując, że rozkład jest normalny.
2. Dla danych -1.09, 0.26, 1.09, 0.56, -1.35, 0.65, oszacować na poziomie ufności 0.9 wartość oczekiwaną i wariancję, przyjmując, że rozkład jest normalny.
- 3*. Napisać procedurę generującą n niezależnych zmiennych losowych $X_i = f(R_i)$, gdzie R_i mają rozkłady jednostajne na odcinku $[0, 1]$, a $f(x)$ jest dowolnie wybraną funkcją, której dziedziną jest odcinek $[0, 1]$. Obliczyć $m = E X_i$ i $\sigma^2 = D^2 X_i$, a następnie z wygenerowanej próby znaleźć przedziały ufności dla m i σ^2 dla poziomów ufności $1 - \alpha = 0.90$, $1 - \alpha = 0.95$ i $1 - \alpha = 0.99$.

6. Testowanie hipotez

6.1. Testy parametryczne

Intuicje testów Testy statystyczne mają za zadanie weryfikację pewnej hipotezy, na podstawie danych statystycznych. Testy parametryczne służą do weryfikacji hipotez o wartościach parametrów w rozkładach zmiennych losowych. Testy nieparametryczne, o których będzie mowa w następnym rozdziale, będą sprawdzać prawdziwość hipotez, w których nie są bądź nie muszą być sprecyzowane wartości parametrów. Testowanie hipotez statystycznych ma (w każdym razie w zakresie tego wykładu) charakterystyczną postać – hipoteza ma postać równości $\theta = \theta_0$, gdzie θ jest prawdziwą, a nam nieznaną wartością parametru rozkładu, natomiast θ_0 jest wartością hipotetyczną. Oznacza to, że taka równość jest sprawdzaną (weryfikowaną) hipotezą, którą można przyjąć (choć tego raczej się nie robi) albo odrzucić i w zamian przyjąć inną, (na przykład $\theta \neq \theta_0$) albo postanowić, że nie ma podstaw do jej odrzucenia, choć nie oznacza to jej przyjęcia, a może oznaczać konieczność przeprowadzenia dalszych badań. Kiedy jesteśmy skłonni hipotezę odrzucić? Intuicyjnie zrobimy tak, gdyby przyjęcie hipotezy oznaczałoby, że zaszło zdarzenie bardzo mało prawdopodobne, na przykład zdarzenie, którego prawdopodobieństwo byłoby mniejsze od $\alpha = 0.05$, czyli takie, które zdarzałoby się średnio 5 razy na 100.

Formalizacja testów Rozumowanie to sprecyzujemy następująco. Niech θ będzie parametrem w pewnym rozkładzie o dystrybuancie $F(x, \theta)$. Niech

$$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta \in \Theta_0$$

oznacza, że stawiamy hipotezę $H_0 : \theta = \theta_0$ zwaną hipotezą zerową którą możemy odrzucić na korzyść hipotezy $H_1 : \theta \in \Theta_0$ zwanej hipotezą alternatywną, którą przyjmujemy. Znak \in oznacza tu jeden z trzech operatorów: $=$, $<$ lub $>$. Z rozkładem $F(x, \theta)$ i parametrem θ wiążujemy statystykę

$$Z = Z(X_1, X_2, \dots, X_n)$$

oraz obszar $Q \subset \mathbb{R}$ służące do weryfikacji hipotezy H_0 w ten sposób, aby przy założeniu prawdziwości H_0 była spełniona równość

$$\Pr(Z \in Q) = \alpha. \tag{6.1.1}$$

Wtedy odrzucamy H_0 i przyjmujemy H_1 o ile istotnie zdarzy się, że $z = Z(\omega) \in Q$, czyli gdy zajdzie zdarzenie mało prawdopodobne. W praktyce statystycznej przyjmuje się zwykle, że $\alpha = 0.05$ albo czasem $\alpha = 0.01$ lub ewentualnie $\alpha = 0.1$.

Obszar krytyczny i poziom istotności Obszar Q nazywa się obszarem krytycznym, a liczbę α nazywa się poziomem istotności. A więc hipotezę zerową odrzucamy na korzyść alternatywnej, gdy wartość związanej z hipotezą statystyki znajdzie się w obszarze krytycznym. Może się oczywiście zdarzyć, że $z \in Q$ mimo, że hipoteza H_0 jest prawdziwa. Zdarzy się to jednak z małym prawdopodobieństwem α . Popelniamy wtedy

6.1 Testy parametryczne

błąd polegający na odrzuceniu hipotezy prawdziwej, zwany błędem pierwszego rodzaju. Przyjęcie hipotezy H_0 fałszywej stanowi błąd drugiego rodzaju. W przyjętej tutaj procedurze nie ma jednak przyjmowania H_0 , co najwyżej stanowi się, że nie ma podstaw do jej przyjęcia. Taką procedurę postępowania przyjęto, gdyż nie precyzuje się tu prawdopodobieństwa popełnienia błędu drugiego rodzaju. W następnych dwóch punktach omówimy dwa przykłady testów statycznych, gdzie nieznanymi parametrami próby będą wartość oczekiwana i wariancja. Liczne przykłady innych testów można znaleźć w książce [3].

6.1.1. Testy dla średniej

Podobnie jak dla przedziałów ufności, rozpatrujemy trzy modele: rozkład normalny i znana wariancja, rozkład normalny i nieznana wariancja, rozkład dowolny ze skończoną wariancją i duża próba. We wszystkich modelach n oznacza liczebność próby.

*Rozkład
normalny,
 σ znane*

Model I. Populacja generalna ma rozkład $N(m, \sigma)$, odchylenie standardowe jest znane. Nieznany jest parametr m , dla którego stawiamy hipotezę $H : m = m_0$ przeciwko jednej z trzech hipotez:

$$H_1 : m \neq m_0,$$

$$H_1 : m > m_0,$$

$$H_1 : m < m_0.$$

Statystyka służąca do weryfikacji hipotezy dana jest wzorem

$$U = \frac{\bar{X} - m_0}{\sigma} \sqrt{n}, \quad (6.1.2)$$

która przy założeniu prawdziwości hipotezy H_0 ma rozkład $N(0, 1)$.

W przypadku hipotezy alternatywnej $H_1 : m \neq m_0$ obszar krytyczny jest dwustronny, symetryczny i dla poziomu istotności α określony jest wzorem $Q = (-\infty, -u_\alpha) \cup (u_\alpha, \infty)$, gdzie u_α wyznaczone jest z zależności $\Pr(|U| > u_\alpha) = \alpha$. Dla hipotezy alternatywnej $H_1 : m < m_0$ obszar krytyczny jest lewostronny i określony jest wzorem $Q = (-\infty, u_\alpha)$, a dla $H_1 : m > m_0$ obszar krytyczny jest prawostronny i określony jest wzorem $Q = (u_\alpha, \infty)$ gdzie u_α wyznaczone jest z zależności $\Pr(U > u_\alpha) = \alpha$. Zmienna losowa U ma tutaj rozkład normalny $N(m, \sigma)$.

*Rozkład
normalny,
 σ nieznane*

Model II. Populacja generalna ma rozkład $N(m, \sigma)$, odchylenie standardowe jest nieznane. Hipoteza zerowa i hipotezy alternatywne są takie same jak w poprzednim modelu. Ponieważ jednak σ nie jest znane, to statystyka służąca do weryfikacji hipotezy dana jest wzorem

$$t = \frac{\bar{X} - m_0}{s} \sqrt{n-1} = \frac{\bar{X} - m_0}{\hat{s}} \sqrt{n}, \quad (6.1.3)$$

która przy założeniu prawdziwości hipotezy H_0 ma rozkład t -Studenta o $n-1$ stopniach swobody. Wobec tego u_α jest zastąpione przez t_α wyznaczone ze wzorów $\Pr(|t| > t_\alpha) = \alpha$ lub $\Pr(t > t_\alpha) = \alpha$.

*Rozkład
normalny,
duża próba*

Ponieważ dostępne tablice statyczne podają wartość t_α dla danych α i n , to przy jednostronnych (lewo i prawostronnych) obszarach krytycznych trzeba skorzystać z zależności $2 \Pr(t > t_\alpha) = \Pr(|t| > t_\alpha)$.

Model III. Populacja generalna ma rozkład $N(m, \sigma)$, parametr σ może, ale musi być znany, natomiast próba jest duża, (n co najmniej kilkadziesiąt). Wzory takie same jak w modelu I lub II, to znaczy (6.1.2) lub (6.1.3), gdzie σ jest zastąpione przez s^2 lub \hat{s}^2 .

Przykład. Posłóźmy się danymi ze strony 53. Dla takich danych mamy $\bar{x} = 2.0031$ i $\hat{s} = 0.1967$. Jeżeli postawimy hipotezę $H_0 : m = 2$ przeciw hipotezie $H_1 : m \neq 2$ na poziomie istotności $\alpha = 0.05$, to obliczona ze wzoru (6.1.2) statystyka $|u| = 0.16 < u_\alpha = 1$.

6.1 Testy parametryczne

Obliczając otrzymujemy $\hat{s}^2 = 0.004779$ oraz $\hat{s} = 0.06913$. Stąd i ze wzoru (6.1.4) mamy $\chi^2 = 17.20$. Ponieważ dla 9 stopni swobody $\chi_\alpha^2 = 16.919$, to odrzucamy hipotezę zerową na korzyść alternatywnej czyli przyjmujemy, że dokładność jest gorsza niż 0.05 jednostki. Jeżeli jednak zadowala nas dokładność 0.1 jednostki, to wtedy $\chi^2 = 4.30$ i stwierdzamy, że nie ma podstaw do odrzucenia takiej hipotezy.

6.1.3. Testy dla dwóch średnich

6.1.4. Zadania

1. Cecha w populacji generalnej ma rozkład normalny. Z pomiarów otrzymano 4 wyniki: 121, 120, 133, 122. Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę, że $m = 125$ oraz hipotezę, że $\sigma = 5$. Na poziomie ufności $1 - \alpha$ oszacować metodą przedziałową parametry m i σ . Powtórzyć obliczenia dla $\alpha = 0.01$ i $\alpha = 0.02$.
2. Cecha w populacji generalnej ma rozkład normalny $N(m, 1.2)$. Z pomiarów otrzymano 5 wyników: 2.3922, 3.9655, 5.2769, 2.2171, 1.9742. Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę, że $m = 3$. Na poziomie ufności $1 - \alpha$ oszacować metodą przedziałową parametr m . Powtórzyć obliczenia dla $\alpha = 0.01$ i $\alpha = 0.02$.
3. Z próby 100 elementowej obliczono $\bar{x} = 1.28$ i $s^2 = 0.21$. Czy na poziomie istotności $\alpha = 0.01$ można zweryfikować hipotezę, że $m = 1.25$ oraz hipotezę, że $\sigma^2 = 0.2$? Na poziomie ufności $1 - \alpha$ oszacować metodą przedziałową parametr m .
- 4*. Dla danych z zadania 3 zweryfikować hipotezę, że m i σ mają wartości obliczone jako parametry teoretyczne. Przyjąć poziom istotności $\alpha = 0.1$, $\alpha = 0.05$ i $\alpha = 0.01$.

6.2. Testy nieparametryczne

6.2.1. Testy zgodności

Testy zgodności służą do weryfikacji hipotez o postaci rozkładów, a dokładniej o postaci dystrybuant rozkładów. Hipoteza zerowa, którą będziemy weryfikować będzie miała postać $H_0 : F(x) = F_0(x)$ lub $H_0 : F(x, \theta) = F_0(x, \theta)$, przeciw hipotezie alternatywnej $H_1 : F(x) \neq F_0(x)$ lub $H_1 : F(x, \theta) \neq F_0(x, \theta)$, gdzie θ jest parametrem rozkładu, które wartość też może być weryfikowana.

Przykład. Hipotezą może być $H_0 : F(x) \sim N(m, \sigma)$, gdzie m i σ są pewnymi, nie interesującymi nas parametrami. Hipoteza może też być postaci $H_0 : F(x) \sim N(0, \sigma)$ albo też $H_0 : F(x) \sim N(0, 1)$.

*Test zgodności
Pearsona*

Test służący do weryfikacji hipotezy o postaci dystrybuanty rozkładu, powinien mierzyć rozbieżności pomiędzy dystrybuantą hipotetyczną $F_0(x)$ a wynikami otrzymanymi z próby. Pierwszym takim testem jest test chi-kwadrat Pearsona. Polega on na tym, że oś liczbową dzielimy na rozłączne przedziały Δ_i punktami d_i , $i = 1, 2, \dots, r - 1$. Otrzymujemy w ten sposób n przedziałów,

$$\Delta_1 = (-\infty, d_1), \Delta_2 = [d_1, d_2), \dots, \Delta_r = [d_{r-1}, \infty).$$

Oznaczmy przez n_i liczbę wyników w przedziale Δ_i , $n = n_1 + n_2 + \dots + n_r$. Niech $p_i = \Pr(X \in \Delta_i)$, gdzie X jest zmienną losową (cechą w populacji generalnej), a $F(x)$ jej dystrybuantą (hipotetyczną). Statystyka

$$\chi^2 = \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i} \quad (6.2.1)$$

mierzy taką właśnie rozbieżność między wynikami otrzymanymi a dystrybuantą hipotetyczną. Należy jednak zwrócić uwagę, że dla efektywnego obliczenia prawdopodobieństw p_i należy znać również wszystkie parametry występujące w dystrybuanecie $F(x)$. Jeżeli nie są one jednak znane, to trzeba je estymować.

Twierdzenie 6.2.1. *Jeżeli nieznane parametry dystrybuanty $F(x)$ są oszacowane metodą największej wiarygodności, to dystrybuenta statystyki określonej wzorem (6.2.1) jest zbieżna dla $n \rightarrow \infty$ do dystrybunaty rozkładu chi-kwadrat Pearsona o $r - k - 1$ stopniach swobody, gdzie k jest liczbą nieznanich parametrów.*

Jeżeli n jest dostatecznie duża (n równe kilkadziesiąt, na przykład $n = 60$) oraz w każdej klasie Δ_i jest co najmniej 8 wyników, $n_i \geq 8$, to można na podstawie twierdzenia 6.2.1 przyjąć, że statystyka określona wzorem (6.2.1) ma w przybliżeniu rozkład chi-kwadrat. Jeżeli w jakiejś klasie jest mniej niż 8 wyników, to taką klasę należy połączyć z sąsiednią. Obszar krytyczny jest postaci przedziału (χ_α^2, ∞) , a więc hipoteza zerowa zostanie odrzucona na poziomie istotności α , gdy $\chi^2 > \chi_\alpha^2$, a χ^2 jest wyznaczone ze wzoru (i odczytane z tablic) $\Pr(\chi^2 > \chi_\alpha^2) = \alpha$.

*Test zgodności
Kołmogorowa*

Innym testem zgodności jest test Kołmogorowa. W porównaniu z testem Pearsona ma on liczne ograniczenia. Wymaga on wyników dokładnych, nie

pogrupowanych w klasy, dużej próby i można nim sprawdzać tylko hipotezy dotyczące rozkładów ciągłych, w których sprecyzowano wartości parametrów. Idea testu Kołmogorowa polega na mierzeniu odchylenia dystrybuanty teoretycznej od empirycznej. Stawiamy hipotezę $H_0 : F(x) = F_0(x)$, gdzie $F_0(x)$ jest dystrybuantą typu ciągłego, z wszystkimi sprecyzowanymi parametrami. Niech $F_n(x)$ będzie dystrybuantą empiryczną. Oznaczmy

$$\lambda = \sqrt{n} \sup_x |F_n(x) - F(x)|. \quad (6.2.2)$$

Twierdzenie 6.2.2. *Statystyka określona wzorem (6.2.2) ma w granicy przy założeniu prawdziwości hipotezy H_0 oraz dla $n \rightarrow \infty$, rozkład niezależny od postaci $F(x)$.*

Rozkład graniczny statystyki λ określonej wzorem (6.2.2) nazywa się rozkładem Kołmogorowa. Jest on stabilizowany, przy czym najczęściej podaje się wartości λ_α , przy których $\Pr(\lambda \geq \lambda_\alpha) = \alpha$. Obszar krytyczny tworzy przedział (λ_α, ∞) , czyli hipotezę H_0 odrzucamy, gdy $\lambda > \lambda_\alpha$.

6.2.2. Testy niezależności

Populację generalną badamy ze względu na dwie cechy, X i Y . Testujemy hipotezę zerową $H_0 : X$ i Y są niezależne, czyli $H_0 : F(x, y) = F_X(x)F_Y(y)$ przeciwko hipotezie alternatywnej $H_1 : F(x, y) \neq F_X(x)F_Y(y)$.

Oznaczmy przez n liczebność próby. Wartości cechy X dzielimy na r klas, a wartości cechy Y na s klas. W ten sposób wszystkie elementy dzielimy na rs klas. Przez n_{ij} oznaczmy liczbę elementów, które ze względu na cechę X są w klasie i -tej, a ze względu na cechę Y są w klasie j -tej. Określmy ponadto liczebności brzegowe

$$n_{i\cdot} = \sum_{j=1}^s n_{ij}, \quad n_{\cdot j} = \sum_{i=1}^r n_{ij}.$$

Wtedy

$$n = \sum_{i=1}^r \sum_{j=1}^s n_{ij} = \sum_{i=1}^r n_{i\cdot} = \sum_{j=1}^s n_{\cdot j}.$$

Z liczebności brzegowych szacuje się prawdopodobieństwa brzegowe

$$p_{i\cdot} = \frac{n_{i\cdot}}{n}, \quad p_{\cdot j} = \frac{n_{\cdot j}}{n}.$$

Zakładając prawdziwość H_0 , to znaczy niezależność cech X i Y , oblicza się prawdopodobieństwa hipotetyczne

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

Stąd podobnie jak we wzorze (6.2.1) konstruuje się statystykę

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - np_{ij})^2}{np_{ij}}. \quad (6.2.3)$$

Twierdzenie 6.2.3. *Dystrybuanta statystyki określonej wzorem (6.2.3) jest zbieżna dla $n \rightarrow \infty$ do dystrybucyjności rozkładu chi-kwadrat Pearsona o $(r - 1)(s - 1)$ stopniach swobody.*

Obszar krytyczny jest określony tak samo jak dla statystyki (6.2.1). Ze względu na to, że statystyka (6.2.3) ma rozkład chi-kwadrat tylko przy $n \rightarrow \infty$, to n musi być bardzo duże, a wszystkie $n_{ij} \geq 8$.

Jeżeli $r = s = 2$, to tablica $[n_{ij}]$ nazywa się tablicą czteropolową. W tym przypadku mamy jeden stopień swobody statystyki χ^2 .

6.2.3. Zadania

1. Przy pomocy testu chi-kwadrat Pearsona sprawdzić, czy cecha w populacji generalnej ma rozkład jednostajny, gdy z próby uzyskano następujące wyniki: 2.07, 2.35, 1.52, 2.73, 2.30, 2.56, 1.46, 2.61, 2.86, 1.82, 1.81, 2.56, 1.97, 1.21, 2.16, 1.66, 2.62, 2.07, 1.29, 2.32. Czy można odrzucić hipotezę, że rozkład jest jednostajny na odcinku $[1, 3]$ lub $[1.2, 2.2]$? Przyjąć poziom istotności $\alpha = 0.05$.

2. Dane z próby n -elementowej zostały pogrupowane w tabeli, gdzie

Przedział	Liczba wyników
0 – 1	44
1 – 2	36
2 – 3	20
3 – 4	14
4 – 5	10
5 – 6	8
6 – 7	4
7 – 8	6
8 – 9	5
9 – 10	3

Na poziomie istotności $\alpha = 0.02$ zweryfikować hipotezę, że rozkład jest a) wykładniczy, b) wykładniczy z parametrem $\lambda = 0.5$.

3*. Wygenerować w dowolny sposób 100 liczb o rozkładzie normalnym z parametrami $N(m, \sigma)$, a następnie zweryfikować przy pomocy testu chi-kwadrat Pearsona, że jest to a) rozkład normalny, b) rozkład normalny z parametrem m , c) rozkład normalny z parametrem σ , d) rozkład normalny z parametrami m i σ .

4*. Dla danych z zadania 3 zweryfikować przy pomocy testu Kołmogorowa hipotezę, że są to dane z rozkładu normalnego z parametrami m i σ .

5. Niech x_i będzie liczbą rzutów kostką, w których wyrzucono i oczek. W 180 rzutach otrzymano $x_1 = 25$, $x_2 = 31$, $x_3 = 24$, $x_4 = 32$, $x_5 = 28$. Czy taki wynik można na poziomie istotności $\alpha = 0.05$ usprawiedliwić przypadkiem?

6. Każdy element ma dwie cechy, cechę $X \in (0, 1)$ oraz cechę Y przyjmującą tylko wartości 0 i 1. Czeropolowa tablica wygląda następująco:

	$X < 0.5$	$X \geq 0.5$
$Y = 0$	72	29
$Y = 1$	53	26

Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę o niezależności cech X i Y .

Literatura

- [1] F. Clegg, Po prostu statystyka, Wyd. Szk. i Pedagog., Warszawa 1994.
- [2] M. Fisz, Rachunek prawdopodobieństwa, PWN, (wiele wydań).
- [3] J. Greń, Statystyka matematyczna, modele i zadania, PWN, (wiele wydań)
- [4] T. Inglot, T. Ledwina, Z. Ławniczak, Materiały do ćwiczeń z rachunku prawdopodobieństwa i statystyki matematycznej, skrypt P. Wr., 1979.
- [5] W. Feller, Wstęp do rachunku prawdopodobieństwa, t. I, PWN, 1966.
- [6] W. Feller, Wstęp do rachunku prawdopodobieństwa, t. II, PWN, 1969.